When does a minor outbreak become a major epidemic? Linking the risk from invading pathogens to practical definitions of a major epidemic

AUTHORS

R.N. Thompson[1,2,3,4,*], C.A. Gilligan[4], N.J. Cunniffe[4]

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

AFFILIATIONS

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

[2]Department of Zoology, University of Oxford, Oxford, UK

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

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

ABSTRACT

Forecasting whether or not initial reports of disease will be followed by a major epidemic is an important component of disease management, guiding optimal deployment of limited resources for surveillance and control. For example, the probability that undetected cases arriving in different countries would lead to a major epidemic was estimated during the 2014-16 Ebola epidemic in West Africa, and in the ongoing epidemic in DR Congo. Standard epidemic risk estimates involve assuming that infections occur according to a branching process. Surprisingly, however, these calculations are carried out without the factors differentiating major epidemics from minor outbreaks being defined precisely. We assess implications of this lack of explicitness by considering three practically relevant

potential definitions of a major epidemic; namely, an outbreak in which: i) a large number of hosts are infected simultaneously; ii) a large number of infections occur in total; and iii) disease remains in the population for a long period. We show that the major epidemic probability under these definitions can be similar to, or different from, the standard estimate. This holds in a range of systems, highlighting that careful consideration of what constitutes a “major epidemic” in each outbreak is vital for accurate quantification of risk.

KEYWORDS

mathematical modelling; infectious disease epidemiology; major epidemic; forecasting

SHORT TITLE

Assessing the practically relevant risk of a major epidemic

1. INTRODUCTION

Infectious disease epidemics in populations of humans, animals and plants represent a recurring risk worldwide [1–7]. An important question for policy-makers towards the start of an outbreak is whether initial cases will lead on to a major epidemic, or whether the pathogen will rapidly die out instead [8,9]. An important practical consequence is that, if an outbreak is likely to simply 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 well 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 can be derived by assuming that infections occur according to a branching process (see Methods). In some models in which the infected class is sub-divided into different compartments, the value of $I(0)$ must instead be interpreted as the total number of individuals infected at the current time. For example, for the Susceptible-Exposed-Infectious-Removed model, the exponent in equation (1) would in fact become $E(0) + I(0)$ [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 also 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 also considered in the context of flare-ups in new locations for the ongoing Ebola epidemic in the Democratic Republic of the Congo [26].

However, the probability of a major epidemic in equation (1) is derived without specifying a precise definition of a major epidemic. 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 [36–38]). 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. This provides a natural definition for “minor outbreaks” and “major epidemics”, motivated largely by the elegance of the mathematical modelling analysis. However, for practical application it would often be more appropriate for the notion of a major epidemic to instead be grounded in consequences for control of disease, depending on the specific system and outbreak under consideration.

Here we consider three potential definitions of a major epidemic that might be practically relevant in different outbreak scenarios. Specifically, these are:

93 **Concurrent size.** An outbreak in which the number of individuals simultaneously
94 infected exceeds the capacity for treatment.
95 **Total infections**. An outbreak in which the total number of infections exceeds the
96 number of available treatments.
97 **Duration**. An outbreak that is not contained quickly and therefore persists for an
98 unacceptably long period.
99
100 We investigate under which circumstances the probability of a major epidemic according
101 to each of these definitions coincides with the branching process estimate. In our
102 analyses, as examples we consider three stochastic epidemiological models that are
103 representative of different host responses to infection and capture different routes of
104 transmission. Specifically, we consider the SIS model, the SIR model, and a host-vector
105 model parameterised for Zika virus transmission. For the SIS and SIR models, the
106 standard branching process estimate corresponds to equation (1), and in the case of Zika
107 virus the standard branching process estimate is an adapted version of equation (1) that
108 accounts for host-vector-host transmission (see Methods).
109
110 To motivate our analyses, we note that the three definitions of a major epidemic above
111 might each be applicable in different scenarios. For example, it might be natural to
112 assume that, if the number of individuals infected at any time always remains below the
113 capacity for treatment, then the outbreak is minor since medical care is available for all
114 sick individuals. The threshold capacity might derive from the number of available hospital

beds [39] or the availability of care workers [40]. This motivates the “Concurrent size” definition above.

However, a definition based on numbers of simultaneously infected hosts cannot be applied ubiquitously. 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 over the course of 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 [41]. Another possible definition of a major epidemic is therefore an outbreak in which the total number of infections exceeds a critical value (the “Total infections” definition, above). This critical value might be set by the stock of available treatments for use during the outbreak.

Finally, we consider a third definition of a major epidemic (the “Duration” definition). Under this definition, a major epidemic is an outbreak that persists for an unacceptably long time. 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 be considered minor. For example, the first Ebola outbreak in the Democratic Republic of the Congo in 2018 resulted in 53 cases, but was not considered a major epidemic due to its fast containment [42], leading to commendation of the success of public health

measures. Consequently, an outbreak might only be classified as a major epidemic if it persists for a threshold length of time.

It is important to recognise that these definitions do not always coincide. In 1665-66, plague affected the village of Eyam in the UK, which famously isolated itself via a self-imposed quarantine [43,44]. 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 ambiguity particularly regarding the size of the at-risk population [45]). However, model fits suggest that a maximum of only around 30 people were ever infected simultaneously [46–48]. As a result, this epidemic might have been classified as a major epidemic according to the “Total infections” and “Duration” definitions, yet not under the “Concurrent size” definition, depending on the precise values of the thresholds set in each case. This highlights the need to define a major epidemic carefully, since an individual outbreak may or may not qualify as a major epidemic, depending on the definition used.

We will show that the probability of a major epidemic depends on precisely how a major epidemic is defined. The probability of a major epidemic under practical definitions may or may not match the branching process estimate. The definition to use should therefore be considered carefully before the major epidemic risk is estimated at the beginning of an emerging outbreak. Only once the notion of a major epidemic has been formally defined – based on criteria of practical relevance – can this risk be properly assessed.

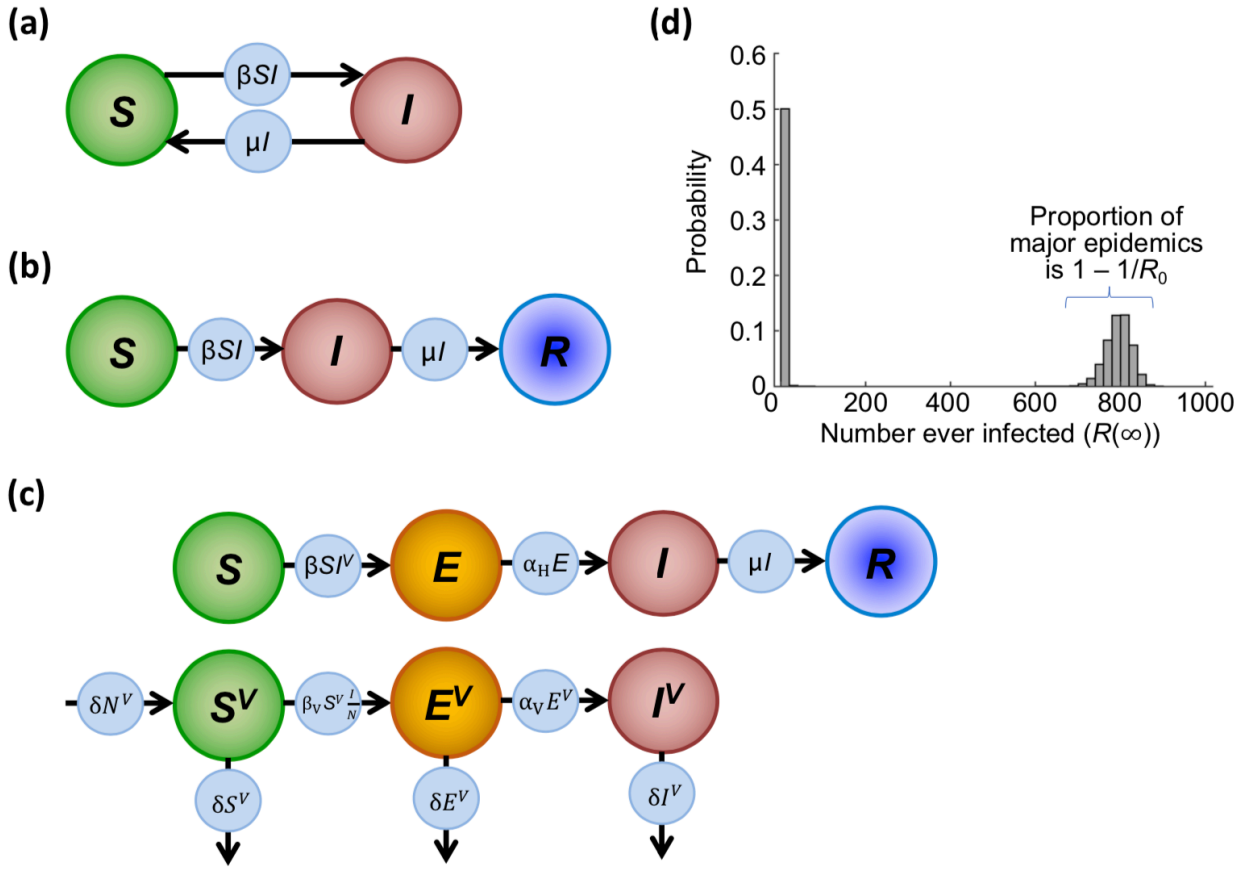

Figure 1. Schematic diagrams illustrating the population structures for the different models considered, and an example distribution of final sizes for the stochastic SIR model. (a) The SIS model. (b) The SIR model. (c) The model of Zika virus transmission. (d) Distribution of final sizes in the stochastic 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 calculate the probability of a major epidemic under the "Concurrent size" definition for the stochastic SIS model, the

SIR model and the Zika host-vector model. Our final two analyses focus on the stochastic SIS model. We calculate the probability of a major epidemic under other definitions of a major epidemic (the “Total infections” and “Duration” definitions). In each of these five analyses, we compare the probability of a major epidemic for the particular model-major epidemic definition pairing considered to the commonly used branching process approximation for the probability of a major epidemic which does not require a major epidemic to be defined formally.

Here, we describe the epidemiological models that we use, calculation of the branching process approximation to the probability of a major epidemic for each model, and how the probability of a major epidemic can be calculated under the “Concurrent size” definition for each of the models considered. We then explain how the probability of a major epidemic under the other practically relevant definitions of a major epidemic 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 rate of infection 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 rate of removal. 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 the host-vector model of Kucharski *et al*. [49], 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. The deterministic version of this model is given by

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

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

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

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

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

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

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

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

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

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 $R_0^{VH} = \frac{1}{\delta}\beta N$ is the expected number of humans infected by a single infectious vector.

The basic reproduction number is given by

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

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 [50,51]. We note that in some studies, e.g. [49], the square root is omitted from the definition of R_0. In contrast to the expression calculated by Kucharski *et al*. [49], to facilitate simulation of the stochastic model we also 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 probability of a major epidemic when a pathogen first arrives in a host population [8,9,21,27,29–35] 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 probability of a major epidemic is given by

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

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 probability of a major epidemic can be derived for more complex systems by considering the stochastic version of the model of Zika virus 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 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

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

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

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 major epidemic (“Concurrent size” definition)</u>

As described in the introduction, we first define a major epidemic to be an outbreak in which the maximum number of simultaneously infected individuals 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.

Deterministic SIS and SIR models

In the deterministic SIS model, the maximum number of simultaneously infected individuals over the course of the outbreak is given by

$$I_{\max} = \begin{cases} N\left(1 - \frac{1}{R_0}\right) \text{ for } I(0) \leq N\left(1 - \frac{1}{R_0}\right), \\ I(0) \text{ for } I(0) > N\left(1 - \frac{1}{R_0}\right), \end{cases}$$

as shown in Text S1.

In the deterministic SIR model, on the other hand,

$$I_{\max} = \begin{cases} I^* \text{ for } I(0) \leq I^*, \\ I(0) \text{ for } I(0) > I^*, \end{cases}$$

where $I^* = -\frac{N}{R_0} + \frac{N}{R_0}\ln\left(\frac{N}{S(0)R_0}\right) + S(0) + I(0)$, and again this is shown in Text S1.

The probability of a major epidemic according to the “Concurrent size” definition under the deterministic SIS and SIR models is then

$$\text{Prob(major epidemic)} = \begin{cases} 1 \text{ for } M \leq I_{\max}, \\ 0 \text{ for } M > I_{\max}, \end{cases}$$

where I_{max} is given by the corresponding expressions above depending which model is used.

Stochastic SIS model

Under the stochastic SIS model, however, calculating the probability of a major epidemic is more challenging [52]. Nevertheless, we can calculate this value analytically, which is advantageous since approximating this quantity using model simulations can be time consuming given that outbreaks under the SIS model can persist for indefinitely long periods.

We denote the probability of no major epidemic starting with *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 a major epidemic certainly does not occur if there are initially no infected individuals, and certainly does 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 major epidemic is then given by

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

Stochastic SIR model

Under the stochastic SIR model, the probability of a major epidemic according to the "Concurrent size" definition starting from any state (*I*,*R*) is calculated using an iterative approach. Denoting the probability of a major epidemic starting from state (*I*,*R*) by $p_{I,R}$, then conditioning on what happens next 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 $p_{M,R} = 1$. To do this, the probability of a major epidemic 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 major epidemic under the “Concurrent size” definition is approximated by simulating the model 10,000 times using the Gillespie direct method [53], and then calculating the proportion of simulations in which the number of infected human hosts exceeds *M* at any point during the simulation.

Probability of a major epidemic (“Total infections” and “Duration” definitions)

We also consider the probability of a major epidemic according to the deterministic and stochastic SIS model for the “Total infections” and “Duration” definitions of a major epidemic. Under the “Total infections” definition, a major epidemic is assumed to be an outbreak in which at least *F* infections occur over the course of the outbreak. Under the “Duration” definition, a major epidemic is defined to be an outbreak that persists for at least *T* days.

In the deterministic SIS model, whenever $R_0 > 1$, the outbreak persists indefinitely with an infinite number of infection events. As a result, any outbreak is a major epidemic under the “Total infections” and “Duration” definitions.

In contrast, in any simulation of the stochastic SIS model, the number of infected individuals will always reach zero, even if this takes a long time. As a result, the probability of a major epidemic under the “Total infections” and “Duration” definitions is not simply one or zero depending on the value of R_0. We approximate the probability of a major epidemic under these definitions by simulating the model 10,000 times using the Gillespie

direct method [53], and recording the proportion of simulations in which there are at least *F* infections or that have a duration of at least *T* days.

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). In a deterministic setting, the SIS model predicts the largest number of individuals simultaneously infected 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 the probability of a major epidemic soon after the pathogen enters the system. For any definition of "major epidemic", according to a deterministic model the corresponding probability is either zero or one depending on the values of model parameters. 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 major epidemic for the stochastic SIS model under the "Concurrent size" definition. We then consider different epidemiological models, as well as different definitions of a major epidemic. In each case, the probability of a major epidemic for the particular epidemiological model-definition of a major epidemic pair under consideration is compared to the branching process approximation to the probability of a major epidemic for that model. Results are shown in Figs 3-5, as well as summarised in Tables 1 and 2.

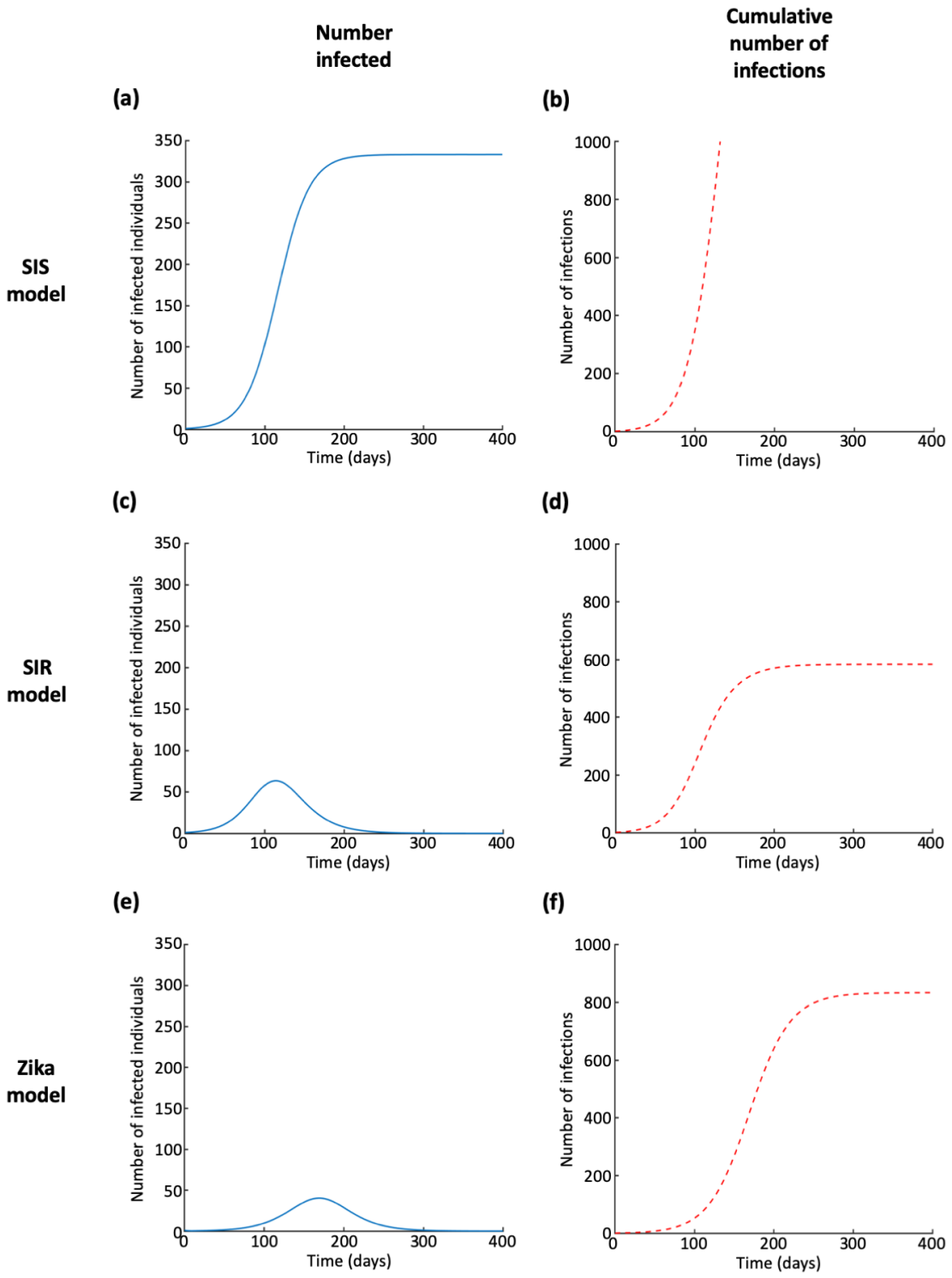

Figure 2. Numerical solutions of the deterministic SIS, SIR and Zika virus transmission models when the basic reproduction number is $R_0 = 1.5$. (a) Number of infected individuals through time according to the deterministic SIS model. (b) Cumulative number of infections through time according to the 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, $\beta = 6.15 \times 10^{-5}$ per day, $\beta_V = 0.22$ per day [49].

The probability of a major epidemic

We calculated the probability of a major epidemic according to the stochastic SIS model under the "Concurrent size" definition of a major epidemic – i.e. an outbreak in which a threshold number of simultaneously infected individuals is exceeded. In this case, as described in Methods, it is possible to calculate the probability of a major epidemic analytically.

We show the probability of a major epidemic for a range of values of the major epidemic threshold, *M*, in Fig 3a. For R_0 much larger than one, we found that the probability of a major epidemic was approximated closely by the standard branching process estimate for many values of the major epidemic 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 control strategies is usually to reduce the reproduction number below one [54,55].

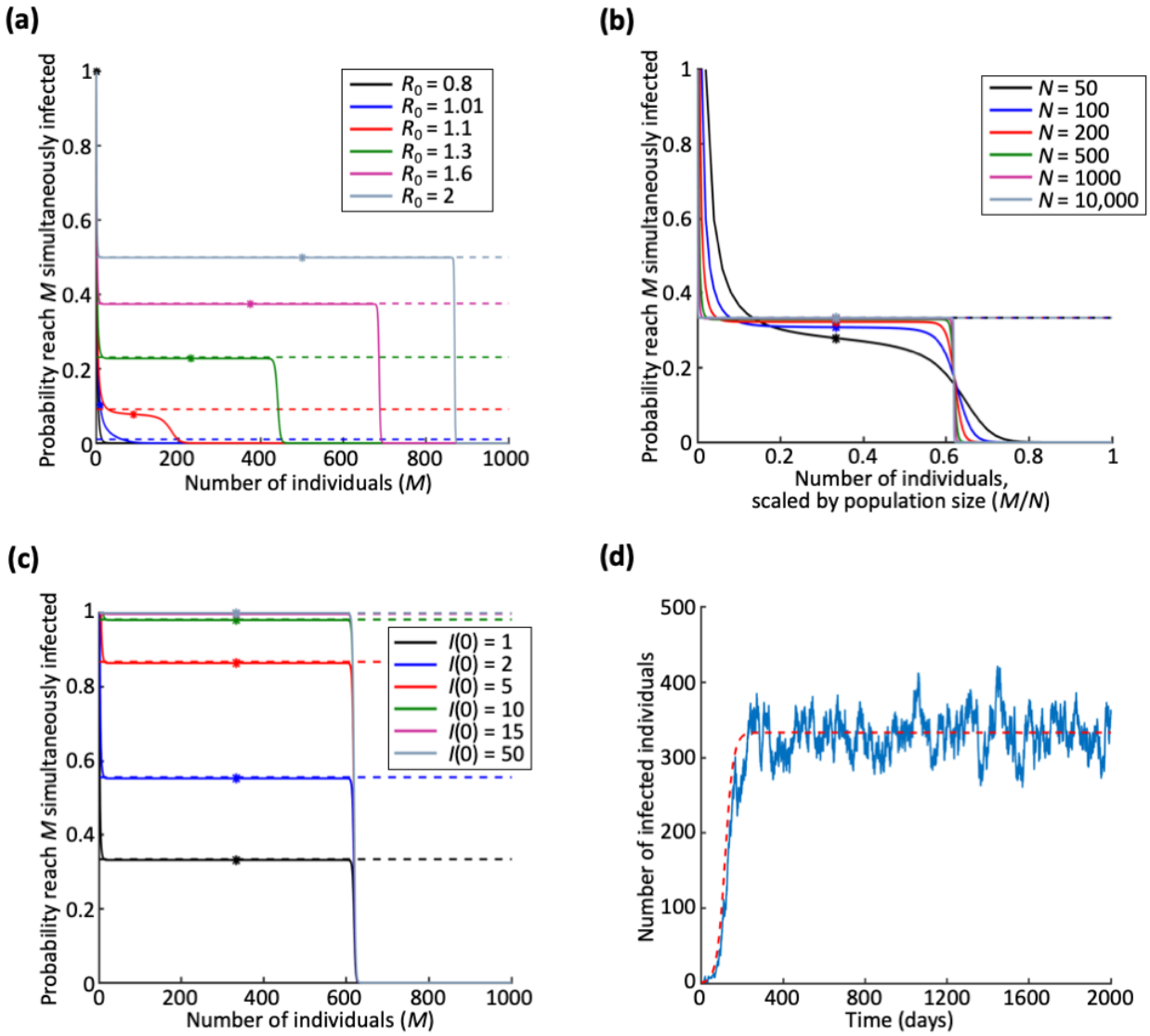

Figure 3. Probability of a major epidemic under the SIS model, where a major epidemic is defined as an outbreak in which *M* simultaneously infected individuals is exceeded ("Concurrent size" definition). (a) Dependence on R_0. Solid lines represent the true probability of a major epidemic (system of equations (7)), dotted lines represent the branching process estimate (equation (1)), and dots show the maximum number simultaneously infected in the analogous deterministic models (for values of *M* below this, the probability of a major epidemic in the relevant deterministic model is 1). 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 major 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, 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, *I* will fluctuate approximately symmetrically around the deterministic equilibrium value (Fig 3d). It is likely, then, that by the time the pathogen

509 dies out by reaching $I = 0$, the number of infected individuals will at some stage also have
510 reached approximately double the value it fluctuated around too. In Fig 3b, when N =
511 10,000, this approximation gives a maximum proportion of the population simultaneously
512 infected of 0.67, when the true value is 0.62.
513
514 We also note that, if the expression above for max(I) is reformulated to give the maximum
515 proportion of the population that is simultaneously infected, the resulting expression is
516 independent of the population size N. This can also be seen graphically – in Fig 3b, if the
517 pathogen successfully invaded the population then the maximum proportion of hosts that
518 were simultaneously infected was approximately independent of the size of the host
519 population, so long as N was sufficiently large.
520
521
522 Different epidemiological models
523
524 We considered the probability of a major epidemic ("Concurrent size" defintion) under the
525 SIR and Zika virus transmission models. For the stochastic SIR model, we used an
526 iterative method to calculate the probability of a major epidemic as described in Methods.
527
528 For the stochastic Zika virus transmission model, we simulated the model in a population
529 of N = 1,000 human hosts and N^V = 10,000 vectors using the Gillespie direct algorithm
530 [53], using parameter values from Kucharski *et al.*, 2016 [49] – 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.

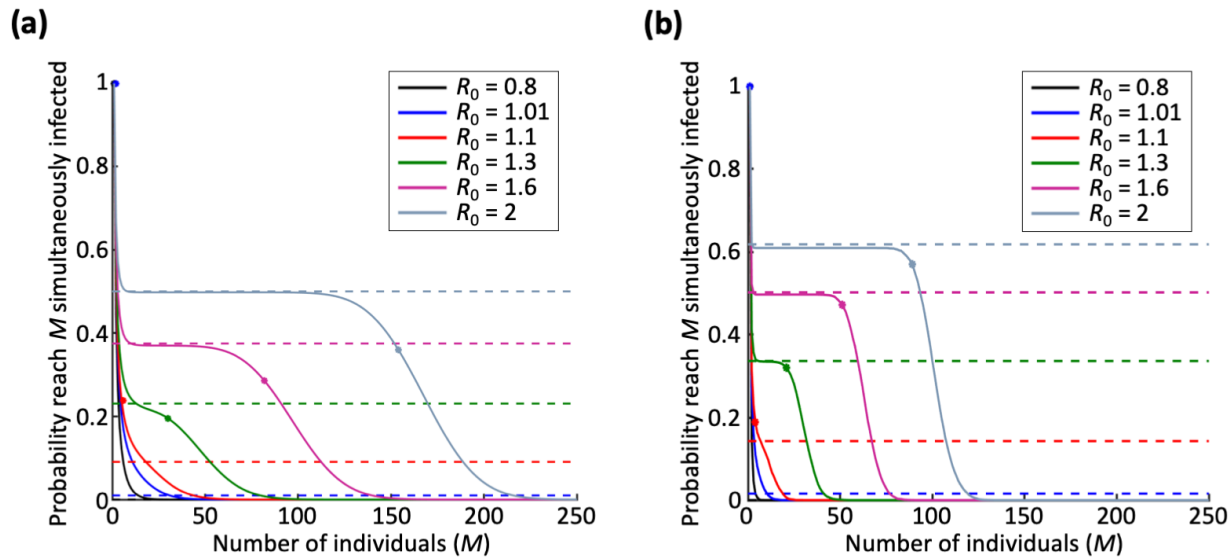

Figure 4. Probability of a major epidemic under the "Concurrent size" definition of a major epidemic, for the SIR and Zika virus transmission models. (a) SIR model. Solid lines represent the true probability of a major epidemic calculated using the iterative method described in Methods, dotted lines represent the branching process estimate (equation (1)), and dots show the maximum number simultaneously infected in the analogous deterministic models (for values of *M* below this, the probability of a major epidemic in the relevant deterministic model is 1). R_0 is varied by changing the value of β. (b) Equivalent to a, but for the Zika virus transmission model (where *M* refers to the number of simultaneously infected hosts). For the Zika transmission model, the true probability of a major epidemic is calculated by simulation, and the branching process estimate is given by equation (5). For both models, $N = 1{,}000$. Other parameters for the Zika virus 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, $\beta_V = 0.22$ per day [49]. Initial conditions for both models comprise of a single infected host, with all other individuals (for the Zika transmission model, hosts and vectors) susceptible.

549 Under the stochastic SIR and Zika models, for R_0 larger than and not close to one, the
550 maximum number of simultaneously infected individuals whenever the pathogen invaded
551 the host population was typically smaller than under the SIS model (*cf*. Fig 2).
552 Nonetheless, we found qualitatively similar behaviour in these cases – the probability of
553 a major epidemic approximated using a branching process corresponded to a wide range
554 of values of the major epidemic threshold when R_0 was high (Fig 4). However, even if that
555 is the case, the practically relevant value of the major epidemic threshold (e.g. the number
556 of available hospital beds) may not give a probability of a major epidemic that matches
557 the branching process estimate. For example, if R_0 = 2 and 250 beds are available, for
558 the SIR model the probability of a major epidemic under the "Concurrent size" definition
559 is 0 (solid grey line in Fig 4a), yet the branching process estimate for the probability of a
560 major epidemic is 0.5 (dotted grey line in Fig 4a).
561
562 <u>Alternative definitions of a major epidemic</u>
563
564 For the stochastic SIS model, we then calculated the probability of a major epidemic for
565 different definitions of a major epidemic – specifically, outbreaks in which there are at
566 least *F* infection events (the "Total infections" definition – Fig 5a) or outbreaks that persist
567 for at least *T* days (the "Duration" definition – Fig 5b).
568

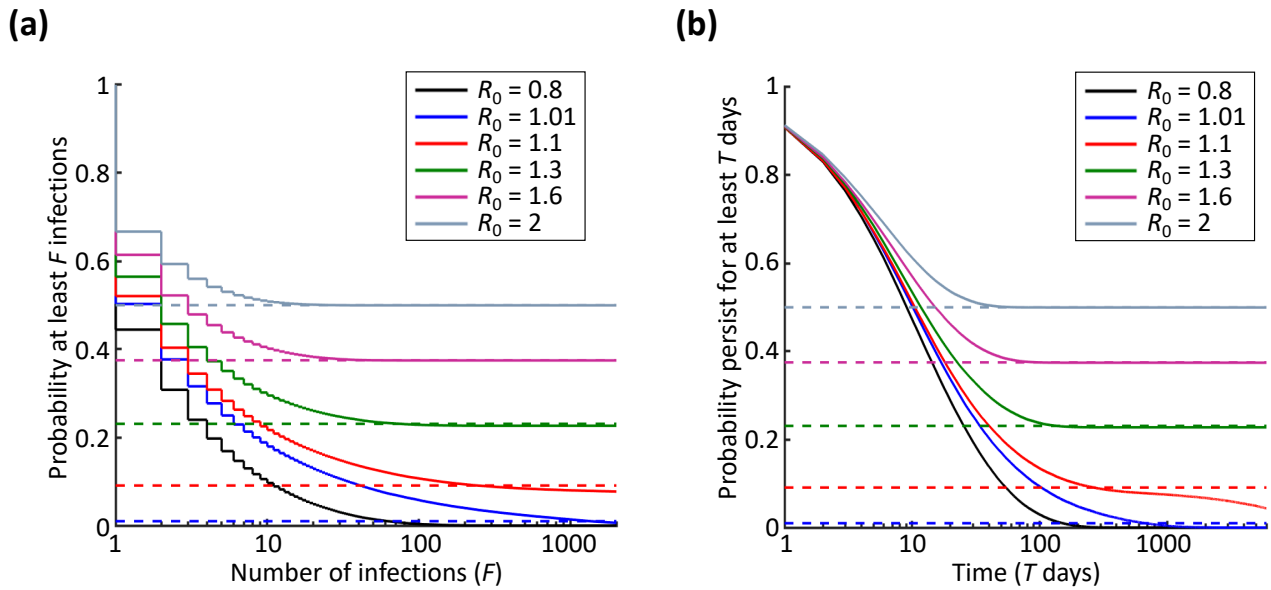

Figure 5. Probability of a major epidemic under the SIS model, for alternative definitions of a major epidemic. (a) A major epidemic is defined as an outbreak in which at least *F* infections occur ("Total infections" definition). (b) A major epidemic is defined as an outbreak that persists for at least *T* days ("Duration" definition). Solid lines represent the true probability of a major epidemic assessed via simulation of the stochastic model, and dotted lines represent the branching process estimate (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 branching process estimate corresponded to a very wide range of major epidemic thresholds under the "Total infections" or "Duration" definitions (i.e. values of *F* or *T*) compared to under the "Concurrent size" definition. As a

result, in these specific cases (i.e. when the stochastic SIS model was used and a major epidemic was defined according to the "Total infections" or "Duration" definitions) it can be concluded that the branching process approximation often leads to sensible estimates of the risk posed by invading pathogens. Nonetheless, even in these cases, for small or very large values of the major epidemic thresholds the probability of a major epidemic does not match the branching process estimate, particularly when R_0 was 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 major epidemic is the branching process approximation to the probability of a major epidemic accurate? Range of values here are those for which the branching process approximation 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 definition of a major epidemic. For which values of the thresholds in the more practically relevant definitions of a major epidemic is the branching process approximation to the probability of a major epidemic accurate? Range of values here are those for which the branching process approximation 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

Determining the risk of an emerging outbreak developing into a major epidemic is vital for planning whether or not intervention and/or containment strategies will be necessary. When a pathogen arrives in a new location, the probability of a major epidemic can be approximated by assuming that infections occur according to a branching process. For simple models such as the stochastic SIS and SIR models, this leads to the probability of a major epidemic in equation (1). It is also possible to estimate the probability of a major epidemic according to the branching process approximation using models with more complexity, as we showed by considering the case of host-vector transmission (see equations (4) and (5)).

However, the branching process estimate for the probability of a major epidemic is not necessarily accurate when definitions of a major epidemic are used that address practical aspects of disease control (Fig 3). The branching process estimate corresponds to a range of choices of the epidemic thresholds in our definitions when R_0 is much greater than one, or when the population size is extremely large (see e.g. different values of M in Fig 3a,b). However, when R_0 is close to one and the population is not large, the standard estimate can correspond to a single choice of the epidemic threshold (see e.g. blue and

red lines in Fig 3a, and Tables 1 and 2). For specific outbreaks, even when both R_0 and the population size are large, the branching process estimate may not be relevant – since the range of choices of the major epidemic threshold that the standard estimate corresponds to may or may not include the specific threshold of practical importance in the outbreak, for example the number of hospital beds available. Consequently, using the standard branching process estimate for the probability of a major epidemic could lead to the risk of a major epidemic being incorrectly assessed, potentially including scenarios in which a major epidemic develops when previously deemed unlikely. Our main conclusion that the branching process estimate for the probability of a major epidemic may or may not match the true probability of a major epidemic when a practically relevant definition is used holds for a range of epidemiological systems (Fig 4) as well as different definitions of a major epidemic that apply in alternative settings (Fig 5). We note that exactly how well the branching process estimate reflects a range of threshold values depends on the underlying model as well as the relevant quantity differentiating major epidemics from minor outbreaks (*cf*. Figs 3a and 5).

We considered practical definitions of a major epidemic that were based on thresholds such as the availability of treatment. A previous study defined major epidemics according to a threshold in the percentage of the population ever infected, and concluded that epidemiological modellers should report the precise cutoff used to define a major epidemic in model simulations [56]. Their conclusion was based on the observation that different thresholds in the percentage of hosts ever infected corresponded to wide variations in the other outputs of model simulations including the number of dead hosts

655 or the time of the epidemic peak. We support this conclusion, and indeed some authors
656 have reported the definition of a major epidemic they used clearly – for example, Keeling
657 *et al*. [57] define a major epidemic to be an outbreak in which at least one-third of the
658 population becomes infected. However, we also emphasise that the precise type of
659 threshold, and the value used, should be chosen according to practical relevance in the
660 particular system under consideration, rather than simply an arbitrary threshold in the
661 number of individuals ever infected.
662
663 Under the first definition of a major epidemic that we considered (the "Concurrent size"
664 definition), the probability of a major epidemic was assessed in the context of the capacity
665 for treatment by estimating whether or not a threshold number of simultaneously infected
666 individuals was likely to be exceeded. This definition may be practically relevant in a range
667 of scenarios. For example, real-time analysis of a diphtheria epidemic in Cox Bazar's in
668 Bangladesh involved assessing the number of hospital beds that were needed [39]. The
669 number of beds required was approximated in that study by using a model to forecast
670 disease incidence, assuming that 15% of reported cases would require treatment as
671 inpatients with an average hospital stay of five days for each case. The number of hospital
672 beds that were already available might have provided a practically relevant major
673 epidemic threshold. Another example for which this type of threshold might apply is citrus
674 greening disease in Brazil, for which a law was introduced stating that a citrus grove must
675 be destroyed if 28% of trees in the grove were infected and symptomatic [58,59]. Major
676 epidemics could therefore be defined as outbreaks in which more than 28% of trees in a
677 grove are infected and symptomatic concurrently. Other examples for which interventions

678 were introduced as soon as a threshold in the number simultaneously infected was
679 reached include the development of the National Chlamydia Screening Programme in the
680 United Kingdom in 2002 in response to the large size of the infected population [60].
681
682 However, no single definition of a major epidemic will be relevant in all situations. To
683 illustrate the ambiguity to which this might lead, we also considered two other definitions
684 of a major epidemic. In one of these (the “Total infections” definition), whether or not an
685 outbreak was classified as a major epidemic referred to the total number of infection
686 events over the course of the outbreak, rather than the maximum number simultaneously
687 infected. This might correspond to the total number of treatments required, which may be
688 an important threshold if treatments have been stockpiled prior to the outbreak [41]. This
689 definition might also be relevant if, for example, a policy-maker has to choose how to
690 deploy resources between two different epidemics. If there are only sufficient resources
691 to contain one outbreak, and both epidemics are equally controllable, then it might be
692 preferable to choose the one that is likely to generate more infections. In other real-world
693 scenarios, alternative definitions might be appropriate. We also considered major
694 epidemics defined as outbreaks that persist for a threshold length of time (the “Duration”
695 definition). Different definitions of a major epidemic might appear contradictory – for
696 example, treatment can act to reduce the total number of infections yet increase the
697 duration of the outbreak [61], making a major epidemic less likely under the “Total
698 infections” definition of a major epidemic but more likely under the “Duration” definition.
699

Our intention here was to use very simple models to demonstrate the principle that different definitions of a major epidemic lead to different probabilities of a major epidemic. Although simple models are commonly used, accurate outbreak forecasts require a model carefully matched to the epidemiology of the host-pathogen system, potentially including asymptomatic transmission [9,62,63] or spread between spatially distinct regions [29,64]. For certain definitions, it may be necessary to include convalescent hosts in the model explicitly. For example, if convalescent individuals require resources, such as beds in treatment rooms or hospitals, and the definition of a major epidemic is linked to the availability of resources (as in the case of the “Concurrent size” definition), then these individuals should be modelled, potentially by including them in a new compartment following the infectious class. More complex definitions of major epidemics could also be used, for example requiring multiple criteria to be satisfied for an outbreak to be classified as a major epidemic. In these more complicated scenarios, analytic calculations of the probability of a major epidemic might not be possible. Model simulations can then be used to assess the probability of a major epidemic, as we showed for a host-vector model of Zika virus transmission (Fig 4b).

We note that practical use of the methods presented here at the start of an emerging outbreak to assess the major epidemic risk would require the wide range of interventions that are introduced in outbreak response settings to be integrated into the models explicitly. One way in which control can be included is to consider the effective reproduction number when the pathogen arrives in the system instead of the basic reproduction number, since the effective reproduction number accounts for interventions

[26,54,55,65–67]. In that case, the results that we presented would be unchanged (except that e.g. the lines in Fig 3a would correspond to different values of the effective reproduction number). Temporal changes in interventions once an outbreak is underway have been approximated in epidemiological models by assuming that the values of the parameters governing transmission change during the outbreak, either by assuming that transmissibility changes at a single timepoint [68,69] or continuously as the outbreak progresses [70,71]. However, for detailed descriptions of control to be included in estimates of the major epidemic risk, more complex interventions should be included in model simulations. Since models are often used to test possible control strategies [12,20,72–74], this is a simple extension of the results presented here.

In summary, we have shown that how precisely to define a major epidemic should be considered carefully in future studies. The definition of a major epidemic should be designed to match the questions of interest in the particular setting being considered. Standard approximations based on branching processes should not always be used as the default. Only once a “major epidemic” has been defined precisely can the probability of a major epidemic occurring be properly assessed. Providing an explicit demonstration of the consequences of not considering a practically relevant definition in evaluating the risk of an epidemic is the key contribution of this paper.

SUPPLEMENTARY MATERIAL

Text S1.

Figure S1. Schematic showing how the probability of a major 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.

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.

FUNDING

This research was funded by BBSRC via a G2O PhD studentship (RNT), and by Christ Church, Oxford, via a Junior Research Fellowship (RNT).

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.

769 Emerging infectious diseases of plants: pathogen pollution, climate change and
770 agrotechnology drivers. Trends Ecol Evol. 2004;19: 535–544.
771 doi:10.1016/j.tree.2004.07.021
772 3. Taylor LH, Latham SM, Woolhouse MEJ. Risk factors for human disease
773 emergence. Philos Trans R Soc B. 2001;356: 983–989.
774 doi:10.1098/rstb.2001.0888
775 4. Morens DM, Folkers GK, Fauci AS. The challenge of emergin and re-emerging
776 infectious diseases. Nature. 2004;430: 242–249.
777 5. Jones KE, Patel NG, Levy MA, Storeygard A, Balk D, Gittleman JL, et al. Global
778 trends in emerging infectious diseases. Nature. 2008; 990–994.
779 doi:10.1038/nature06536
780 6. Fisher MC, Henk DA, Briggs CJ, Brownstein JS, Madoff LC, McCraw SL, et al.
781 Emerging fungal threats to animal, plant and ecosystem health. Nature. 2012;484:
782 186–194. doi:10.1038/nature10947
783 7. Thompson RN, Brooks-Pollock E. Detection, forecasting and control of infectious
784 disease epidemics: modelling outbreaks in humans, animals and plants. Philos
785 Trans R Soc B. 2019;374: 20190038.
786 8. Craft ME, Beyer HL, Haydon DT. Estimating the probability of a major outbreak
787 from the timing of early cases: an indeterminate problem ? PLoS One. 2013;8:
788 e57878. doi:10.1371/journal.pone.0057878
789 9. Thompson RN, Gilligan CA, Cunniffe NJ. Detecting presymptomatic infection is
790 necessary to forecast major epidemics in the earliest stages of infectious disease
791 outbreaks. PLoS Comput Biol. 2016;12: e1004836.

792 doi:10.1371/journal.pcbi.1004836
793 10. Ramsay ME. Measles: the legacy of low vaccine coverage. Arch Dis Child.
794 2013;98: 752–754. doi:10.1136/archdischild-2013-304292
795 11. Rohani P, Earn DJD, Grenfell BT. Impact of immunisation on pertussis
796 transmission in England and Wales. Lancet. 2000;355: 285–286.
797 12. Ferguson NM, Donnelly CA, Anderson RM. The foot-and-mouth epidemic in Great
798 Britain: Pattern of spread and impact of interventions. Science (80-). 2001;292:
799 1155–1160.
800 13. Kao RR. The role of mathematical modelling in the control of the 2001 FMD
801 epidemic in the UK. Trends Microbiol. 2002;10: 279–286.
802 14. Keeling MJ. Models of foot-and-mouth disease. Proc R Soc B Biol Sci. 2005;272:
803 1195–1202.
804 15. Keeling MJ, Woolhouse MEJ, Shaw DJ, Matthews L, Chase-Topping M, Haydon
805 DT, et al. Dynamics of the 2001 UK foot and mouth epidemic: Stochastic
806 dispersal in a heterogeneous landscape. Science (80-). 2001;294: 813–818.
807 16. Pautasso M, Aas G, Queloz V, Holdenrieder O. European ash (Fraxinus
808 excelsior) dieback – A conservation biology challenge. Biol Conserv. Elsevier Ltd;
809 2013;158: 37–49. doi:10.1016/j.biocon.2012.08.026
810 17. Cunniffe NJ, Laranjeira FF, Neri FM, DeSimone RE, Gilligan CA. Cost-effective
811 control of plant disease when epidemiological knowledge is incomplete: Modelling
812 Bahia bark scaling of citrus. PLoS Comput Biol. 2014;10: e1003753.
813 doi:10.1371/journal.pcbi.1003753
814 18. Chan M-S, Jeger MJ. An analytical model of plant virus disease dynamics with

815 roguing and replanting. J Appl Ecol. 1994;31: 413–427.
816 19. Wu JT, Cowling BJ, Lau EHY, Ip DKM, Ho L-M, Tsang T, et al. School closure
817 and mitigation of pandemic (H1N1) 2009, Hong Kong. Emerg Infect Dis. 2010;16:
818 538. doi:10.3201/eid1603.091216
819 20. Thompson RN, Gilligan CA, Cunniffe NJ. Control fast or control smart: When
820 should invading pathogens be controlled? PLoS Comput Biol. 2018;14.
821 doi:10.1371/journal.pcbi.1006014
822 21. Keeling MJ, Rohani P. Modeling infectious diseases in humans and animals.
823 Princeton University Press; 2008.
824 22. Lloyd AL, Zhang J, Root AM. Stochasticity and heterogeneity in host-vector
825 models. J R Soc Interface. 2007;4: 851–863. doi:10.1098/rsif.2007.1064
826 23. Lloyd-Smith JO, Schreiber SJ, Kopp PE, Getz WM. Superspreading and the effect
827 of individual variation on disease emergence. Nature. 2005;438: 355–359.
828 doi:10.1038/nature04153
829 24. Nishiura H, Cook AR, Cowling BJ. Assortativity and the probability of epidemic
830 extinction: a case study of pandemic influenza A (H1N1-2009). Interdiscip
831 Perspect Infect Dis. 2011; 194507.
832 25. Anderson D, Watson R. On the spread of a disease with gamma distributed latent
833 and infectious periods. Biometrika. 1980;67: 191–198.
834 26. Thompson RN, Jalava K, Obolski U. Sustained transmission of Ebola in new
835 locations: more likely than previously thought. Lancet Infect Dis. 2019;19: 1058–
836 1059. doi:10.1016/S1473- 3099(19)30360-3
837 27. Swinton J. Extinction times and phase transitions for spatially structured closed

838 epidemics. Bull Math Biol. 1998;60: 215–230.
839 28. Merler S, Ajelli M, Fumanelli L, Parlamento S, Pastore y Piontti A, Dean NE, et al.
840 Containing Ebola at the Source with Ring Vaccination. PLoS Negl Trop Dis.
841 2016;10: 1–11. doi:10.1371/journal.pntd.0005093
842 29. Thompson RN, Thompson C, Pelerman O, Gupta S, Obolski U. Increased
843 frequency of travel in the presence of cross-immunity may act to decrease the
844 chance of a global pandemic. Philos Trans R Soc B. 2019;374: 20180274.
845 doi:10.1098/rstb.2018.0274
846 30. Althaus CL, Low N, Musa EO, Shuaib F, Gsteiger. Ebola virus disease outbreak
847 in Nigeria: Transmission dynamics and rapid control. Epidemics. 2015;11: 80–84.
848 31. Abdullah N, Kelly JT, Graham SC, Birch J, Gonçalves-Carneiro D, Mitchell T, et
849 al. Structure-guided identification of a nonhuman morbillivirus with zoonotic
850 potential. J Virol. 2018;92: e01248-18.
851 32. Keeling MJ, Danon L. Mathematical modelling of infectious diseases. Br Med Bull.
852 2009;92: 33–42. doi:10.1093/bmb/ldp038
853 33. Keeling MJ, Ross J V. On methods for studying stochastic disease dynamics. J R
854 Soc Inferface. 2008;5: 171–181. doi:10.1098/rsif.2007.1106
855 34. Kessler DA, Shnerb NM. Solution of an infection model near threshold. Phys Rev
856 E. 2007;76: 010901. doi:10.1103/PhysRevE.76.010901
857 35. Park AW, Gubbins S, Gilligan CA. Invasion and persistence of plant parasites in a
858 spatially structured host population. Oikos. 2001;94: 162–174.
859 36. Kendall DG. Deterministic and stochastic epidemics in closed populations. Proc
860 3rd Berkeley Symp Math Stat Prob. 1956;4: 149–165.

861 37. de Jong MCM. Mathematical modelling in veterinary epidemiology: why model
862 building is important. Prev Vet Med. 1995;25: 183–193.
863 38. Nasell I. 'The threshold concept in stochastic epidemic and endemic models' in D.
864 Mollison, Epidemic models: their structure and relation to data. Cambridge
865 University Press; 1995.
866 39. Finger F, Funk S, White K, Siddiqui MR, Edmunds WJ, Kucharski AJ. Real-time
867 analysis of the diphtheria outbreak in forcibly displaced Myanmar nationals in
868 Bangladesh. BMC Med. 2019;17: 58. doi:10.1038/d41586-018-05810-w
869 40. Lamontagne F, Clément C, Fletcher T, Jacob ST, Fischer II WA, Fowler RA.
870 Doing today's work superbly well - treating Ebola with current tools. N Engl J Med.
871 2014;371: 1565–1566.
872 41. O'Dowd A. Government says it would stockpile Tamiflu again. Br Med J.
873 2014;349: g6386. doi:10.1136/bmj.g6386
874 42. Butler D. Africa's latest Ebola outbreak comes to swift end. Nature. 2018; 2018–
875 2020.
876 43. Wood W. The History and Antiquities of Eyam. Whitefish, MT: Kessinger
877 Publishing; 1859.
878 44. Scott S, Duncan CJ. Biology of plagues: evidence from historical populations.
879 Cambridge University Press; 2001.
880 45. Whittles LK, Didelot X. Epidemiological analysis of the Eyam plague outbreak of
881 1665-1666. Proc R Soc B Biol Sci. 2016;283: 20160618.
882 46. Brauer F. "Compartmental models in epidemiology" in F. Brauer, P. van den
883 Driessche, J. Wu - Mathematical Epidemiology. Springer; 2014. doi:10.1007/978-

3-540-78911-6

47. Raggett GF. A stochastic model of the Eyam plague. J Appl Stat. 1982;9: 212–225. doi:10.1080/02664768200000021

48. Cullerne JP, French A, Poon D, Baxter A, Thompson RN. The pedagogical power of context: extending the epidemiology of Eyam. Phys Educ. 2019;

49. Kucharski AJ, Funk S, Eggo RM, Mallet H-P, Edmunds WJ, Nilles EJ. Transmission Dynamics of Zika Virus in Island Populations : A Modelling Analysis of the 2013 – 14 French Polynesia Outbreak. PLoS Negl Trop Dis. 2016;10: e0004726. doi:10.1371/journal.pntd.0004726

50. van den Driessche P, Watmough J. Further notes on the basic reproduction number. In “Mathematical Epidemiology.” Springer; 2008.

51. Heffernan JM, Smith RJ, Wahl LM. Perspectives on the basic reproductive ratio. J R Soc Inferface. 2005;2: 281–293. doi:10.1098/rsif.2005.0042

52. 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. doi:10.1016/j.cam.2009.11.003

53. Gillespie DT. Exact stochastic simulation of coupled chemical reactions. J Phys Chem. 1977;8: 2340–2361.

54. Cori A, Ferguson NM, Fraser C, Cauchemez S. A new framework and software to estimate time-varying reproduction numbers during epidemics. Am J Epidemiol. 2013;178: 1505–1512. doi:10.1093/aje/kwt133

55. Thompson RN, Stockwin JE, Gaalen RD Van, Polonsky JA, Kamvar ZN, Demarsh

907 PA, et al. Improved inference of time-varying reproduction numbers during
908 infectious disease outbreaks. Epidemics. 2019;19: 100356.
909 doi:10.1016/j.epidem.2019.100356
910 56. Orbann C, Sattenspiel L, Miller E, Dimka J. Defining epidemics in computer
911 simulation models: How do definitions influence conclusions? Epidemics. Elsevier
912 B.V.; 2017;19: 24–32. doi:10.1016/j.epidem.2016.12.001
913 57. Keeling MJ. The implications of network structure for epidemic dynamics. Theor
914 Popul Biol. 2005;67: 1–8. doi:10.1016/j.tpb.2004.08.002
915 58. Craig AP, Cunniffe NJ, Parry M, Laranjeira FF, Gilligan CA. Grower and regulator
916 conflict in management of the citrus disease Huanglongbing in Brazil: A modelling
917 study. J Appl Ecol. 2017;55: 1956–1965. doi:10.1111/1365-2664.13122
918 59. Belasque Jr. J, Bassanezi RB, Yamamoto PT, Ayres AJ, Tachibana A, Violante
919 AR, et al. Letter to the editor: Lessons from Huanglongbing management in Sao
920 Paulo state, Brazil. J Plant Pathol. 2010;92: 285–302.
921 60. LaMontagne DS, Fenton KA, Randall S, Anderson S, Carter P. Establishing the
922 National Chlamydia Screening Programme in England: results from the first full
923 year of screening. Sex Transm Infect. 2004;80: 335–341.
924 61. Hollingsworth TD, Klinkenberg D, Heesterbeek H, Anderson RM. Mitigation
925 strategies for pandemic influenza A: Balancing conflicting policy objectives. PLoS
926 Comput Biol. 2011;7: e1001076. doi:10.1371/journal.pcbi.1001076
927 62. Hyatt-Twynam SR, Parnell S, Stutt ROJH, Gottwald TR, Gilligan CA, Cunniffe NJ.
928 Risk-based management of invading plant disease. New Phytol. 2017;214: 1317–
929 1329.

930 63. Thompson RN, Morgan OW, Jalava K. Rigorous surveillance is necessary for
931 high confidence in end-of-outbreak declarations for Ebola and other infectious
932 diseases. Philos Trans R Soc B. 2019; doi:10.1098/rstb.2018.0431
933 64. Kraemer MUG, Faria NR, Reiner RC, Golding N, Nikolay B, Stasse S, et al.
934 Spread of yellow fever virus outbreak in Angola and the Democratic Republic of
935 the Congo 2015–16: a modelling study. Lancet Infect Dis. 2017;17.
936 doi:10.1016/S1473-3099(16)30513-8
937 65. Nishiura H, Chowell G. The effective reproduction number as a prelude to
938 statistical estimation of time-dependent epidemic trends. Mathematical and
939 Statistical Estimation Approaches in Epidemiology. 2009. doi:10.1007/978-90-
940 481-2313-1_5
941 66. Delamater PL, Street EJ, Leslie TF, Yang YT, Jacobsen KH. Complexity of the
942 basic reproduction number (R0). Emerg Infect Dis. 2019;25.
943 doi:10.3201/eid2501.171901
944 67. Anderson RM. The concept of herd immunity and the design of community-based
945 immunization programmes. Vaccine. 1992;10: 928–935. doi:10.1016/0264-
946 410X(92)90327-G
947 68. Hart WS, Hochfilzer LFR, Cunniffe NJ, Lee H, Nishiura H, Thompson RN.
948 Accurate forecasts of the effectiveness of interventions against Ebola may require
949 models that account for variations in symptoms during infection. Epidemics. 2019;
950 69. Meakin SR, Tildesley MJ, Davis E, Keeling MJ. A metapopulation model for the
951 2018 Ebola outbreak in Equateur province in the Democratic Republic of the
952 Congo. bioRxiv. 2018; 1–30.

70. Chowell G, Nishiura H. Characterizing the transmission dynamics and control of Ebola Virus Disease. PLoS Biol. 2015;13: 1002057. doi:10.1371/journal.pbio.1002057

71. Chowell G, Hengartner NW, Castillo-Chavez C, Fenimore PW, Hyman JM. The basic reproductive number of Ebola and the effects of public health measures: The cases of Congo and Uganda. J Theor Biol. 2004;229: 119–126. doi:10.1016/j.jtbi.2004.03.006

72. Thompson RN, Hart WS. Effect of confusing symptoms and infectiousness on forecasting and control of Ebola outbreaks. Clin Infect Dis. 2018;67. doi:10.1093/cid/ciy248

73. Thompson RN, Cobb RC, Gilligan CA, Cunniffe NJ. Management of invading pathogens should be informed by epidemiology rather than administrative boundaries. Ecol Modell. 2016;324. doi:10.1016/j.ecolmodel.2015.12.014

74. 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. doi:10.1073/pnas.1602153113