Will an outbreak exceed available resources for control? Estimating the risk from invading pathogens using practical definitions of a severe epidemic

AUTHORS

R.N. Thompson[1,2,*], C.A. Gilligan[3], N.J. Cunniffe[3]

*Correspondence to: robin.thompson@chch.ox.ac.uk

AFFILIATIONS

[1]Mathematical Institute, University of Oxford, Oxford, UK

[2]Christ Church, University of Oxford, Oxford, UK

[3]Department of Plant Sciences, University of Cambridge, Cambridge, UK

ABSTRACT (200 words)

Forecasting whether or not initial reports of disease will be followed by a severe epidemic is an important component of disease management. Standard epidemic risk estimates involve assuming that infections occur according to a branching process and correspond to the probability that the outbreak persists beyond the initial stochastic phase. However, an alternative assessment is to predict whether or not initial cases will lead to a severe epidemic in which available control resources are exceeded. We show how this risk can be estimated by considering three practically relevant potential definitions of a severe epidemic; namely, an outbreak in which: i) a large number of hosts are infected simultaneously; ii) a large total number of infections occur; and iii) the pathogen remains in the population for a long period. We show that the probability of a severe epidemic

under these definitions often coincides with the standard branching process estimate for the major epidemic probability. However, these practically relevant risk assessments can also be different from the major epidemic probability, as well as from each other. This holds in different epidemiological systems, highlighting that careful consideration of what constitutes a severe epidemic in an ongoing outbreak is vital for accurate risk quantification.

KEYWORDS

mathematical modelling; infectious disease epidemiology; major epidemic; severe epidemic; forecasting

SHORT TITLE

Assessing the practically relevant risk of a severe epidemic

1. INTRODUCTION

Infectious disease epidemics in populations of humans, animals and plants represent a recurring risk worldwide [1–7]. An important aim for policy-makers near the start of an outbreak is to assess the risk posed by the invading pathogen, including whether initial cases will lead on to a major epidemic or whether the pathogen will die out rapidly instead [8,9]. An important practical consequence is that, if an outbreak is likely simply to fade out, then costly interventions such as vaccination [10,11], culling/felling/roguing of plants

or agricultural animals [12–18] and workplace or school closure [19] may be unnecessary [20].

There is a well-known estimate for the probability of a major epidemic when a pathogen is newly arrived in a host population, which in its simplest form is given by

$$\text{Prob(major epidemic)} = 1 - \left(\frac{1}{R_0}\right)^{I(0)}, \qquad (1)$$

in which R_0 is the basic reproduction number of the pathogen and $I(0)$ is the number of individuals that are currently infected. The estimate in equation (1) applies to a wide range of models, including the commonly used Susceptible-Infected-Susceptible (SIS) and Susceptible-Infected-Removed (SIR) models [21]. It is derived by assuming that infections occur according to a branching process (see Methods). For the commonly used Susceptible-Exposed-Infectious-Removed model, the exponent in equation (1) would change from $I(0)$ to $E(0) + I(0)$ [9]. More sophisticated estimates that are based on similar branching process approximations can be derived for models including additional epidemiological detail, such as more complex population structure [22–24] and/or infectious periods that are not exponentially distributed [25,26].

The quantity in equation (1), and particularly the version in which $I(0) = 1$, is used extensively in the epidemiological modelling literature [8,9,21,26–35]. It is increasingly used in real-time during emerging outbreaks. For example, it was used during the 2014-16 epidemic of Ebola virus disease in West Africa to estimate the chance that, if the virus

arrived in Nigeria, sustained transmission would follow in that country [30]. It was considered in the context of flare-ups in new locations for the recent Ebola epidemic in the Democratic Republic of the Congo [26]. Branching process models were also used at the start of the COVID-19 pandemic before cases were detected outside China to assess the risk of epidemics elsewhere [36,37], including the application of equation (1) [36].

However, while the major epidemic probability in equation (1) is useful to assess whether or not an outbreak is likely to persist beyond the initial stochastic phase, becoming a major epidemic does not guarantee that the outbreak will overwhelm available control resources. Over many outbreaks under identical conditions, if the population size is large and R_0 is much greater than one, then the distribution of possible epidemic sizes is bimodal according to simple epidemic models such as the stochastic SIR model (Fig 1d – see also [38–42]). In other words, the final size of any single outbreak is almost always in one of two possible ranges. For example, in Fig 1d, virtually all outbreaks either lead to 1-20 hosts ever infected or to 700-860 hosts ever infected, where the precise ranges depend on the population size and the value of R_0. The estimate for the probability of a major epidemic in equation (1) corresponds approximately to the proportion of outbreaks that have a final size in the higher of these ranges. The outbreaks within the higher range, however, do not necessarily represent outbreaks in which available control resources are exceeded. For practical assessments of the threat from an invading pathogen, it would often therefore be appropriate for the notion of a severe epidemic to be grounded in consequences for disease control, depending on the specific system and outbreak under consideration.

Here, we assess whether or not outbreaks are likely to develop into severe epidemics according to three possible metrics that might be practically relevant in different outbreak scenarios. Specifically, these are:

Concurrent size. In this assessment, a severe epidemic is an outbreak in which the number of individuals infected simultaneously exceeds the capacity for treatment.

Total infections. In this assessment, a severe epidemic is an outbreak in which the total number of infections exceeds the number of available treatments.

Duration. In this assessment, a severe epidemic is an outbreak that is not contained quickly and therefore persists for an unacceptably long period.

We compare the probability of a severe epidemic under each of these definitions, as well as calculate the branching process estimate for probability of a major epidemic (hereafter, we differentiate between the “probability of a severe epidemic” calculated using one of the metrics above and the “probability of a major epidemic” calculated by assuming that infections occur according to a branching process). In our analyses, as examples we consider three stochastic epidemiological models that are representative of different host responses to infection and capture different routes of transmission. Specifically, we consider the SIS model, the SIR model, and a host-vector model parameterised for Zika virus transmission. For the SIS and SIR models, the probability of a major epidemic corresponds to equation (1), and in the case of Zika virus the probability of a major epidemic is given by an adapted version of equation (1) that accounts for transmission

between hosts by vectors (we present an approach for deriving these well-known formulae in Methods).

To motivate our analyses, we note that estimation of the risk of the outbreak going on to have a large concurrent size, a large total number of infections or a long duration might be the appropriate risk assessment in different scenarios. For example, it might be natural to assume that, if the number of individuals infected at any time always remains below the capacity for treatment, then the outbreak is not severe since medical care is available for all individuals that require treatment. Indeed, in the ongoing COVID-19 pandemic, one of the main aims of interventions in the UK has been to ensure that the number of individuals requiring intensive care unit beds remains below the total number of beds available [43–45]. More generally, the threshold capacity might derive from the number of available beds in hospitals or treatment units [46,47], or the availability of care workers [48]. This motivates consideration of the “concurrent size” metric above.

However, assessing the outbreak risk based on numbers of hosts infected simultaneously will not always be appropriate. Policy-makers often have to make decisions concerning how much treatment to stockpile; if all cases must be treated, this corresponds to the total number of infections during the outbreak. For example, in response to growing awareness of the threat of a potential influenza pandemic, between 2006 and 2013 policy-makers in the UK stockpiled around 40 million units of antivirals at a cost of £424 million. This led to severe criticism when only 2.4 million units were needed, the majority of which were used during the 2009 H1N1 influenza pandemic [49]. Another possible risk assessment is

therefore whether or not the total number of infections will exceed a critical value (the "total infections" metric above). This critical value might be set by the stock of available treatments for use during the outbreak.

Finally, we consider a third possible risk assessment (using the "duration" metric). In this scenario, we evaluate whether or not an outbreak is likely to persist for an unacceptably long period. An outbreak that fades out quickly may escape public attention. Even if an outbreak leads to a significant number of hosts infected, if it ends relatively quickly then it might not be considered a severe epidemic. For example, the first Ebola outbreak in the Democratic Republic of the Congo in 2018 resulted in 53 cases, but was not considered a severe epidemic due to its fast containment, leading to commendation of the success of public health measures [50]. Consequently, an outbreak might only be classified as a severe epidemic if it persists for a threshold length of time.

An outbreak might be classified as a severe epidemic according to one of the metrics above, yet not be a severe epidemic if another metric is used. In 1665-66, plague affected the village of Eyam in the UK, which famously isolated itself via a self-imposed quarantine [51,52]. The outbreak in the village was long-running, and a large number of individuals were killed (most reports suggest 250-260 out of a total of 350 in the village died, although there is some uncertainty particularly regarding the size of the at-risk population [53]). However, model fits suggest that a maximum of only around 30 people were ever infected simultaneously [54–56]. As a result, this epidemic might have been classified as severe according to the "total infections" and "duration" metrics, yet not the "concurrent size"

metric, depending on the precise values of the thresholds set in each case. This highlights the need to consider the appropriate metric for defining severe epidemics in the particular ongoing outbreak under consideration in order to perform the most practically relevant risk assessment.

A large body of theoretical work exists relating to the metrics for defining severe epidemics that we consider. For example, for the stochastic SIS and SIR models, probability distributions for the maximum number of individuals infected concurrently prior to epidemic extinction have been derived previously [57–59]. For R_0 significantly greater than one, a related quantity (the quasi-stationary distribution – the distribution of the number of infected individuals in the long phase of the epidemic prior to extinction) has been studied in detail for models in which the pathogen persists long-term, including the stochastic SIS model [60–65] and the stochastic SIR model with births and deaths [66]. Analytic expressions and approximations have been found for the total number of infections over the course of an outbreak for a range of epidemic models [25,41,59,67–69] and different approaches exist for calculating probability mass functions describing the possible final sizes of a stochastic epidemic (for a review of approaches for the stochastic SIR model, see [59]). The duration of an epidemic has also been well-studied [60,70–75], as has the duration of the initial stochastic phase of outbreaks that go on to become major epidemics [76].

However, previously developed mathematical theory is not the focus of our analyses. Instead, the novelty of the research that we present is to compare assessments of the

risk from invading pathogens evaluated in different ways. We demonstrate the general principle that the precise definition of a severe epidemic (i.e. the metric chosen to define a severe epidemic, or the choice to use the standard branching process estimate for the major epidemic probability) can affect the assessments of the risk whenever a pathogen arrives in a new host population. The probabilities that an outbreak has a large concurrent size, a large total number of infections or a long duration may not coincide and depend on precisely which values of the relevant thresholds are set. These probabilities may or may not match the major epidemic probability assessed in the standard way. Careful consideration of precisely how a severe epidemic is classified is therefore necessary whenever the risk from an invading pathogen is estimated at the beginning of an emerging outbreak. Only once the notion of a severe epidemic has been formally defined – based on criteria of practical relevance – can this risk be properly assessed.

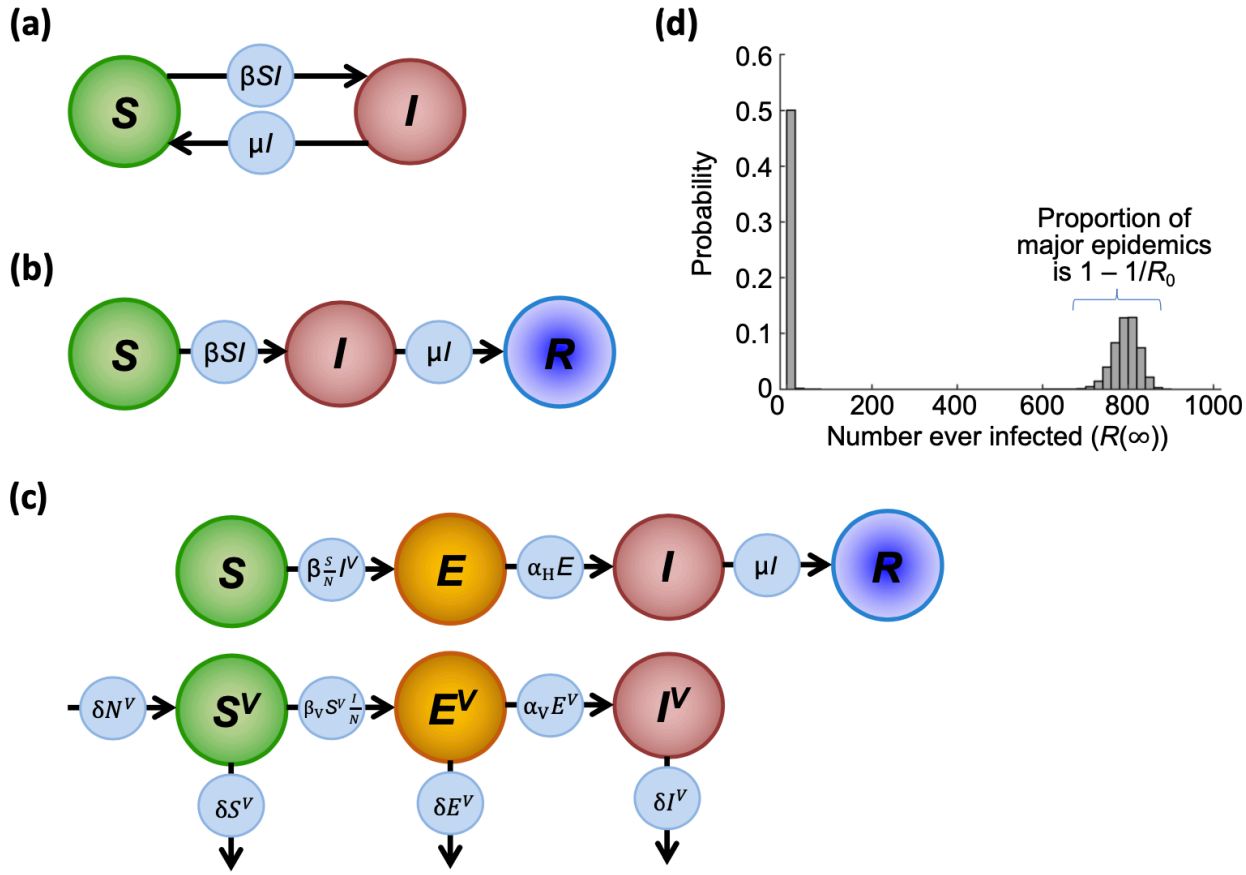

Figure 1. Schematic diagrams illustrating the population structures for the different models considered, and an example distribution of final sizes for the stochastic SIR model. (a) The SIS model. (b) The SIR model. (c) The host-vector model of Zika virus transmission. (d) Distribution of final sizes in the stochastic SIR model, with population size N = 1000, R_0 = 2, $I(0)$ = 1 and the rest of the population susceptible initially. The x-axis has been split into bars of width 20 (so that, for example, the first bar corresponds to the probability that between 1 and 20 individuals are ever infected).

2. METHODS

We conduct five analyses. In the first three analyses, we consider the stochastic SIS model, the SIR model and the Zika host-vector model, and assess the risk that a single

initial case will lead to an outbreak with a large “concurrent size”. Our final two analyses focus on the stochastic SIS model. We calculate the probability of an outbreak going on to exceed a pre-specified total number of infections (the “total infections” metric) or time (the “duration” metric).

Here, we describe the epidemiological models that we use, calculation of the branching process estimate of the major epidemic probability for each model, and how the probability of a severe epidemic can be calculated under the “concurrent size” metric for each of the models considered. We then explain how the probability of a severe epidemic under the other practically relevant metrics can be obtained for the SIS model, although our methodology generalises immediately to any model for which a method of stochastic simulation is available.

Epidemiological models

Susceptible-Infected-Susceptible (SIS) model

According to the SIS model, at any time each individual in the population is classified to be either (*S*)usceptible to or (*I*)nfected by the pathogen. The deterministic SIS model is given by

$$\frac{\mathrm{d}S}{\mathrm{d}t} = -\beta IS + \mu I,$$

$$\frac{\mathrm{d}I}{\mathrm{d}t} = \beta IS - \mu I, \quad (2)$$

where β represents the infection rate between each susceptible-infected pair and μ is the rate at which each infected host recovers and becomes susceptible again. We use the analogous stochastic model in most of our analyses, in which the net rate at which any epidemiological event occurs is $\beta IS + \mu I$. At any time prior to the end of the outbreak, the probability that this next event is an infection is $\frac{\beta IS}{\beta IS + \mu I}$ and the probability that the next event is a recovery is $\frac{\mu I}{\beta IS + \mu I}$.

In this model, if the total population size is *N*, the basic reproduction number is given by

$$R_0 = \frac{\beta N}{\mu}.$$

Susceptible-Infected-Removed (SIR) model

Under the SIR model, at any time each individual in the population is classified according to whether they are (*S*)usceptible to infection, (*I*)nfected, or (*R*)emoved and no longer spreading the pathogen or available for infection. The deterministic SIR model is given by

$$\frac{\mathrm{d}S}{\mathrm{d}t} = -\beta IS,$$

$$\frac{\mathrm{d}I}{\mathrm{d}t} = \beta IS - \mu I,$$

$$\frac{\mathrm{d}R}{\mathrm{d}t} = \mu I, \quad (3)$$

in which β is again a parameter governing the infection rate, and μ is the removal rate. In the analogous stochastic model, the net rate at which any epidemiological event

occurs is still $\beta IS + \mu I$, and the probability that the next event is an infection event is similarly unchanged at $\frac{\beta IS}{\beta IS+\mu I}$. However, the other possible next event is a removal, which occurs with probability $\frac{\mu I}{\beta IS+\mu I}$. The basic reproduction number is again

$$R_0 = \frac{\beta N}{\mu}.$$

Zika transmission model

We consider the transmission of Zika virus according to a host-vector model [77], which we chose to demonstrate how the probability of a major epidemic can be calculated in a relatively complex epidemiological setting. In the model, the numbers of the *N* hosts that are (*S*)usceptible, (*E*)xposed, (*I*)nfectious and (*R*)emoved are tracked, as well as the number of the N^V vectors that are (S^V)usceptible, (E^V)xposed and (I^V)nfectious. We adapt the version of the model as presented by Kucharski *et al.* [77] slightly to a more standard formulation in which all transmission terms are proportional to the relevant number of vectors and density of hosts (i.e., when most hosts are susceptible early in the outbreak, hosts are sufficiently abundant that the number of hosts that each vector bites is independent of the total number of hosts). The deterministic version of this model is then

$$\frac{\mathrm{d}S}{\mathrm{d}t} = -\beta I^V \frac{S}{N},$$

$$\frac{\mathrm{d}E}{\mathrm{d}t} = \beta I^V \frac{S}{N} - \alpha_H E,$$

$$\frac{\mathrm{d}I}{\mathrm{d}t} = \alpha_H E - \mu I,$$

$$\frac{\mathrm{d}R}{\mathrm{d}t} = \mu I,$$

$$\frac{\mathrm{d}S^V}{\mathrm{d}t} = \delta N^V - \beta_V S^V \frac{I}{N} - \delta S^V,$$

$$\frac{\mathrm{d}E^V}{\mathrm{d}t} = \beta_V S^V \frac{I}{N} - (\delta + \alpha_V) E^V,$$

$$\frac{\mathrm{d}I^V}{\mathrm{d}t} = \alpha_V E^V - \delta I^V. \qquad (4)$$

The parameters β and β_V govern the rates at which infectious vectors infect susceptible hosts and infectious hosts infect susceptible vectors, respectively. The mean latent periods of infections in hosts and vectors are given by $1/\alpha_H$ and $1/\alpha_V$. The parameter μ is the rate of removal of infectious hosts, and δ describes the death rate of every vector. In the analogous stochastic model, the number of infected human hosts arising from a single infected human (accounting for human-vector-human transmission) in an otherwise entirely susceptible population of humans and vectors is given by

$$R_0^{HV} \times \rho^{E^V \to I^V} \times R_0^{VH} = \frac{\beta_V \alpha_V \beta N^V}{\mu(\delta + \alpha_V)\delta N},$$

where $R_0^{HV} = \frac{1}{\mu}\frac{\beta_V N^V}{N}$ is the expected number of vectors infected (and going on to enter the exposed class) by a single infectious human, $\rho^{E^V \to I^V} = \frac{\alpha_V}{\delta + \alpha_V}$ is the proportion of exposed vectors that become infectious and $R_0^{VH} = \frac{\beta}{\delta}$ is the expected number of humans infected by a single infectious vector.

The basic reproduction number is given by

$$R_0 = \sqrt{\frac{\beta_V \alpha_V \beta N^V}{\mu(\delta + \alpha_V)\delta N}},$$

where the square root accounts for the fact that it takes two generations for infected humans to generate new infections, since new infections require host-vector-host transmission [78,79]. We note that in some studies, e.g. [77], the square root is omitted from the definition of R_0. In contrast to the expression calculated by Kucharski *et al*. [77], to facilitate simulation of the stochastic model we explicitly track the total number of vectors, N^V, rather than the density.

<u>Probability of a major epidemic (branching process estimate)</u>

Standard estimate (stochastic SIS/SIR models)

The commonly used estimate for the major epidemic probability when a pathogen first arrives in a host population [8,9,35,36,21,27,29–34] can be derived by assuming that infections occur according to a branching process, making the assumptions that the susceptible population is large and that infection lineages arising from different infected hosts are independent. When a single infected host arrives in an otherwise susceptible population, the branching process estimate for the major epidemic probability is given by

$$\text{Prob(major epidemic)} \approx \begin{cases} 0 \ \text{ for } R_0 \leq 1, \\ 1 - \dfrac{1}{R_0} \ \text{ for } R_0 > 1. \end{cases}$$

This expression is derived in Text S1.

If instead there are $I(0)$ infected individuals initially rather than one, then for no major epidemic to occur, it is necessary for each initial infection lineage to die out, leading to the approximation given in equation (1) whenever $R_0 > 1$.

Standard estimate (more complex models)

Here we show how the branching process estimate for the major epidemic probability can be derived for more complex systems by considering the stochastic version of the host-vector model given by the system of equations (4). Invasion probabilities have been estimated previously for vector-borne pathogens, for example by considering the distributions of secondary infections from infected hosts/vectors and their associated probability generating functions [22], however we use an alternative approach here (a "first step analysis", as in [80]) that is straightforward to understand and extend to a range of epidemiological settings.

We denote the probability of no major epidemic occurring starting from i exposed or infectious human hosts, j exposed vectors and k infectious vectors by q_{ijk}. We must consider exposed and infectious vectors separately to account for the possibility that exposed vectors die before becoming infectious.

Starting from a single infectious host introduced into an entirely susceptible population of hosts and vectors, and conditioning on the next event, gives

$$q_{100} = \frac{\mu}{\mu + \frac{\beta_V N^V}{N}} q_{000} + \frac{\frac{\beta_V N^V}{N}}{\mu + \frac{\beta_V N^V}{N}} q_{110},$$

because the next event can then either be recovery of the infectious host (with probability $\frac{\mu}{\mu + \frac{\beta_V N^V}{N}}$) or infection of a susceptible vector (with probability $\frac{\frac{\beta_V N^V}{N}}{\mu + \frac{\beta_V N^V}{N}}$). Similarly, starting from a single exposed or infectious vector gives

$$q_{010} = \frac{\delta}{\delta + \alpha_V} q_{000} + \frac{\alpha_V}{\delta + \alpha_V} q_{001},$$

$$q_{001} = \frac{\delta}{\delta + \beta} q_{000} + \frac{\beta}{\delta + \beta} q_{101}.$$

We again assume that infection lineages are independent, permitting us to approximate terms with two exposed or infectious individuals by non-linear terms involving single exposed or infectious individuals, e.g. $q_{110} \approx q_{100} q_{010}$. Noting that $q_{000} = 1$, the three equations above can be solved to give expressions for q_{100}, q_{010} and q_{001}. In particular, the probability of a major epidemic starting from a single infected host is then

$$1 - q_{100} = \begin{cases} 0 & \text{for } R_0 \leq 1, \\ \dfrac{(R_0)^2 - 1}{(R_0)^2 + R_0^{VH}} & \text{for } R_0 > 1. \end{cases} \qquad (5)$$

In this expression, R_0^{VH} is the expected number of humans infected by a single infectious vector in an otherwise entirely susceptible population of humans and vectors.

<u>Probability of a severe epidemic (“concurrent size” metric)</u>

As described in the introduction, we first define a major epidemic to be an outbreak in which the maximum number of individuals infected simultaneously is above a threshold value, which we denote by *M*. The value of *M* of relevance in practical applications might be set by the capacity for treatment.

Stochastic SIS model

Under the stochastic SIS model, the probability that the number of individuals infected simultaneously is at least *M* at some time prior to epidemic extinction can be calculated analytically [57]. This is advantageous since approximating this quantity using model simulations can be time consuming given that outbreaks under the SIS model can persist for long periods.

We denote the probability that the maximum number of individuals simultaneously infected (also known as the "peak prevalence" [59]) is at least *M,* starting from *I* infected individuals, as q_I, and assume that the rest of the population is susceptible. Conditioning on whether the first event is an infection or recovery event gives

$$q_I = \frac{\beta IS}{\beta IS + \mu I} q_{I+1} + \frac{\mu I}{\beta IS + \mu I} q_{I-1},$$

for $I = 1,2,\ldots,M-1$. We solve this tridiagonal system of equations with boundary conditions $q_0 = 1$ and $q_M = 0$, since the peak number infected will certainly be less than *M* if there are initially no infected individuals (assuming $M > 0$), and will be at least *M* if there are

initially M infected individuals. Noting that $S = N - I$, these equations can be put in the form

$$q_{I+1} - q_I = C_I(q_I - q_{I-1}),$$

where

$$C_I = \frac{\mu I}{\beta I(N - I)}.$$

Iterating this gives

$$q_{I+1} - q_I = A_I(q_1 - q_0), \qquad (6)$$

where

$$A_I = \prod_{j=1}^{I} C_j,$$

$$= \frac{\left(\frac{N}{R_0}\right)^I}{I!\binom{N-1}{I}}.$$

Adding equation (6) to itself for I = 1,2,…,M - 1 and rearranging gives

$$q_1 = \frac{\sum_{I=1}^{M-1} A_I}{1 + \sum_{I=1}^{M-1} A_I}.$$

Instead adding equation (6) to itself for I = 1,2,…,k – 1 gives

$$q_k = \frac{\sum_{I=k}^{M-1} A_I}{1 + \sum_{I=1}^{M-1} A_I}.$$

The probability of a severe epidemic according to the "concurrent size" metric is then given by

$$\text{Prob}(\text{severe epidemic}) = 1 - q_{I(0)} = \begin{cases} 0 \text{ for } I(0) = 0, \\ \dfrac{1}{1 + \sum_{I=1}^{M-1} A_I} \text{ for } I(0) = 1, \\ \dfrac{1 + \sum_{I=1}^{I(0)-1} A_I}{1 + \sum_{I=1}^{M-1} A_I} \text{ for } 1 < I(0) < M, \\ 1 \text{ for } I(0) \geq M. \end{cases} \quad (7)$$

Stochastic SIR model

Under the stochastic SIR model, the probability that the maximum number infected simultaneously is at least *M* (i.e. a severe epidemic occurs, under the "concurrent size" metric) starting from any state (*I*,*R*) is calculated using an iterative approach [58,59]. Denoting the probability of a severe epidemic starting from state (*I*,*R*) by $p_{I,R}$, then conditioning on the next event gives

$$p_{I,R} = \frac{\beta I(N - I - R)}{\beta I(N - I - R) + \mu I} p_{I+1,R} + \frac{\mu I}{\beta I(N - I - R) + \mu I} p_{I-1,R+1}.$$

This system can be solved with boundary conditions $p_{0,R} = 0$, $p_{I,N-M+1} = 0$ and $p_{M,R} = 1$. Doing so does not require this system of equations to be solved simultaneously. Instead, the value of $p_{I,R}$ is deduced for the following states (in order): (*I*,*R*) = (*M*-1,*N*-*M*), (*M*-2,*N*-*M*), …, (1,*N*-*M*), (*M*-1,*N*-*M*-1), …,(1,*N*-*M*-1),… (*M*-1,0),…,(1,0). For a schematic showing the order in which these probabilities are deduced, see Fig S1.

Zika transmission model

The probability of a severe epidemic using the "concurrent size" metric is approximated by simulating the model 10,000 times using the Gillespie direct method [81], and then calculating the proportion of simulations in which the number of infected human hosts is at least *M* at any point during the simulation.

<u>Probability of a severe epidemic ("total infections" and "duration" metrics)</u>

We also consider the probability of a severe epidemic according to the stochastic SIS model for the "total infections" and "duration" metrics. Specifically, we estimate the probability that at least *F* infections occur over the course of the outbreak (prior to outbreak extinction), and the probability that the outbreak persists for at least *T* days.

We approximate these quantities by simulating the model 10,000 times using the Gillespie direct method [81] and recording separately the proportion of simulations in which there are at least *F* infections or that have a duration of at least *T* days. Each simulation is stopped when either of these following two criteria are satisfied: i) the simulated outbreak

has gone extinct (I = 0), or; ii) both the number of infections has reached the maximum value of F considered (F = 2000; see Fig 5a) and the duration has reached the maximum value of T considered (T = 6000; see Fig 5b).

3. RESULTS

To begin to understand outbreak dynamics under the SIS, SIR and Zika transmission models, we first numerically solved the deterministic models given by the systems of equations (2), (3) and (4) with $R_0 = 1.5$ in each case (Fig 2). For the parameter values considered, in a deterministic setting the SIS model predicts the largest number of individuals infected simultaneously as well as the most infections in total. Epidemics persisted forever (i.e. I remained larger than zero) under all three models, although the number of infected hosts tended to zero under the SIR and Zika transmission models.

However, our main focus is assessing the risk from an invading pathogen, according to the “concurrent size”, “total infections” and “duration” metrics for defining severe epidemics. We therefore considered the more realistic stochastic models, in which demographic stochasticity is included. In the following sections, first we calculate the probability of a severe epidemic for the stochastic SIS model under the “concurrent size” metric. We then consider different epidemiological models, as well as other metrics defining a severe epidemic. In each case, the probability of a severe epidemic for the particular epidemiological model-severe epidemic metric pair under consideration is compared with the branching process approximation to the probability of a major epidemic for that model. The rationale for this comparison is that both quantities represent a

possible way to assess the risk from an invading pathogen. Results are shown in Figs 3-5 and summarised in Tables 1 and 2. A chart outlining the model-severe epidemic metric pairs considered in each figure is shown in Fig S2.

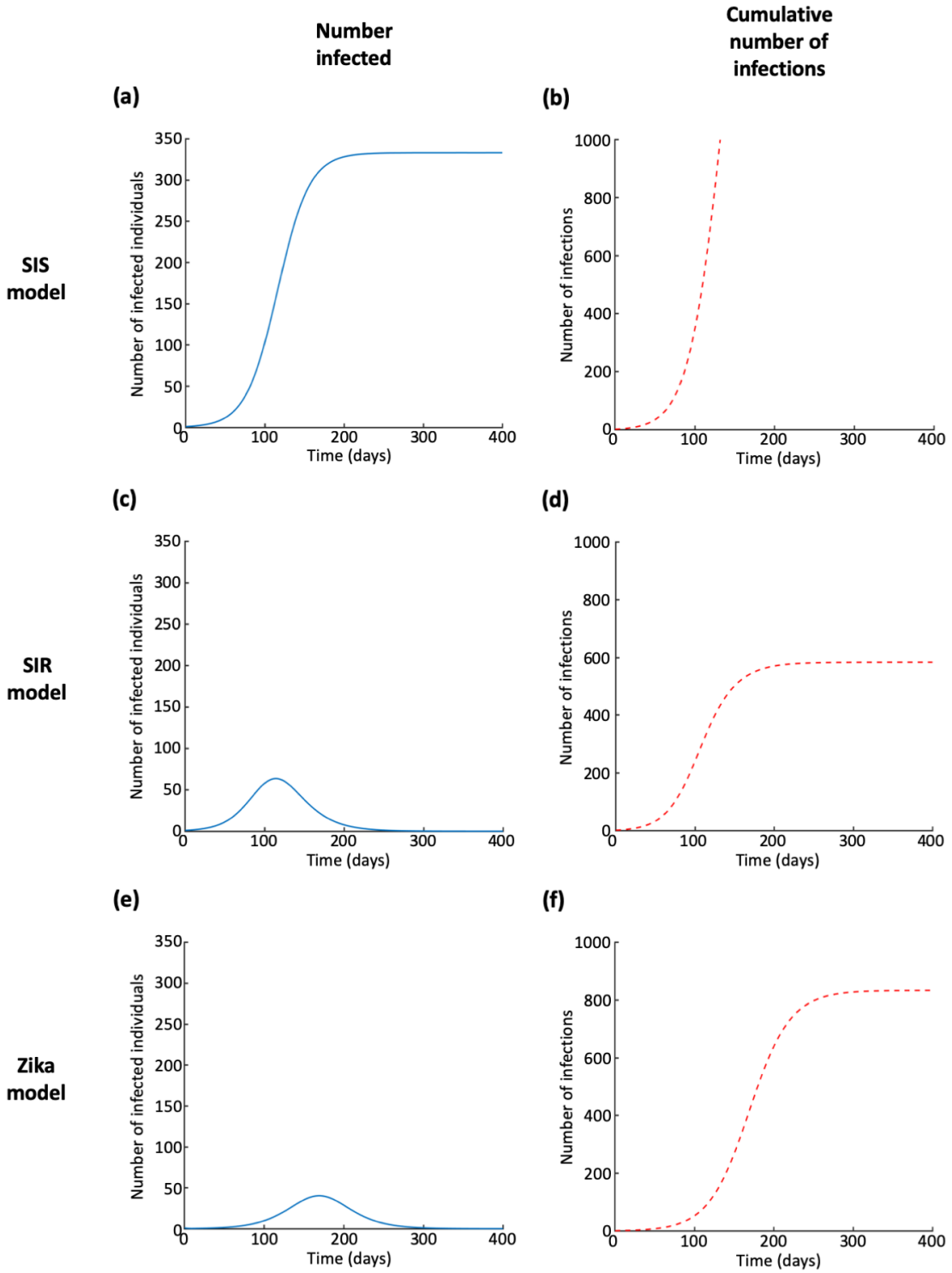

Figure 2. Numerical solutions of the deterministic SIS, SIR and Zika virus transmission models when the basic reproduction number is $R_0 = 1.5$. (a) Temporal changes in the number of infected individuals according to the deterministic SIS model. (b) Temporal changes in the cumulative number of infections according to the deterministic SIS model. (c)-(d) Equivalent to a-b but for the deterministic SIR model. (e)-(f) Equivalent to a-b but for the deterministic Zika virus transmission model. Parameters for deterministic SIS and SIR models: $N = 1{,}000$, $\beta = 0.00015$ per day, $1/\mu = 10$ days. Parameters for deterministic Zika virus transmission model: $N = 1{,}000$, $N^V = 10{,}000$, $1/\alpha_V = 10.5$ days, $1/\alpha_H = 5.9$ days, $1/\mu = 5$ days, $1/\delta = 7.8$ days, $\beta = 6.15 \times 10^{-2}$ per day, $\beta_V = 0.22$ per day [77].

The probability of a major epidemic

We calculated the probability of a severe epidemic according to the stochastic SIS model under the "concurrent size" metric for a severe epidemic – i.e. an outbreak in which the number of individuals infected simultaneously is at least a pre-specified threshold number (*M*) at some time during the outbreak. In this case, as described in Methods, it is possible to calculate the probability of a severe epidemic analytically.

We show the probability of a severe epidemic for a range of values of the threshold *M* in Fig 3a. Of course, for R_0 much larger than one, the probability of a severe epidemic was approximated closely by the standard branching process estimate for the probability of a major epidemic for many values of the threshold, *M*. When, however, R_0 was close to one, the standard estimate corresponded to a single choice of *M* (see e.g. blue lines in Fig 3a, where the solid line is close to the corresponding dotted line in only one place, i.e. for a single value of *M*). The parameter regime in which R_0 is close to one is important in

many real epidemiological systems since the aim of pre-emptive control strategies is often to reduce R_0 below one (or, when an outbreak has started, to reduce the time-varying or effective reproduction number below one [82–85]).

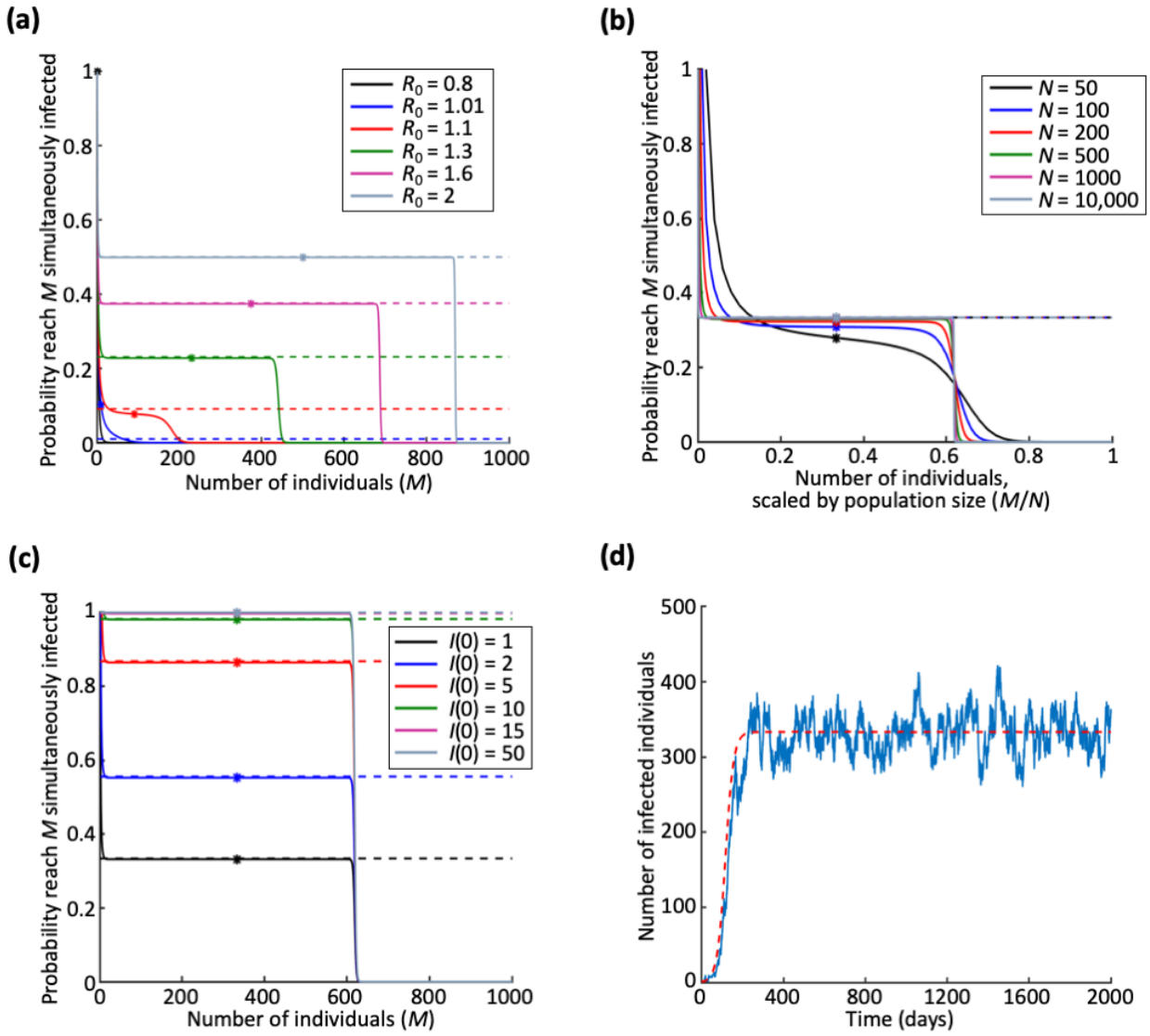

Figure 3. Probability of a severe epidemic under the SIS model, where a severe epidemic is defined as an outbreak in which at least M individuals are infected simultaneously at some time during the outbreak ("concurrent size" metric). (a) Dependence on R_0. Solid lines represent the probability of a severe epidemic (system of equations (7)), dotted lines represent the branching process estimate for the major epidemic

probability (equation (1)), and dots show the maximum number simultaneously infected in the analogous deterministic models (calculated analytically as shown in Text S1). R_0 is varied by changing the value of β. (b) Equivalent to a, but showing dependence on the population size, *N*. (c) Equivalent to a, but showing dependence on the initial number of infected individuals, *I*(0). (d). Single simulation of the stochastic SIS model (blue), and numerical solution of deterministic SIS model (red dotted). The value of *I* in the stochastic simulation will continue to fluctuate about the deterministic value until *I* reaches 0. Parameter values (except where stated): N = 1,000, R_0 = 1.5, $I(0)$ = 1 and $R(0)$ = 0 and the remainder of the population susceptible initially. In panel d, $\beta = 0.000015$ per day and $1/\mu = 10$ days.

In large host populations, the probability of a severe epidemic as a function of *M* took the form of a step function (Fig 3b). This suggested that, if the pathogen successfully invaded the population, then the number of infected individuals would definitely reach a specific maximum value which is determined by R_0. For example, for outbreaks with R_0 = 1.5, the pathogen will invade the population with probability 0.33 (i.e. the probability of a major epidemic), and, if this occurs, then around two-thirds of the population will be infected simultaneously at some time during the epidemic.

An approximation to the maximum value of *I* that will be reached in the stochastic SIS model in the large *N* case is given by

$$\max(I) \approx \max\left(2N\left(1-\frac{1}{R_0}\right), N\right).$$

The expression $2N\left(1-\frac{1}{R_0}\right)$ is twice the maximum value of *I* in the corresponding deterministic model. This approximation involves the assumption that, if the pathogen invades in the stochastic SIS model (so that a major epidemic occurs), *I* will fluctuate approximately symmetrically around the deterministic equilibrium value (Fig 3d). This is consistent with the previous finding that, in the stochastic SIS model, the quasi-stationary distribution can be approximated by the (symmetric) normal distribution when R_0 is larger than and not close to one [62]. It is likely, then, that by the time the pathogen dies out by reaching *I* = 0, the number of infected individuals will at some stage also have reached approximately double the value it fluctuated around too. In Fig 3b, when *N* = 10,000, this approximation gives a maximum proportion of the population simultaneously infected of 0.67, when the true value is 0.62.

We also note that, if the expression above for max(*I*) is reformulated to give the maximum proportion of the population that is simultaneously infected, the resulting expression is independent of the population size *N*. This can also be seen graphically – in Fig 3b, if the pathogen successfully invaded the population then the maximum proportion of hosts that were infected simultaneously was approximately independent of the size of the host population, so long as *N* was sufficiently large.

<u>Different epidemiological models</u>

We considered the probability that the maximum number of individuals infected simultaneously is at least a pre-specified threshold (i.e. a severe epidemic occurs, using the "concurrent size" metric) under the SIR and Zika virus transmission models. For the stochastic SIR model, we used an iterative method to calculate this probability as described in Methods.

For the stochastic Zika virus transmission model, we simulated the model in a population of N = 1,000 human hosts and N^V = 10,000 vectors using the Gillespie direct algorithm [81], using parameter values from Kucharski *et al.*, 2016 [77] – see caption of Fig 4. The value of R_0 was then varied in Fig 4b by altering the parameter β that governs the rate at which infected vectors infect susceptible human hosts.

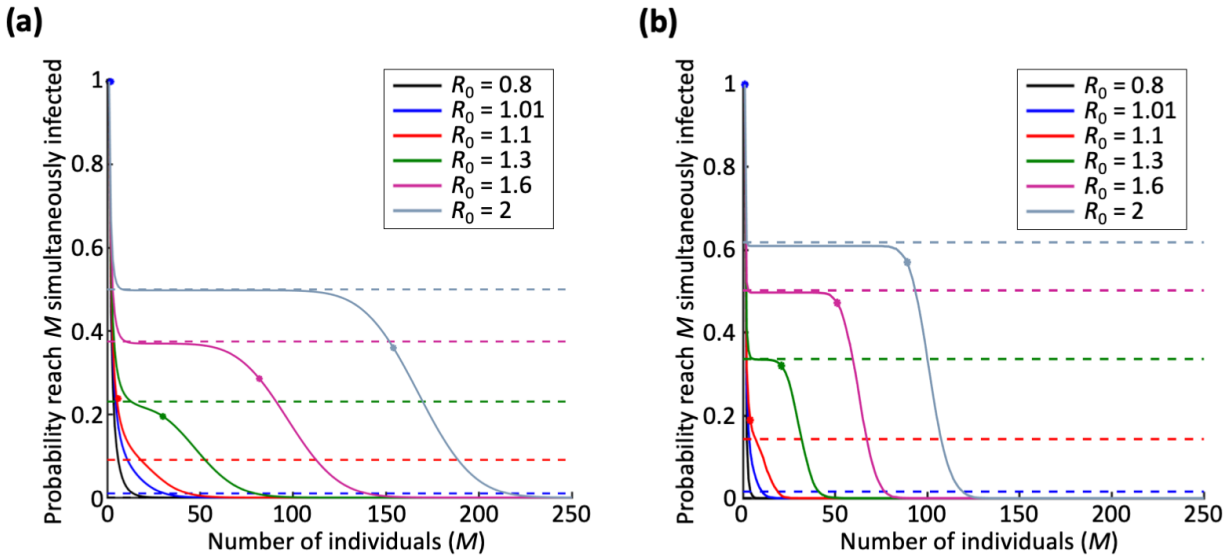

Figure 4. Probability of a severe epidemic under the SIR and Zika virus transmission models, where a severe epidemic is defined as an outbreak in which at least M individuals are infected simultaneously at some time during the outbreak ("concurrent size" metric). (a) SIR model. Solid lines represent the true probability of a severe epidemic calculated using the iterative method described in Methods, dotted lines

562 represent the branching process estimate of the probability of a major epidemic (equation (1)), and dots
563 show the maximum number simultaneously infected in the analogous deterministic model (calculated
564 analytically for the SIR model as shown in Text S1). R_0 is varied by changing the value of β. (b) Equivalent
565 to a, but for the Zika virus transmission model (where *M* refers to the number of simultaneously infected
566 hosts). For the Zika transmission model, the probability of a severe epidemic is calculated by simulation,
567 and the branching process estimate of the major epidemic probability is given by equation (5). The
568 maximum number simultaneously infected is found for the deterministic Zika virus transmission model by
569 numerically solving the model. For both models, $N = 1{,}000$. Other parameters for the Zika virus
570 transmission model: $N^V = 10{,}000$, $1/\alpha_V = 10.5$ days, $1/\alpha_H = 5.9$ days, $1/\mu = 5$ days, $1/\delta = 7.8$ days, $\beta_V =$
571 0.22 per day [77]. Initial conditions for both models comprise of a single infected host, with all other
572 individuals (for the Zika transmission model, hosts and vectors) susceptible.
573
574
575 Under the stochastic SIR and Zika virus transmission models, for R_0 larger than and not
576 close to one, the maximum number of simultaneously infected individuals whenever the
577 pathogen invaded the host population was typically smaller than under the SIS model (*cf*.
578 Fig 2). Nonetheless, we found qualitatively similar behaviour in these cases – the
579 probability of a severe epidemic was similar to the major epidemic probability
580 approximated using a branching process for a wide range of values of the severe
581 epidemic threshold when R_0 was high (Fig 4). However, even if that is the case, the
582 practically relevant value of the severe epidemic threshold (e.g. the number of available
583 hospital beds) may mean that the severe epidemic probability does not match the major
584 epidemic probability. For example, if $R_0 = 2$ and 250 beds are available, for the SIR model
585 the probability of a severe epidemic under the "concurrent size" definition is 0 (solid grey

line in Fig 4a), yet the branching process estimate for the probability of a major epidemic is 0.5 (dotted grey line in Fig 4a).

Alternative definitions of a severe epidemic

For the stochastic SIS model, we then calculated the probability of a severe epidemic using different metrics to define a severe epidemic – specifically, outbreaks in which there are at least *F* infection events (the "total infections" metric – Fig 5a) or outbreaks that persist for at least *T* days (the "duration" metric – Fig 5b).

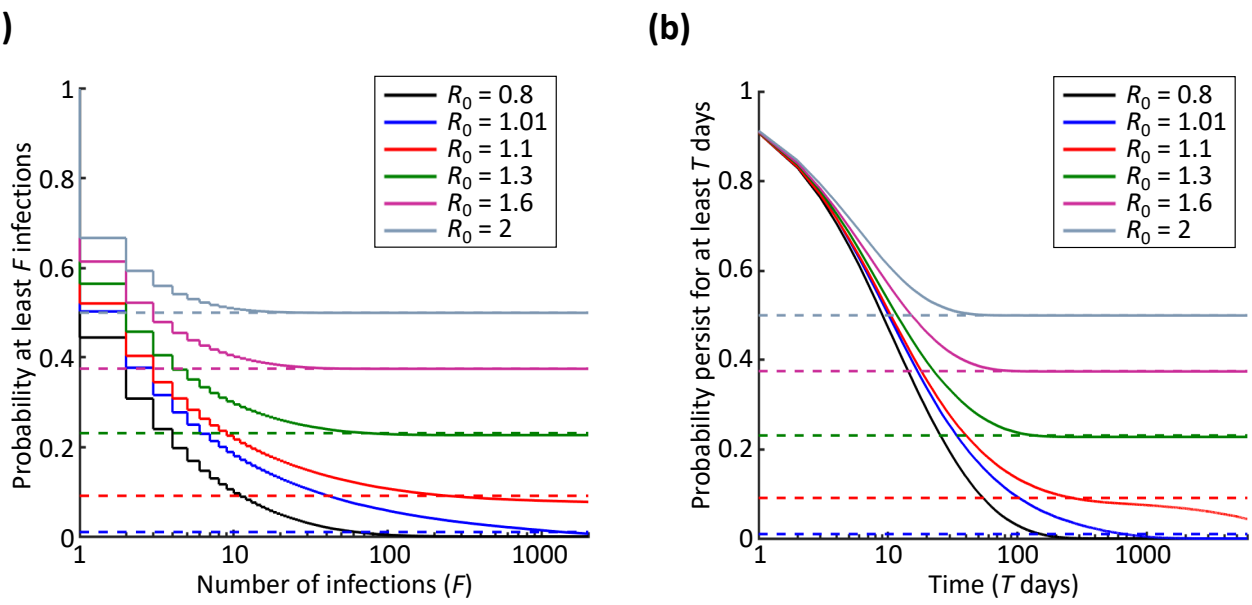

Figure 5. Probability of a severe epidemic under the SIS model, for different definitions of a severe epidemic. (a) A severe epidemic is defined as an outbreak in which at least *F* infections occur ("total infections" metric). (b) A severe epidemic is defined as an outbreak that persists for at least *T* days ("duration" metric). Solid lines represent the probability of a severe epidemic assessed via simulation of the stochastic model, and dotted lines represent the branching process estimate for the probability of a major epidemic (equation (1)). The x-axis is shown on a log-scale, for *F* between 1 and 2000 (panel a) and *T* between 1 and 6000 (panel

b). The step function in panel a reflects the fact that the total number of infections can only take discrete values. Results of the deterministic model are not included in the figure, since under the deterministic SIS model epidemics persist indefinitely and generate an infinite number of infections whenever $R_0 > 1$. Parameter values: N = 1,000, $I(0)$ = 1 and $R(0)$ = 0 and the remainder of the population susceptible initially. In both panels, R_0 is varied by changing the value of β. In panel b, $1/\mu = 10$ days.

In the stochastic SIS model, if the pathogen invaded the host population then it tended to persist for long periods. Consequently, the probability of a severe epidemic using the "total infections" or "duration" metrics is approximately equal to the major epidemic probability for a wide range of values of the severe epidemic thresholds (i.e. values of F or T) compared with under the "concurrent size" definition. However, even in these cases, for small or very large values of the severe epidemic thresholds, the probability of a severe epidemic does not match the major epidemic probability, particularly when R_0 is larger than but close to one (see e.g. red line in Fig 5b).

	SIS model	**SIR model**	**Zika host-vector model**
R_0 = 1.1	$21 \leq M \leq 54$ (Fig 3a)	$18 \leq M \leq 21$ (Fig 4a)	$6 \leq M \leq 7$ (Fig 4b)
R_0 = 1.6	$8 \leq M \leq 681$ (Fig 3a)	$8 \leq M \leq 55$ (Fig 4a)	$3 \leq M \leq 47$ (Fig 4b)
R_0 = 2	$6 \leq M \leq 866$ (Fig 3a)	$7 \leq M \leq 117$ (Fig 4a)	$2 \leq M \leq 83$ (Fig 4b)

Table 1. Effect of underlying epidemiology. For which values of the threshold (number of hosts simultaneously infected, M) in the "concurrent size" definition of a severe epidemic does the probability of a severe epidemic match the branching process estimate for the major epidemic probability? Range of values here are those for which the probability of a severe epidemic is within 0.01 of the probability of a

major epidemic. Results are summarised for the SIS, SIR and Zika virus models, for the parameter values shown in the relevant figure captions.

	Concurrent size	**Total infections**	**Duration**
R_0 = 1.1	$21 \leq M \leq 54$ (Fig 3a)	$138 \leq F \leq 878$ (Fig 5a)	$183 \leq T \leq 458$ (Fig 5b)
R_0 = 1.6	$8 \leq M \leq 681$ (Fig 3a)	$F \geq 18$ (Fig 5a)	$T \geq 51$ (Fig 5b)
R_0 = 2	$6 \leq M \leq 866$ (Fig 3a)	$F \geq 10$ (Fig 5a)	$T \geq 34$ (Fig 5b)

Table 2. Effect of metric used to assess the probability of a severe epidemic. For which values of the thresholds in the practically relevant definitions of a severe epidemic do the probability of a severe epidemic match the branching process estimate for the major epidemic probability? Range of values here are those for which the probability of a severe epidemic is within 0.01 of the probability of a major epidemic. Results are summarised for the SIS model, for the parameter values shown in the relevant figure captions. “total infections” threshold values were tested up to a maximum of F = 2,000 (Fig 5a) and “duration” threshold values were tested up to T = 6,000 (Fig 5b).

4. DISCUSSION

Evaluating the risk from an invading pathogen early in a potential severe epidemic is vital for planning interventions and determining whether or not current control resources are sufficient. When a pathogen arrives in a new location, the probability that initial cases will lead to a major epidemic as opposed to fading out as a minor outbreak can be approximated by assuming that infections occur according to a branching process. This probability represents the risk that the outbreak will persist beyond the initial stochastic phase in which case numbers are low. For simple models such as the stochastic SIS and SIR models, this corresponds to the major epidemic probability in equation (1). This can be extended to estimate the major epidemic probability using models with additional

complexity, as we demonstrated by considering the case of host-vector transmission (see equations (4) and (5)).

However, the threat from an invading pathogen can also be assessed by estimating the probability that the outbreak will become “severe” and overwhelm available control resources. For example, the probability that the peak prevalence exceeds the capacity of treatment facilities (e.g. the number of hospital beds) can be estimated (Figs 3 and 4), as well as the probability that the outbreak exceeds a threshold in the total number of infections or has a long duration (Fig 5). In each case, the probability of a severe epidemic depends on the value of the threshold set (e.g. the exact number of hospital beds) to differentiate severe epidemics from other outbreaks, as well as the metric used to define a severe epidemic. This highlights the need to consider the precise definition of a “severe epidemic” carefully when assessing the risk from an invading pathogen.

When R_0 is much greater than one or when the population size is extremely large, however, the probability of a severe epidemic is constant for a range of values of the threshold differentiating severe epidemics from other outbreaks (see e.g. different values of M in Fig 3a,b). In these cases, the probability of a severe epidemic will often match between definitions (Table 2). This is perhaps unsurprising since, for example, an outbreak with a large total number of infections may well also have a large peak prevalence. The probability of a severe epidemic is then approximately equal to the major epidemic probability calculated in the standard way (Table 2). However, even in this case, the specific threshold of practical importance in the outbreak may correspond to a

672 different probability of a severe epidemic compared with the probability corresponding to
673 a wide range of threshold values. Consequently, if a policy-maker wishes to understand
674 the risk that an invading pathogen will overwhelm available resources for control, then
675 this question should be addressed directly by choosing the relevant metric and threshold
676 value carefully.
677
678 We considered practical definitions of a severe epidemic that were based on thresholds
679 such as the availability of treatment. A previous study defined severe epidemics according
680 to a threshold in the percentage of the population ever infected, and concluded that
681 epidemiological modellers should report the precise cutoff used to define such epidemics
682 in model simulations [86]. Their conclusion was based on the observation that different
683 thresholds in the percentage of hosts ever infected corresponded to wide variations in the
684 other outputs of model simulations including the number of dead hosts or the time of the
685 epidemic peak. We support this conclusion, and indeed some authors have reported the
686 definition of a severe epidemic they used clearly – for example, Keeling *et al*. [87]
687 differentiate between outbreaks in which less than one-third of the population becomes
688 infected and those in which larger numbers of individuals are ever infected. Similarly, a
689 recent study explored the risk of resurgence of COVID-19 when interventions are
690 removed in different countries [88]. In that study, resurgence was said to have occurred
691 when the number of individuals who are symptomatic infectious reached 100. However,
692 while these values were reported clearly, we emphasise that the precise type of threshold
693 and the value used should be chosen according to practical relevance in the particular
694 system under consideration.

Under the first definition of a severe epidemic that we considered (using the “concurrent size” metric), the probability of a severe epidemic was assessed in the context of the capacity for treatment by estimating whether or not a threshold number of simultaneously infected individuals was likely to be exceeded. This definition may be practically relevant in a range of scenarios. For example, real-time analysis of a diphtheria epidemic in Cox Bazar’s in Bangladesh involved assessing the number of hospital beds that were needed [46]. The number of beds required was approximated in that study by using a model to forecast disease incidence, assuming that 15% of reported cases would require treatment as inpatients with an average hospital stay of five days for each case. The number of hospital beds that were already available might have provided a practically relevant severe epidemic threshold. Another example for which this type of threshold might apply is citrus greening disease in Brazil, where a law was introduced stating that a citrus grove must be destroyed if 28% of trees in the grove were infected and symptomatic [89,90]. At a local spatial scale, severe epidemics could therefore be defined as outbreaks in which more than 28% of trees in a grove are infected and symptomatic concurrently. Other examples for which interventions were introduced as soon as a threshold in the number simultaneously infected was reached include the development of the National Chlamydia Screening Programme in the United Kingdom in 2002 in response to the large size of the infected population [91].

However, no single metric for determining whether or not an outbreak is a severe epidemic will be relevant in all situations. To illustrate the ambiguity to which this might

lead, we also considered two other definitions of a severe epidemic. In one of these (when the “total infections” metric was used), whether or not an outbreak was classified as a severe epidemic referred to the total number of infection events over the course of the outbreak, rather than the maximum number simultaneously infected. This might correspond to the total number of treatments required, which may be an important threshold if treatments have been stockpiled prior to the outbreak [49]. This definition might also be relevant if, for example, a policy-maker has to choose how to deploy resources between two different epidemics. If there are only sufficient resources to contain one outbreak, and both epidemics are equally controllable, then – in the absence of other considerations – it might be preferable to choose to contain the epidemic that is likely to generate more infections. In other real-world scenarios, alternative definitions might be appropriate. We also considered classifying severe epidemics as outbreaks that persist beyond a threshold length of time (using the “duration” metric). Different definitions of a severe epidemic might appear contradictory – for example, treatment can act to reduce the total number of infections yet increase the duration of the outbreak [92], making a severe epidemic less likely when the “total infections” metric is used but more likely when instead the “duration” metric is used.

Our intention here was to use very simple models to demonstrate the principle that different approaches for evaluating the threat from an invading pathogen can lead to very different assessments of risk. As described in the introduction, our research builds on a rich history of analyses that relate to the results obtained here. For example, for the stochastic SIS model, the time to extinction varies with R_0. For $R_0 < 1$, outbreaks will

741 certainly end quickly. When $R_0 > 1$, however, if R_0 is increased then the duration of the
742 epidemic also increases (exponentially with R_0 [62]). Exact analytic expressions for the
743 duration of a stochastic SIS epidemic have not been found, and so we assessed the
744 probability of a severe epidemic under the "duration" metric using model simulations (Fig
745 5). However, analytic approximations to the expected duration of a major epidemic exist
746 (e.g. [60,75]), and exploring the relationship between these approximations and the
747 probability of a severe epidemic for different values of the threshold under the "duration"
748 metric represents an interesting avenue for further investigation.
749
750 Another important extension of the research presented here is to explore the risk of a
751 severe epidemic for outbreaks that do not assume that the population is well-mixed. The
752 field of contact network epidemiology provides a framework in which the risk from an
753 invading pathogen can be explored, accounting for the topology of the underlying network
754 when making epidemiological predictions and planning public health measures. In that
755 context, the probability of a major epidemic can be derived, depending on the
756 transmissibility (*T*) of the pathogen rather than R_0 [93]. The value of *T* represents the
757 average probability that an infected host will transmit the pathogen to a susceptible
758 individual that they have contact with. Meyers *et al.* [94] investigated the dependence of
759 the major epidemic probability on the degree of the index case, as well as the major
760 epidemic probability for different numbers of initial cases, in the context of SARS. Those
761 authors, as well as Pourbohloul *et al.* [95], explored the effects of different interventions
762 that reduce the numbers of contacts on quantities including the probability of a major
763 epidemic. The framework underlying these models has also been extended to account

764 for the time evolution of outbreaks [76], which is imported if the evaluation of the risk from
765 an invading pathogen is to be linked to the extinction time of the outbreak (as in the
766 “duration” metric that we considered).
767
768 Although our approach could be extended for different types of models (such as network
769 models), compartmental models (such as the SIS and SIR models) are commonly used
770 for assessing outbreak risks. Accurate outbreak forecasting using a compartmental model
771 requires the model to be carefully matched to the epidemiology of the host-pathogen
772 system, potentially including within-host dynamics [96,97], asymptomatic transmission
773 [9,98,99] or spread between spatially distinct regions [29,100]. For certain definitions of a
774 severe epidemic, it may be necessary to include bed-ridden or convalescent hosts in the
775 model explicitly. For example, if the definition of a severe epidemic is linked to the
776 availability of beds in treatment centres (as may be the case when the “concurrent size”
777 metric is used), then infected individuals in treatment centres could be included in the
778 model explicitly (for an example in which we consider three different models of an Ebola
779 epidemic with different levels of complexity, see Text S3). More complex definitions of
780 severe epidemics could be used, potentially requiring multiple criteria to be satisfied for
781 an outbreak to be classified as a severe epidemic. In these more complicated scenarios,
782 analytic calculations of the probability of a severe epidemic might not be possible. Model
783 simulations can then be used to assess the risk from the invading pathogen, as we
784 showed for a host-vector model of Zika virus transmission (Fig 4b).
785

786 We note that practical use of the methods presented here at the start of an emerging
787 outbreak to assess the outbreak risk might require the parameters governing pathogen
788 transmission to be estimated directly from case notification data. A range of methods exist
789 for estimating reproduction numbers in real-time during outbreaks [83,84,101,102],
790 including those designed for estimation in the early stochastic phase [103,104]. Practical
791 use of the approaches we have developed might also require the wide range of
792 interventions that are introduced in outbreak response settings to be integrated into the
793 models explicitly. One way in which control can be included is to consider the effective
794 reproduction number when the pathogen arrives in the system instead of the basic
795 reproduction number, since the effective reproduction number accounts for interventions
796 [26,82–84,105–107]. In that situation, the results that we presented would be unchanged
797 (except that e.g. the lines in Fig 3a would correspond to different values of the effective
798 reproduction number rather than the basic reproduction number). Temporal changes in
799 interventions once an outbreak is underway have been approximated in epidemiological
800 models by assuming that the values of the parameters governing transmission change
801 during the outbreak, either by assuming that transmissibility changes at single timepoints
802 [108–110] or continuously as the outbreak progresses [111,112]. However, for detailed
803 descriptions of control to be included in estimates of the severe epidemic risk, more
804 complex interventions should be included in model simulations. Since models are often
805 used to test possible control strategies [7,12,20,45,82,113–115], this is a simple
806 extension of the results presented here.
807

In summary, we have shown that the precise definition of a severe epidemic should be considered carefully in future studies that aim to evaluate the risk when a pathogen first arrives in a host population. Only once a “severe epidemic” has been defined precisely for the specific outbreak and setting under consideration can the probability of a severe epidemic occurring be properly assessed. Providing an explicit demonstration of the consequences of not considering the practically relevant definition in evaluating the risk is the key contribution of this paper.

SUPPLEMENTARY MATERIAL

Text S1.

Text S2.

Figure S1. Schematic showing how the probability of a severe epidemic under the “concurrent size” definition can be deduced for the stochastic SIR model. The probability corresponding to each state (I,R) is deduced iteratively in the order 1,2,3... (red). The black arrows indicate which previous values are used to inform each new deduction.

Figure S2. Chart indicating the model-severe epidemic metric pair considered in the figures in the main text.

COMPETING INTERESTS

We have no competing interests.

AUTHORS' CONTRIBUTIONS

All authors designed the study. RNT undertook the research and wrote the first draft of the manuscript. All authors revised the manuscript and approved the final version for submission.

FUNDING

This research was funded by Christ Church, Oxford, via a Junior Research Fellowship (RNT) and the Bill and Melinda Gates Foundation (CAG).

REFERENCES

1. Daszak P, Cunningham AA, Hyatt AD. Emerging infectious diseases of wildlife — threats to biodiversity and human health. Science (80-). 2000;287: 443–450.
2. Anderson PK, Cunningham AA, Patel NG, Morales FJ, Epstein PR, Daszak P. Emerging infectious diseases of plants: pathogen pollution, climate change and agrotechnology drivers. Trends Ecol Evol. 2004;19: 535–544.
3. Taylor LH, Latham SM, Woolhouse MEJ. Risk factors for human disease emergence. Philos Trans R Soc B. 2001;356: 983–989.
4. Morens DM, Folkers GK, Fauci AS. The challenge of emergin and re-emerging infectious diseases. Nature. 2004;430: 242–249.

5. Jones KE, Patel NG, Levy MA, Storeygard A, Balk D, Gittleman JL, et al. Global trends in emerging infectious diseases. Nature. 2008; 990–994.

6. Fisher MC, Henk DA, Briggs CJ, Brownstein JS, Madoff LC, McCraw SL, et al. Emerging fungal threats to animal, plant and ecosystem health. Nature. 2012;484: 186–194.

7. Thompson RN, Brooks-Pollock E. Detection, forecasting and control of infectious disease epidemics: modelling outbreaks in humans, animals and plants. Philos Trans R Soc B. 2019;374: 20190038.

8. Craft ME, Beyer HL, Haydon DT. Estimating the probability of a major outbreak from the timing of early cases: an indeterminate problem ? PLoS One. 2013;8: e57878.

9. Thompson RN, Gilligan CA, Cunniffe NJ. Detecting presymptomatic infection is necessary to forecast major epidemics in the earliest stages of infectious disease outbreaks. PLoS Comput Biol. 2016;12: e1004836.

10. Ramsay ME. Measles: the legacy of low vaccine coverage. Arch Dis Child. 2013;98: 752–754.

11. Rohani P, Earn DJD, Grenfell BT. Impact of immunisation on pertussis transmission in England and Wales. Lancet. 2000;355: 285–286.

12. Ferguson NM, Donnelly CA, Anderson RM. The foot-and-mouth epidemic in Great Britain: Pattern of spread and impact of interventions. Science (80-). 2001;292: 1155–1160.

13. Kao RR. The role of mathematical modelling in the control of the 2001 FMD epidemic in the UK. Trends Microbiol. 2002;10: 279–286.

14. Keeling MJ. Models of foot-and-mouth disease. Proc R Soc B Biol Sci. 2005;272: 1195–1202.

15. Keeling MJ, Woolhouse MEJ, Shaw DJ, Matthews L, Chase-Topping M, Haydon DT, et al. Dynamics of the 2001 UK foot and mouth epidemic: Stochastic dispersal in a heterogeneous landscape. Science (80-). 2001;294: 813–818.

16. Pautasso M, Aas G, Queloz V, Holdenrieder O. European ash (Fraxinus excelsior) dieback – A conservation biology challenge. Biol Conserv. 2013;158: 37–49.

17. Cunniffe NJ, Laranjeira FF, Neri FM, DeSimone RE, Gilligan CA. Cost-effective control of plant disease when epidemiological knowledge is incomplete: Modelling Bahia bark scaling of citrus. PLoS Comput Biol. 2014;10: e1003753.

18. Chan M-S, Jeger MJ. An analytical model of plant virus disease dynamics with roguing and replanting. J Appl Ecol. 1994;31: 413–427.

19. Wu JT, Cowling BJ, Lau EHY, Ip DKM, Ho L-M, Tsang T, et al. School closure and mitigation of pandemic (H1N1) 2009, Hong Kong. Emerg Infect Dis. 2010;16: 538.

20. Thompson RN, Gilligan CA, Cunniffe NJ. Control fast or control smart: When should invading pathogens be controlled? PLoS Comput Biol. 2018;14.

21. Keeling MJ, Rohani P. Modeling infectious diseases in humans and animals. Princeton University Press; 2008.

22. Lloyd AL, Zhang J, Root AM. Stochasticity and heterogeneity in host-vector models. J R Soc Interface. 2007;4: 851–63.

23. Lloyd-Smith JO, Schreiber SJ, Kopp PE, Getz WM. Superspreading and the effect

900 of individual variation on disease emergence. Nature. 2005;438: 355–359.
901 24. Nishiura H, Cook AR, Cowling BJ. Assortativity and the probability of epidemic
902 extinction: a case study of pandemic influenza A (H1N1-2009). Interdiscip
903 Perspect Infect Dis. 2011; 194507.
904 25. Anderson D, Watson R. On the spread of a disease with gamma distributed latent
905 and infectious periods. Biometrika. 1980;67: 191–8.
906 26. Thompson RN, Jalava K, Obolski U. Sustained transmission of Ebola in new
907 locations: more likely than previously thought. Lancet Infect Dis. 2019;19: 1058–
908 59.
909 27. Swinton J. Extinction times and phase transitions for spatially structured closed
910 epidemics. Bull Math Biol. 1998;60: 215–230.
911 28. Merler S, Ajelli M, Fumanelli L, Parlamento S, Pastore y Piontti A, Dean NE, et al.
912 Containing Ebola at the source with ring vaccination. PLoS Negl Trop Dis.
913 2016;10: 1–11.
914 29. Thompson RN, Thompson C, Pelerman O, Gupta S, Obolski U. Increased
915 frequency of travel in the presence of cross-immunity may act to decrease the
916 chance of a global pandemic. Philos Trans R Soc B. 2019;374: 20180274.
917 30. Althaus CL, Low N, Musa EO, Shuaib F, Gsteiger. Ebola virus disease outbreak
918 in Nigeria: Transmission dynamics and rapid control. Epidemics. 2015;11: 80–4.
919 31. Abdullah N, Kelly JT, Graham SC, Birch J, Gonçalves-Carneiro D, Mitchell T, et
920 al. Structure-guided identification of a nonhuman morbillivirus with zoonotic
921 potential. J Virol. 2018;92: e01248-18.
922 32. Keeling MJ, Danon L. Mathematical modelling of infectious diseases. Br Med Bull.

923 2009;92: 33–42.
924 33. Keeling MJ, Ross J V. On methods for studying stochastic disease dynamics. J R
925 Soc Inferface. 2008;5: 171–181.
926 34. Kessler DA, Shnerb NM. Solution of an infection model near threshold. Phys Rev
927 E. 2007;76: 010901.
928 35. Park AW, Gubbins S, Gilligan CA. Invasion and persistence of plant parasites in a
929 spatially structured host population. Oikos. 2001;94: 162–174.
930 36. Thompson RN. Novel coronavirus outbreak in Wuhan, China, 2020: Intense
931 surveillance is vital for preventing sustained transmission in new locations. J Clin
932 Med. 2020;9: 498.
933 37. Hellewell J, Abbott S, Gimma A, Bosse NI, Jarvis CI, Russell TW, et al. Feasibility
934 of controlling COVID-19 outbreaks by isolation of cases and contacts. Lancet
935 Glob Heal. 2020;8: E488–E496.
936 38. Kendall DG. Deterministic and stochastic epidemics in closed populations. Proc
937 3rd Berkeley Symp Math Stat Prob. 1956;4: 149–165.
938 39. de Jong MCM. Mathematical modelling in veterinary epidemiology: why model
939 building is important. Prev Vet Med. 1995;25: 183–193.
940 40. Nasell I. 'The threshold concept in stochastic epidemic and endemic models' in D.
941 Mollison, Epidemic models: their structure and relation to data. Cambridge
942 University Press; 1995.
943 41. Britton T. Stochastic epidemic models: A survey. Math Biosci. 2010;225: 24–35.
944 42. Yan P. Distribution theory, stochastic processes and infectious disease modelling.
945 Math Epidemiol. 2008;1945: 229–293.

946 43. Ferguson NM, Laydon D, Nedjati-Gilani G, Imai N, Ainslie K, Baguelin M, et al.
947 Report 9: Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19
948 mortality and healthcare demand. 2020. Available: www.imperial.ac.uk/mrc-
949 global-infectious-disease-analysis/covid-19/report-9-impact-of-npis-on-covid-19/
950 44. Davies NG, Kucharski AJ, Eggo RM, Gimma A, CMMID COVID-19 working
951 group, Edmunds WJ. The effect of non-pharmaceutical interventions on COVID-
952 19 cases, deaths and demand for hospital services in the UK: a modelling study.
953 2020.
954 45. Thompson RN. Epidemiological models are important tools for guiding COVID-19
955 interventions. BMC Med. 2020;18: 152.
956 46. Finger F, Funk S, White K, Siddiqui MR, Edmunds WJ, Kucharski AJ. Real-time
957 analysis of the diphtheria outbreak in forcibly displaced Myanmar nationals in
958 Bangladesh. BMC Med. 2019;17: 58.
959 47. Kucharski AJ, Camacho A, Flasche S, Glover RE, Edmunds WJ, Funk S.
960 Measuring the impact of Ebola control measures in Sierra Leone. Proc Natl Acad
961 Sci U S A. 2015;112: 14366–14371.
962 48. Lamontagne F, Clément C, Fletcher T, Jacob ST, Fischer II WA, Fowler RA.
963 Doing today's work superbly well - treating Ebola with current tools. N Engl J Med.
964 2014;371: 1565–1566.
965 49. O'Dowd A. Government says it would stockpile Tamiflu again. Br Med J.
966 2014;349: g6386.
967 50. Butler D. Africa's latest Ebola outbreak comes to swift end. Nature. 2018; 2018–
968 2020.

51. Wood W. The History and Antiquities of Eyam. Whitefish, MT: Kessinger Publishing; 1859.
52. Scott S, Duncan CJ. Biology of plagues: evidence from historical populations. Cambridge University Press; 2001.
53. Whittles LK, Didelot X. Epidemiological analysis of the Eyam plague outbreak of 1665-1666. Proc R Soc B Biol Sci. 2016;283: 20160618.
54. Brauer F. “Compartmental models in epidemiology” in F. Brauer, P. van den Driessche, J. Wu - Mathematical Epidemiology. Springer; 2014.
55. Raggett GF. A stochastic model of the Eyam plague. J Appl Stat. 1982;9: 212–225.
56. Cullerne JP, French A, Poon D, Baxter A, Thompson RN. The pedagogical power of context: extending the epidemiology of Eyam. Phys Educ. 2019.
57. Artalejo JR, Economou A, Lopez-Herrero MJ. The maximum number of infected individuals in SIS epidemic models: Computational techniques and quasi-stationary distributions. J Comput Appl Math. 2010;233: 2563–2574.
58. Neuts MF, Li J. An algorithmic study of S-I-R stochastic epidemic models. Athens Conference on Applied Probability and Time Series Analysis: Applied probability, in honor of JM Gani Lecture Notes in Statistics. 1996. pp. 295–306.
59. House T, Ross J V., Sirl D. How big is an outbreak likely to be? Methods for epidemic final-size calculation. Proc R Soc A Math Phys Eng Sci. 2013;469: 20120436.
60. Kryscio RJ, Lefevre C. On the extinction of the S-I-S stochastic logistic model. J Appl Probab. 1989;26: 685–694.

61. Nasell I. On the quasi-stationary distribution of the stochastic logistic epidemic. Math Biosci. 1999;156: 21–40.
62. Nasell I. The quasi-stationary distribution of the closed endemic SIS model. Adv Appl Probab. 1996;28: 895–932.
63. Clancy D, Pollett PK. A note on quasistationary distributions of birth–death processes and the SIS logistic epidemic. J Appl Probab. 2003;40: 821–825.
64. Clancy D, Mendy ST. Approximating the quasi-stationary distribution of the SIS model for endemic infection. Methodol Comput Appl Probab. 2011;13: 603–618.
65. Clancy D. Approximating quasistationary distributions of birth–death processes. J Appl Probab. 2012;49: 1036–1051.
66. Nasell I. On the time to extinction in recurrent epidemics. J R Stat Soc B. 1999;61: 309–330.
67. Miller JC. A note on the derivation of epidemic final sizes. Bull Math Biol. 2012;74: 2125–2141.
68. Ma J, Earn DJD. Generality of the final size formula for an epidemic of a newly invading infectious disease. Bull Math Biol. 2006;68: 679–702.
69. Ball F. A unified approach to the distribution of total size and total area under the trajectory of infectives in epidemic models. Adv Appl Probab. 1986;18: 289–310.
70. van Herwaarden OA, Grasman J. Stochastic epidemics: major outbreaks and the duration of the endemic period. J Math Biol. 1995;33: 581–601.
71. Allen LJS, Allen EJ. A comparison of three different stochastic population models with regard to persistence time. Theor Popul Biol. 2003;64: 439–449.
72. Grasman J. Stochastic epidemics: The expected duration of the endemic period in

1015 higher dimensional models. Math Biosci. 1998;152: 13–27.
1016 73. Tuckwell HC, Williams RJ. Some properties of a simple stochastic epidemic
1017 model of SIR type. Math Biosci. 2007;208: 76–97.
1018 74. Barbour AD. The duration of the closed stochastic epidemic. Biometrika. 1975.
1019 75. Nasell I. Extinction and quasi-stationarity in the stochastic logistic SIS model.
1020 Springer; 2011.
1021 76. Miller JC, Davoudi B, Meza R, Slim AC, Pourbohloul B. Epidemics with general
1022 generation interval distributions. J Theor Biol. 2010;262: 107–115.
1023 77. Kucharski AJ, Funk S, Eggo RM, Mallet H-P, Edmunds WJ, Nilles EJ.
1024 Transmission dynamics of Zika virus in island populations: A modelling analysis of
1025 the 2013-14 French Polynesia outbreak. PLoS Negl Trop Dis. 2016;10:
1026 e0004726.
1027 78. van den Driessche P, Watmough J. Further notes on the basic reproduction
1028 number. In “Mathematical Epidemiology.” Springer; 2008.
1029 79. Heffernan JM, Smith RJ, Wahl LM. Perspectives on the basic reproductive ratio. J
1030 R Soc Inferface. 2005;2: 281–293.
1031 80. Bartlett MS. The relevance of stochastic models for large-scale epidemiological
1032 phenomena. Appl Stat. 1964;13: 2–8.
1033 81. Gillespie DT. Exact stochastic simulation of coupled chemical reactions. J Phys
1034 Chem. 1977;8: 2340–2361.
1035 82. Thompson RN, Hollingsworth TD, Isham V, Arribas-Bel D, Ashby B, Britton T, et
1036 al. Key questions for modelling COVID-19 exit strategies. Proc R Soc B Biol Sci.
1037 2020.

1038 83. Cori A, Ferguson NM, Fraser C, Cauchemez S. A new framework and software to
1039 estimate time-varying reproduction numbers during epidemics. Am J Epidemiol.
1040 2013;178: 1505–12.
1041 84. Thompson RN, Stockwin JE, Gaalen RD Van, Polonsky JA, Kamvar ZN, Demarsh
1042 PA, et al. Improved inference of time-varying reproduction numbers during
1043 infectious disease outbreaks. Epidemics. 2019;19: 100356.
1044 85. Abbott S, Hellewell J, Thompson RN, Sherratt K, Gibbs HP, Bosse NI, et al.
1045 Estimating the time-varying reproduction number of SARS-CoV-2 using national
1046 and subnational case counts. Wellcome Open Res. 2020;5: 112.
1047 86. Orbann C, Sattenspiel L, Miller E, Dimka J. Defining epidemics in computer
1048 simulation models: How do definitions influence conclusions? Epidemics.
1049 2017;19: 24–32. Available: http://dx.doi.org/10.1016/j.epidem.2016.12.001
1050 87. Keeling MJ. The implications of network structure for epidemic dynamics. Theor
1051 Popul Biol. 2005;67: 1–8.
1052 88. Hao X, Cheng S, Wu D, Wu T, Lin X, Wang C. Reconstruction of the full
1053 transmission dynamics of COVID-19 in Wuhan. Nature. 2020.
1054 89. Craig AP, Cunniffe NJ, Parry M, Laranjeira FF, Gilligan CA. Grower and regulator
1055 conflict in management of the citrus disease Huanglongbing in Brazil: A modelling
1056 study. J Appl Ecol. 2017;55: 1956–1965.
1057 90. Belasque Jr. J, Bassanezi RB, Yamamoto PT, Ayres AJ, Tachibana A, Violante
1058 AR, et al. Letter to the editor: Lessons from Huanglongbing management in Sao
1059 Paulo state, Brazil. J Plant Pathol. 2010;92: 285–302.
1060 91. LaMontagne DS, Fenton KA, Randall S, Anderson S, Carter P. Establishing the

1061 National Chlamydia Screening Programme in England: results from the first full
1062 year of screening. Sex Transm Infect. 2004;80: 335–341.
1063 92. Hollingsworth TD, Klinkenberg D, Heesterbeek H, Anderson RM. Mitigation
1064 strategies for pandemic influenza A: Balancing conflicting policy objectives. PLoS
1065 Comput Biol. 2011;7: e1001076.
1066 93. Newman MEJ. Spread of epidemic disease on networks. Phys Rev E. 2002;66:
1067 016128.
1068 94. Meyers LA, Pourbohloul B, Newman MEJ, Skowronski DM, Brunham RC.
1069 Network theory and SARS: Predicting outbreak diversity. J Theor Biol. 2005;232:
1070 71–81.
1071 95. Pourbohloul B, Meyers LA, Skowronski DM, Krajden M, Patrick DM, Brunham RC.
1072 Modeling control strategies of respiratory pathogens. Emerg Infect Dis. 2005;11:
1073 1249–1256.
1074 96. Almocera AES, Hernandez-Vargas EA. Coupling multiscale within-host dynamics
1075 and between-host transmission with recovery (SIR) dynamics. Math Biosci.
1076 2019;309: 34–41.
1077 97. Hart WS, Maini PK, Yates CA, Thompson RN. A theoretical framework for
1078 transitioning from patient-level to population-scale epidemiological dynamics:
1079 influenza A as a case study. J R Soc Interface. 2020;17: 20200230.
1080 98. Hyatt-Twynam SR, Parnell S, Stutt ROJH, Gottwald TR, Gilligan CA, Cunniffe NJ.
1081 Risk-based management of invading plant disease. New Phytol. 2017;214: 1317–
1082 1329.
1083 99. Thompson RN, Morgan OW, Jalava K. Rigorous surveillance is necessary for

1084 high confidence in end-of-outbreak declarations for Ebola and other infectious
1085 diseases. Philos Trans R Soc B. 2019;374: 20180431.
1086 100. Kraemer MUG, Faria NR, Reiner RC, Golding N, Nikolay B, Stasse S, et al.
1087 Spread of yellow fever virus outbreak in Angola and the Democratic Republic of
1088 the Congo 2015–16: a modelling study. Lancet Infect Dis. 2017;17.
1089 101. Wallinga J, Teunis P. Different epidemic curves for severe acute respiratory
1090 syndrome reveal similar impacts of control measures. Am J Epidemiol. 2004;160:
1091 509–516.
1092 102. Obadia T, Haneef R, Boëlle PY. The R0 package: A toolbox to estimate
1093 reproduction numbers for epidemic outbreaks. BMC Med Inform Decis Mak.
1094 2012;12: 1–9.
1095 103. Davoudi B, Miller JC, Meza R, Meyers LA, Earn DJD, Pourbohloul B. Early real-
1096 time estimation of the basic reproduction number of emerging infectious diseases.
1097 Phys Rev X. 2012;2: 031005.
1098 104. Kwok KO, Davoudi B, Riley S, Pourbohloul B. Early real-time estimation of the
1099 basic reproduction number of emerging or reemerging infectious diseases in a
1100 community with heterogeneous contact pattern: Using data from Hong Kong 2009
1101 H1N1 Pandemic Influenza as an illustrative example. PLoS One. 2015;10:
1102 e0137959.
1103 105. Nishiura H, Chowell G. The effective reproduction number as a prelude to
1104 statistical estimation of time-dependent epidemic trends. Mathematical and
1105 Statistical Estimation Approaches in Epidemiology. 2009. pp. 103–121.
1106 106. Delamater PL, Street EJ, Leslie TF, Yang YT, Jacobsen KH. Complexity of the

1107 basic reproduction number (R0). Emerg Infect Dis. 2019;25: 1–4.
1108 107. Anderson RM. The concept of herd immunity and the design of community-based
1109 immunization programmes. Vaccine. 1992;10: 928–935.
1110 108. Hart WS, Hochfilzer LFR, Cunniffe NJ, Lee H, Nishiura H, Thompson RN.
1111 Accurate forecasts of the effectiveness of interventions against Ebola may require
1112 models that account for variations in symptoms during infection. Epidemics. 2019.
1113 109. Meakin SR, Tildesley MJ, Davis E, Keeling MJ. A metapopulation model for the
1114 2018 Ebola outbreak in Equateur province in the Democratic Republic of the
1115 Congo. bioRxiv. 2018; 1–30.
1116 110. Flaxman S, Mishra S, Gandy A, Unwin HJT, Mellan TA, Coupland H, et al.
1117 Estimating the effects of non-pharmaceutical interventions on COVID-19 in
1118 Europe. Nature. 2020.
1119 111. Chowell G, Nishiura H. Characterizing the transmission dynamics and control of
1120 Ebola virus disease. PLoS Biol. 2015;13: 1002057.
1121 112. Chowell G, Hengartner NW, Castillo-Chavez C, Fenimore PW, Hyman JM. The
1122 basic reproductive number of Ebola and the effects of public health measures:
1123 The cases of Congo and Uganda. J Theor Biol. 2004;229: 119–126.
1124 113. Thompson RN, Hart WS. Effect of confusing symptoms and infectiousness on
1125 forecasting and control of Ebola outbreaks. Clin Infect Dis. 2018;67: 1472–1474.
1126 114. Thompson RN, Cobb RC, Gilligan CA, Cunniffe NJ. Management of invading
1127 pathogens should be informed by epidemiology rather than administrative
1128 boundaries. Ecol Modell. 2016;324: 28–32.
1129 115. Cunniffe NJ, Cobb RC, Meentemeyer RK, Rizzo DM, Gilligan CA. Modeling when,

where, and how to manage a forest epidemic, motivated by sudden oak death in California. Proc Natl Acad Sci USA. 2016;113: 5640–5645.