INTERFACE

royalsocietypublishing.org/journal/rsif

Research

Cite this article: Nieddu GT, Forgoston E, Billings L. 2022 Characterizing outbreak vulnerability in a stochastic *SIS* model with an external disease reservoir. *J. R. Soc. Interface* **19**: 20220253.
https://doi.org/10.1098/rsif.2022.0253

Received: 30 March 2022
Accepted: 15 June 2022

Subject Category:
Life Sciences–Mathematics interface

Subject Areas:
systems biology, biomathematics

Keywords:
zoonosis, reservoir, stochastic modelling, outbreak, disease dynamics

Author for correspondence:
Garrett T. Nieddu
e-mail: gnieddu@gmail.com

Characterizing outbreak vulnerability in a stochastic *SIS* model with an external disease reservoir

Garrett T. Nieddu[1], Eric Forgoston[2] and Lora Billings[2]

[1]Quantitative Pharmacology and Pharmacometrics, Merck & Co., Inc., Rahway, NJ 07065, USA
[2]Department of Applied Mathematics and Statistics, Montclair State University, 1 Normal Avenue, Montclair, NJ 07043, USA

GTN, 0000-0003-0407-3795; EF, 0000-0003-0694-9981

In this article, we take a mathematical approach to the study of population-level disease spread, performing a quantitative and qualitative investigation of an *SIS*κ model which is a susceptible-infectious-susceptible (*SIS*) model with exposure to an external disease reservoir. The external reservoir is non-dynamic, and exposure from the external reservoir is assumed to be proportional to the size of the susceptible population. The full stochastic system is modelled using a master equation formalism. A constant population size assumption allows us to solve for the stationary probability distribution, which is then used to investigate the predicted disease prevalence under a variety of conditions. By using this approach, we quantify outbreak vulnerability by performing the sensitivity analysis of disease prevalence to changing population characteristics. In addition, the shape of the probability density function is used to understand where, in parameter space, there is a transition from disease free, to disease present, and to a disease endemic system state. Finally, we use Kullback–Leibler divergence to compare our semi-analytical results for the *SIS*κ model with more complex susceptible-infectious-recovered (*SIR*) and susceptible-exposed-infectious-recovered (*SEIR*) models.

1. Introduction

Throughout history, human populations have been subjected to infectious diseases which reside in reservoirs external to the population. When a susceptible human comes into contact with the reservoir, it is possible for the disease to be transmitted from the reservoir to the individual, with possible ongoing transmission of the disease throughout the human population. These reservoirs may be animal based [1] or environmentally based (plants, water sources, soil or air) [2,3]. Recent zoonotic diseases which have arisen from reservoirs include West Nile fever [4], SARS [5], Ebola virus disease [6] and COVID-19 [7].

Our ability to anticipate and react to novel infectious diseases requires that we build modelling frameworks to better understand the outbreak and spread of infectious disease. A wide array of mathematical modelling approaches has been developed specifically for studying infectious diseases [8–10]. The foundation of many of these modelling approaches is based on constructing deterministic compartmental models that typically consist of ordinary or partial differential equations. The analysis of the models can provide the basic reproduction number, which enables one to determine if a system can support a stable endemic state, along with other useful quantitative and qualitative observations.

The basic reproduction number, R_0, is used to gauge the infectiousness of a disease and can also be used to determine whether a disease will be self-sustaining. The formal definition of the basic reproduction number is the number of secondary infections that are generated on average by a single infectious individual in an otherwise fully susceptible population. In the absence of an external source of disease, an R_0 greater than one suggests that the disease will spread, while an R_0 less than one indicates that the disease will die out.

THE ROYAL SOCIETY PUBLISHING

In a deterministic disease model such as a standard susceptible-infectious-susceptible (*SIS*) model, R_0 describes the stability of the disease-free equilibrium (DFE). For $R_0 > 1$, the DFE is unstable, while for $R_0 < 1$, the DFE is stable. A stable DFE generally corresponds to deterministic disease extinction.

The inclusion of infections from the disease reservoir removes the DFE, and therefore, R_0 can no longer be interpreted in the same strict mathematical way. It has previously been shown that for populations connected weakly to a reservoir, the R_0 of the deterministic mean-field model can be used as an indicator of outbreak vulnerability for the associated stochastic model [11]. For the *SIS*κ system with a very weak connection to the disease reservoir, R_0 for the *SIS* system does a good job of indicating when a susceptible population is most vulnerable to external disease invasion.

To properly understand disease introduction into a population, we must account for human exposure through a connection to the reservoir. However, disease introduction is fundamentally a stochastic event that cannot be captured by deterministic models. Although little work has been devoted to understanding stochastic population dynamics in the presence of a reservoir, one recent article discusses, among other things, outbreak vulnerability to disease in the context of Ebola virus disease [11]. Much of this work was computational in nature. To better understand the vulnerability of populations to disease outbreak, in this article we consider a simplified model with a connected reservoir so that the model is analytically tractable. The analytical results presented later do in fact provide insight into the numerical results of the Ebola virus disease model [11], but additionally provide fundamental insight into the mechanisms which underpin outbreak vulnerability in similar stochastic systems connected to a reservoir. The analytical results of the *SIS*κ model allow one to build intuition regarding disease invasion and how different parameters affect the type of outbreak not only for *SIS*-type models but also for susceptible-infectious-recovered (*SIR*)-type and susceptible-exposed-infectious-recovered (*SEIR*)-type models (including the Ebola virus model mentioned earlier which is of *SEIR*-type with additional compartments for hospitalization and deceased individuals). Moreover, the analytical results provide a technical foundation and context for future data-driven studies.

When studying the introduction of disease into a population, it is necessary to define how the population of interest is divided. In the simplest case that we consider in this article, there is one population (the reservoir) that harbours the disease, and there is another, separate population to which the disease is introduced on occasion. Of course, if the situation is such that the reservoir is introducing disease at such a high rate that there is essentially no difference between reservoir transmission and community transmission within the non-reservoir population, then we could consider these two connected sub-populations as a single population.

More generally, the approach to defining appropriate sub-populations should be based on the specific scientific investigation. In the mathematical study of disease spread, the partitioning of human populations into distinct, well-mixed and spatially non-explicit sub-populations is common practice [8,12–14] and has been used to study a variety of epidemiological problems. For instance, it is useful to study vaccination optimization strategies in explicitly connected metapopulation models [15]. For the study of global disease extinction or the spread of disease between human populations, it is necessary to consider multiple separate sub-populations [16,17]. However, in general, there is a noted lack of rigorous study when it comes to determining how a population should be divided into the appropriate sub-populations.

To develop effective disease intervention strategies, we must be able to quantify how disease is spread from one population to another. To perform the latter analysis, we must improve our understanding of how different populations are connected to one another. Although this sounds quite straightforward, it is known that some populations are more closely linked through their behaviour than their proximity. For instance, it has been shown that under vaccination, synchrony of measles outbreaks in geographically close metapopulations has become irregular [18]. Moreover, geographically distant populations may be synchronous. For example, two cities with major air travel hubs may have enough population mixing so that an infectious disease outbreak occurs in both nearly simultaneously.

In this article, we investigate the introduction and persistence behaviour of an infectious disease in a completely susceptible population from an external source; this encroachment is called disease invasion. We take a basic and generic approach to the investigation of disease invasion by considering a simple, but novel, mathematical model for the case of a sub-population connected to a disease reservoir. We define the disease-susceptible population using an *SIS*κ model, a non-standard *SIS* model that includes a term which accounts for exposure from an external source (the reservoir). The disease is not self-sustaining within the population, and so the primary source of infection is exposure to the external disease reservoir. The vulnerability of the population to disease introduction and sustained disease presence, termed outbreak vulnerability, is investigated both qualitatively and quantitatively. We determine the probability distribution for the infected class in the most simple one-dimensional *SIS*κ case and show the extent to which it can be informative for more complex systems by comparing our closed-form solution for the *SIS*κ system with more complicated *SIR* and *SEIR* models.

2. Methods

2.1. *SIS*κ deterministic model

Before considering the full stochastic system, we investigate the mean field dynamics of the *SIS*κ system. The *SIS*κ ordinary differential equations model, including infection from the external disease reservoir, is given as follows:

$$\frac{\mathrm{d}S}{\mathrm{d}t} = \mu N - \mu S - \beta \frac{IS}{N} - \kappa S + \gamma I \tag{2.1}$$

and

$$\frac{\mathrm{d}I}{\mathrm{d}t} = \beta \frac{IS}{N} + \kappa S - (\mu + \gamma) I. \tag{2.2}$$

We will assume a constant population size so that $N = S + I$. This allows one to reduce the system to a single-state variable with governing equation given as follows:

$$\frac{\mathrm{d}I}{\mathrm{d}t} = \beta I \left(1 - \frac{I}{N}\right) + \kappa (N - I) - (\mu + \gamma) I. \tag{2.3}$$

A standard *SIS* model has two physically meaningful equilibria: the extinct state and, when the disease is self-sustaining ($R_0 > 1$), the disease endemic state. When a non-zero and positive κS is included, there will be a steady introduction of disease from an external source. This means that the disease-free state is never

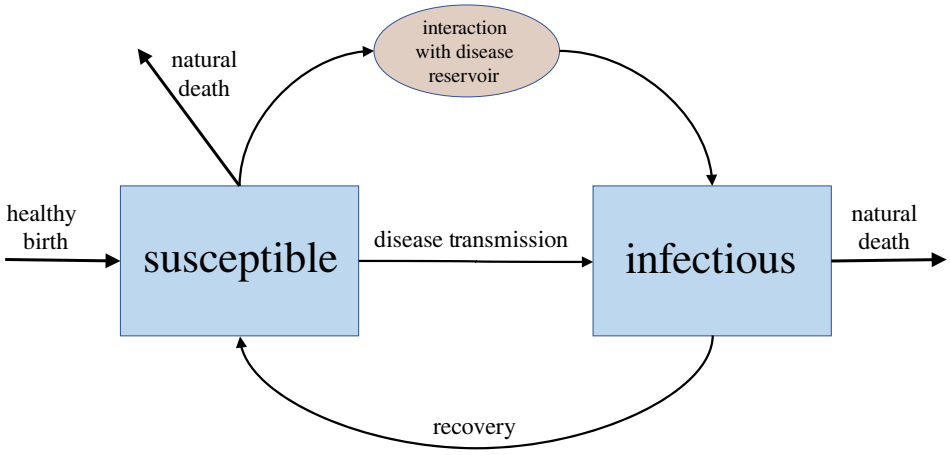

Figure 1. Flow chart showing the movement of individuals within an *SISκ* population.

stable. As a result, the *SISκ* system with $R_0 < 1$ has only one physically meaningful equilibrium point given by

$$I^* = \frac{N[\beta - (\mu + \gamma + \kappa)] + N\sqrt{[\beta - (\mu + \gamma + \kappa)]^2 + 4\beta\kappa}}{2\beta}, \quad (2.4)$$

which is also the mean of the stationary probability distribution associated with the stochastic system.

2.2. *SISκ* stochastic model

The model assumes that the susceptible population is well mixed and non-spatially explicit and that disease can be introduced to the population from an external source/reservoir. We mathematically model this scenario using a standard *SIS* model that is connected to a disease reservoir. Figure 1 shows a diagram representing the compartments and possible transitions that may occur within this novel *SISκ* model. Table 1 provides details about these transitions and their associated rate constants.

A disease reservoir may be unknown or poorly understood, as is the case for the Ebola virus disease in Africa [19,20] or the COVID-19 pandemic. Because of the difficulties associated with modelling a poorly understood reservoir, we choose to account for disease introduction from an external source via a generic infection from the reservoir [11].

In table 1, this generic infection event is labelled 'disease transmission (reservoir),' and its associated rate is κS. The rate includes κ, which indicates how strongly the population is coupled with the external disease reservoir, and S, which is the number of susceptible individuals in the population. This article will consider the non-endemic case, so that in the absence of external exposure ($\kappa = 0$), the population will be (or will quickly become) disease free since the disease is not self-sustaining in the population. Mathematically the deterministic system that is associated with our disease model has a basic reproduction number less than one ($R_0 < 1$) as mentioned previously.

An equal birth and death rate constrain the model to a stationary average population size N. If N is approximated as a constant, then the system is overdetermined, and the equation $I + S = N$ is used to reduce the *SISκ* system to one dimension—this leaves us only dynamically modelling the change in the number of infected individuals over time. Disease transfer and demographic events are assumed to be discrete continuous-time Markov processes, with no autocorrelation in time. The population can thus be modelled using the discrete stochastic master equation given as follows:

$$\frac{d\mathbf{P}(\mathbf{X}, \mathbf{t})}{dt} = \sum_r [W(\mathbf{X} - \mathbf{r}; \mathbf{r})\mathbf{P}(\mathbf{X} - \mathbf{r}, t) - W(\mathbf{X}; \mathbf{r})\mathbf{P}(\mathbf{X}, t)], \quad (2.5)$$

where $W(\mathbf{X}; \mathbf{r})$ is the transition rate from $\mathbf{X}$ to $\mathbf{X} + \mathbf{r}$, with $\mathbf{r}$ consisting of positive or negative integer values, and $\mathbf{P}$ is the probability function for the state variable $\mathbf{X}$. In the one-dimensional *SISκ* system, the state variable, $\mathbf{X}$, tracks the number of infectious individuals I and is real valued. Therefore, $\mathbf{X} = I$ and $\mathbf{P} = [P_0, P_1, P_2, P_3, \ldots, P_N]$. From equation (2.3) one can see that the transition rates for the one-dimensional *SISκ* system are $W(I; +1) = \beta I + \kappa N$ and $W(I; -1) = \beta I^2/N + (\kappa + \mu + \gamma)I$. By substituting these and associated transition rates into equation (2.5), one finds that each equation takes the form

$$\frac{dP_I}{dt} = [\beta(I-1) + \kappa N]P_{I-1} - \left[\beta I + \kappa N + \frac{\beta I^2}{N} + (\kappa + \mu + \gamma)I\right]P_I + \left[\frac{\beta(I+1)^2}{N} + (\kappa + \mu + \gamma)(I+1)\right]P_{I+1}. \quad (2.6)$$

Note that the probability that the population size is less than zero is zero ($P_{k<0} = 0$), as is the probability that the population size is greater than N ($P_{k>N} = 0$). Details of the derivation of equation (2.6) can be found in appendix A.

Table 1. The model as described by possible transitions and their corresponding average transition rates. Here, N is the average population size, I is the number of individuals in the infectious compartment, S is the number of individuals in the susceptible compartment, μ is the birth and death rate, β is the contact rate within the population, γ is the recovery rate, and κ is the strength with which the population is coupled to the external disease reservoir.

description	transition	rate
healthy birth	$(S, I) \to (S+1, I)$	μN
natural death	$(S, I) \to (S-1, I)$	μS
natural death	$(S, I) \to (S, I-1)$	μI
disease transmission (human-to-human)	$(S, I) \to (S-1, I+1)$	$\frac{\beta IS}{N}$
disease transmission (reservoir)	$(S, I) \to (S-1, I+1)$	κS
recovery	$(S, I) \to (S+1, I-1)$	γI

3. Results and discussion

3.1. Outbreak vulnerability

Outbreak vulnerability refers to a susceptible population's vulnerability to the introduction of infectious disease from

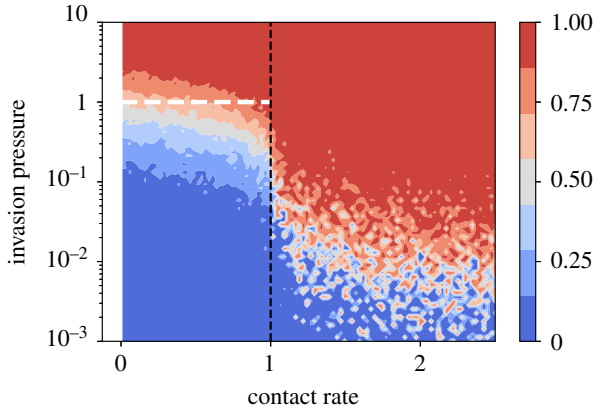

Figure 2. Contour plot showing the proportion of time a population spends with a disease present, as determined using a stochastic simulation (see appendix B for details of the numerical method) of the *SISκ* system with transition rates shown in table 1. Simulations were run out to 2000 days, and observations were recorded every 2 days. The proportion of time spent 'disease present' is assumed to be the same as the proportion of observations for which at least one diseased individual was present in the population. The parameters used for these simulations are $\{\mu, \gamma, \beta, \kappa, N\} = \{5 \times 10^{-5}, 1, \beta, \kappa, N\}$, where the product $N\kappa$ is 'invasion pressure' (shown on vertical axis) and β is given on the horizontal axis as 'contact rate'. The black dashed line is $R_0 = 1$. The white dashed line is $N\kappa = 1$(infections/day).

an external source. Although it is difficult to quantify, the concept is straightforward. Populations that are strongly connected to an external source of disease will be more vulnerable to outbreak than those with a relatively weak coupling to the external source. Similarly, populations with fewer disease-susceptible individuals are relatively less vulnerable to disease outbreak. This can be seen in figure 2 which shows the proportion of time a population spends with a disease present. In figure 2, the horizontal axis is associated with the within-population contact rate, β, and the vertical axis is associated with the strength of external exposure to disease in a disease-naive population, κN, also called the invasion pressure. The figure shows that for any particular contact rate, the proportion of time spent with disease present increases with invasion pressure. Notice that two dashed lines have been included to help guide the eye. The black dashed line is $R_0 = 1$, and the white dashed line is $N\kappa = 1$(infections/day). The region to the right of the $R_0 = 1$ line contains the parameter sets for which the disease is self-sustaining and also is the region of non-validity for our analysis. We assume that the primary cause of infection is connection to an external disease reservoir and not from contact-based infections occurring within the population. The white dashed line gives us a reference when considering the contour plot. It is clear that for the example population, one new infection from an external reservoir per day results in nearly continuous disease presence in the population. Note that this is dependent on the frequency of observation, recovery rate and (as can be seen in the figure) contact rate, among other considerations.

Although it might seem obvious that an increase in the contact rate or invasion pressure will lead to more disease, figure 2 provides a classical sensitivity analysis that allows one to understand the sensitivity of a disease severity metric, namely, the time spent with disease present, to changes in model parameters. One can easily determine what the region of non-validity looks like along with precisely why it is the region of non-validity. Specifically, when the contact rate becomes high enough so that $R_0 > 1$, disease will spread rapidly within the population, and the disease severity metric is insensitive to changes in invasion pressure. The figure also shows that the metric is insensitive to changes in the contact rate when $R_0 < 1$ and allows one to see how the system reacts to changes in invasion pressure.

3.2. Stationary distribution

Although the master equation shown in equation (2.5) cannot be solved analytically in most cases, it is possible to find an analytic solution for one-dimensional, single-step processes such as the *SISκ* system shown in equation (2.6) [21]. The solution is a probability distribution that describes the probable number of infectious individuals in the *SISκ* population. This distribution is centred around the stable equilibrium point shown in equation (2.4) and is given in its general form as follows:

$$P(I) = \pi_0 \prod_{n=1}^{I} \frac{\lambda(n-1)}{\delta(n)}, \tag{3.1}$$

where $\lambda(n)$ is the rate of increase to the infectious class, $\delta(n)$ is the rate of decrease to the infectious class and π_0 is a prefactor. The *SISκ* rates are provided in table 1. Recalling that in the constrained one-dimensional system the susceptible class is represented in terms of the infectious class as $S = N - I$, one finds that

$$\lambda(I) = \frac{\beta I(N-I)}{N} + \kappa(N-I) \tag{3.2}$$

and

$$\delta(n) = (\mu + \gamma)I. \tag{3.3}$$

Using Mathematica and equations (3.2)–(3.3), the exact solution of equation (3.1) for the constrained *SISκ* system is determined to be

$$P(I) = \pi_0 \frac{\Gamma(N+1)\Gamma((N\kappa/\beta)+I)}{\Gamma(N-I+1)\Gamma(I+1)\Gamma\left(\frac{N\kappa}{\beta}\right)} \left(\frac{\beta}{N(\gamma+\mu)}\right)^I, \tag{3.4}$$

where $\Gamma(x)$ is the Gamma function.

Nine example distributions are shown in figure 3, each corresponding to a different parameter set. The parameter sets were chosen such that in the absence of a connection to the external disease reservoir, the populations were non-endemic with respect to the disease, i.e. $R_0 < 1$. Each subfigure contains a probability distribution as determined by equation (3.4) and a histogram based on stochastic simulation. In each case, we can see that the distribution and simulation match well, which offers confidence in the analytical result. Each subfigure corresponds to different invasion pressures; subfigures closer to the top correspond to larger κ-values, and subfigures farther to the right have more susceptible individuals in the disease-naive population as captured by the parameter N. As N and κ increase, the invasion pressure becomes larger, and the population is more likely to contain disease. The simulated data presented in red is associated with high invasion pressure. In those populations, the disease will be effectively endemic even though it is not self-propagating within the population.

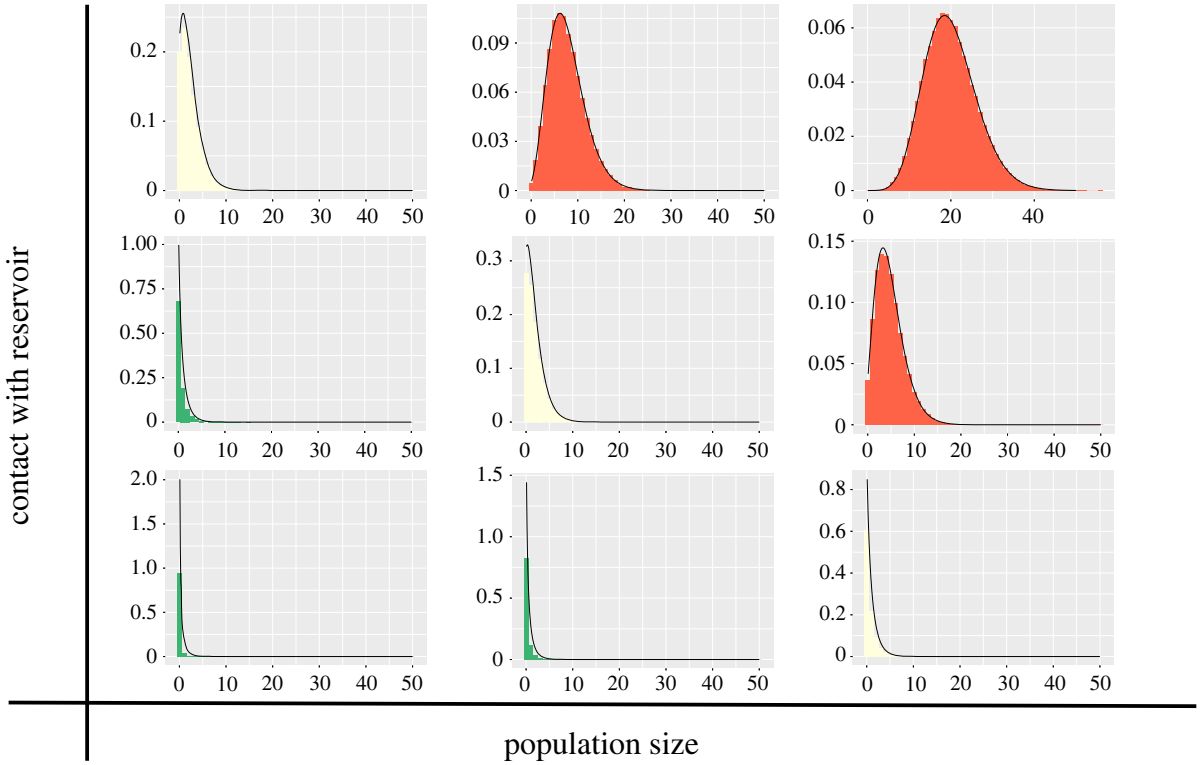

Figure 3. In each sub-panel the black curve corresponds to a probability distribution calculated using equation (3.4), and each histogram is the empirical distribution, calculated using stochastic simulation (see appendix B). The panels that are farther right have larger population sizes, and those closer to the top have larger κ-values (more contact with the reservoir).

3.3. Connectivity of the reservoir

The relationship between disease persistence and population size has been studied for over 50 years [22]. In 'Deterministic and stochastic models for recurrent epidemics', Bartlett aptly stated that 'This notion of a critical size is difficult to discuss quantitatively···' (p. 97), before giving what he described as a crude argument for a threshold [23]. In part, this difficulty is due to the indefinite nature of random processes. Invasion and persistence behaviour are closely related, and it is therefore not entirely surprising that a dependence on population size is seen in the invasion system given by the master equation (equation (2.6)). While critical community size has been shown to relate to birth rate, here we will show that the outbreak behaviour is analogously dependent on the strength with which the system is coupled to a disease's external reservoir.

In our current study, we have investigated disease persistence in a susceptible population in the presence of an external disease reservoir. Although classification of different outbreak behaviours is subjective, we offer an approach to differentiating between different kinds of outbreak behaviours. As shown in figures 3 and 4, different invasion pressure results in a different shape and position of the probability distribution. In this section, we will draw connections between population properties and the qualitatively different outbreak behaviours shown in figure 4. The population properties are mathematically represented by the model's parameters which are associated with the demographic and infection events.

On the basis of the numerical simulations for the $SIS\kappa$ system as well as previous work modelling Ebola virus disease by the authors [11], we have chosen to recognize three qualitatively different outbreak behaviour types: (i) outbreaks that are small and have no great impact on the population, and which can therefore be largely ignored; (ii) outbreaks that are large, occur frequently and result in a significant disease presence in the population, and which therefore are of interest to the affected community as well as to epidemiologists and policy-makers; and (iii) reservoir-driven large outbreaks and/or consistent disease. In scenario (iii), the $SIS\kappa$ model gives uninteresting results since the inexhaustible external source of disease will consistently maintain a high level of infection within the population of interest indefinitely. Instead, one should dynamically model the source of disease for scenario (iii).

Written more succinctly, the three different types of outbreak behaviours are as follows: (i) rare outbreaks, (ii) sporadic outbreaks, and (iii) frequent outbreaks. These correspond to parameter combinations that display weak, intermediate and strong external force of infection, respectively. This differentiation can be seen in figure 3, which shows nine example theoretical distributions and the corresponding stochastically simulated histograms that show the frequency of infectious individuals, and in figure 4, which shows examples of stochastic simulations under low, medium and high invasion pressure. As invasion pressure and population size are increased, so does the force of infection. As one moves upward and rightward in figure 3, the force of infection increases as does the outbreak frequency.

3.3.1. Calculating the normalization constant for the stationary distribution

The probability distribution given by equation (3.4) fully describes the stochastic behaviour of the $SIS\kappa$ system while at quasi-equilibrium, noting that a true equilibrium cannot be achieved due to the stochastic nature of the problem. However, the solution provided is incomplete, as there is no analytic closed form for the normalization constant π_0. Indeed, for most problems, the normalization constant, or prefactor,

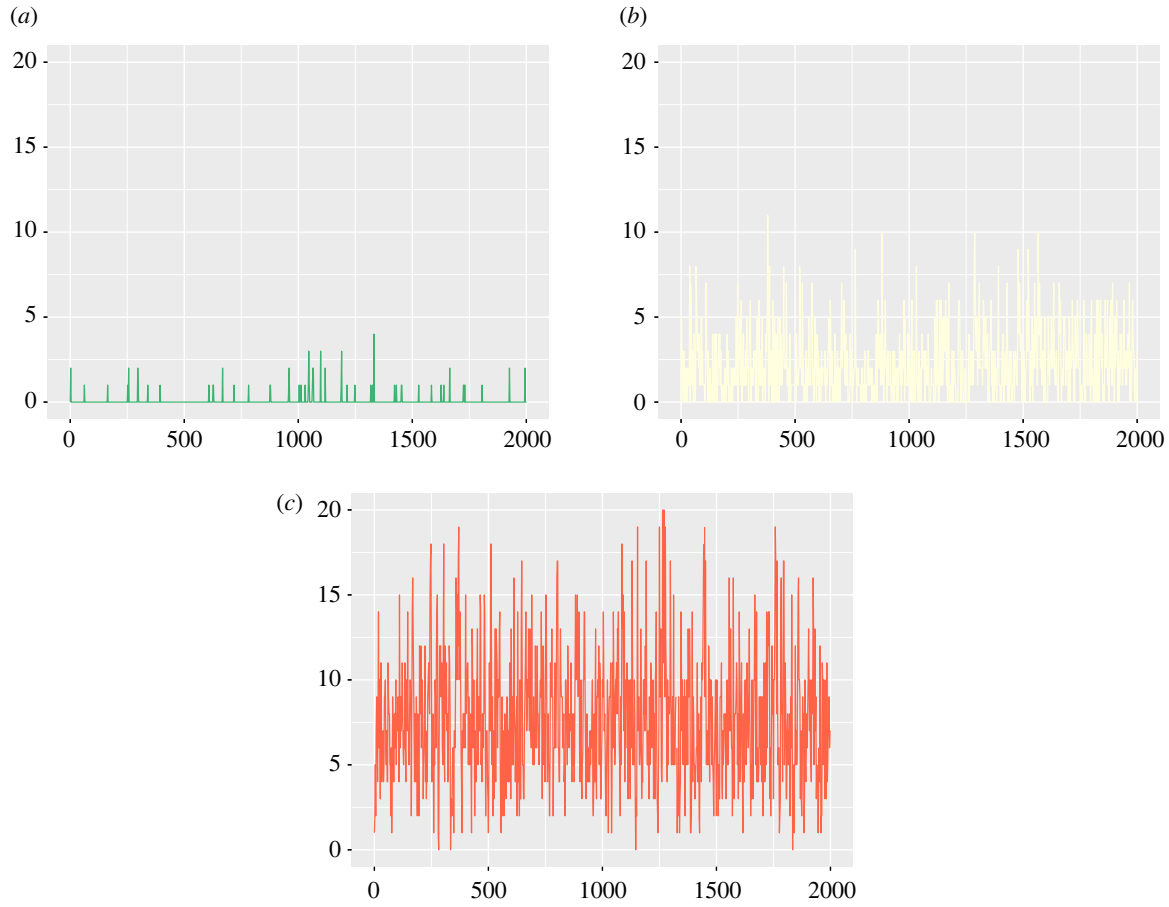

Figure 4. Three examples of stochastic simulations (as described in appendix B) for the *SISκ* system, with (*a*) low invasion pressure, (*b*) medium invasion pressure and (*c*) high invasion pressure.

cannot be found analytically and instead must be calculated numerically. For our model, we can determine the probability that an *SISκ* population contains between I_a and I_b infectious individuals. This probability is given as follows:

$$\mathbb{P}_{ab} = \frac{\sum_{I_a}^{I_b} \hat{P}(x)}{\sum_{0}^{\infty} \hat{P}(x)}, \tag{3.5}$$

where $\hat{P}(x) = P(x)/\pi_0$ is the unnormalized distribution. By definition, the sum of the probability distribution across all possible system states will equal one so that

$$\mathbb{P}_{\text{tot}} = \sum_{0}^{\infty} P(x) = \pi_0 \sum_{0}^{\infty} \hat{P}(x) = 1, \tag{3.6}$$

where $P(x)$ is the normalized probability distribution. Solving for π_0 gives

$$\pi_0 = \left[\sum_{0}^{\infty} \hat{P}(x)\right]^{-1}. \tag{3.7}$$

Although a closed-form solution of π_0 is difficult to obtain, equation (3.7) provides a way to numerically determine the normalization constant. Additional details on the derivation of π_0 for the *SISκ* system are provided in appendix C.

The probability, $\mathbb{P}_{ab}$ given by equation (3.5) can be used to develop intuition about a population's expected outbreak behaviour. Consider the infectious class distribution around the point $I = 0$. The greater the value at $I = 0$, the more probable it is to find the population disease free at any given time. In addition, the more the distribution is dispersed away from $I = 0$, the more disease cases will occur on average. This means that a narrow distribution encompassing $I = 0$ corresponds to a high probability of there being zero infectious individuals in the population. As the centre of the distribution moves rightward, or as the distribution becomes more spread out, the probability of having more than zero infectious individuals increases. This can be seen in figure 3. The three green distributions in the lower left have narrow distributions around $I = 0$. The three yellow distributions are dispersed slightly rightward, and the red distributions have a significantly increased spread and central tendency.

The red distributions are peaked away from $I = 0$ because of high invasion pressure from the reservoir. The parameters used for those simulations correspond to frequent outbreaks. Such parameters comprise the region (in parameter space) of non-validity for this modelling approach. A key assumption of the *SISκ* model is that the disease is randomly but consistently being introduced into the system from an inexhaustible external source. If the connection to that source is also very strong, then the population becomes indistinguishable from the disease reservoir. Any disease dynamics unique to the system are drowned out by the flood of infections coming from the external source.

Figure 5 shows the value of the normalization constant for 4800 different parameter sets. In each case, the value of π_0 was found numerically using the sum in equation (3.7). The value of the normalization constant decreases as $N\kappa$ increases in size.

Note that moving up and to the right on the figure (corresponding to an increase in population size and connectivity) means that one is moving closer to an externally forced endemic state. Near this endemic state, contact with

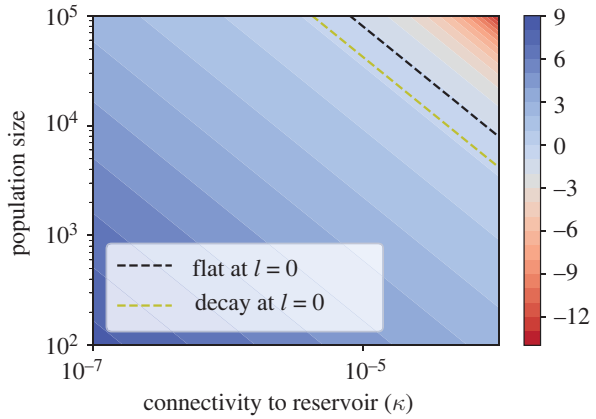

Figure 5. Contour plot of the logarithm of the numerically calculated normalization factor π_0 for $\{\mu, \gamma, \beta, \kappa, N\} = \{5 \times 10^{-5}, 1.0, 0.5, \kappa, N\}$. The black dashed line denotes where the distribution is peaked at $I = 0$, and the yellow dashed line denotes where the slope of the distribution at $I = 0$ corresponds to that of exponential decay.

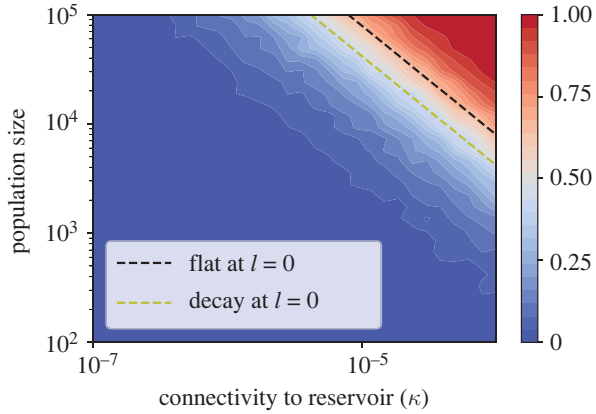

Figure 6. Contour plot of the proportion of time spent disease-present over 2000 simulation-days (see appendix B for details of the numerical method). The parameters used for these realizations are $\{\mu, \gamma, \beta, \kappa, N\} = \{5 \times 10^{-5}, 1.0, 0.5, \kappa, N\}$. The black dashed line denotes where the distribution is peaked at $I = 0$, and the yellow dashed line denotes where the slope of the distribution at $I = 0$ corresponds to that of exponential decay.

the external reservoir would cause enough disease cases so that there are always infectious individuals in the population. This endemic state is not well defined for the stochastic system presented here, and the boundary between 'outbreak-extinction cycles' and 'endemic' is fuzzy. In the following section, we discuss one approach to quantifying the boundary, the results of which can be seen as dashed lines on figures 5 and 6.

Figure 6 shows the proportion of time spent with the disease for the same 4800 parameter sets used in figure 5. For each parameter set, a stochastic Monte Carlo simulation was run for 2000 simulation-days with the system state recorded every 2 days. The time periods during which the population had a non-zero number of infectious individuals was identified and summed. This value was scaled by the total time. In the upper right of figure 6, where $N\kappa$ is largest, the population spends approximately all of its time with the disease; for these parameter combinations, the population is perpetually endemic with the disease, so that the region of parameter space will be referred to as the perpetually endemic region.

3.3.2. Quantifying the boundary

As previously touched upon, the barrier that defines the perpetually endemic region is not a simple threshold. Although there is the extreme case when a population is clearly disease endemic, there is also a fuzzy or ambiguous region in parameter space where the disease may become endemic or may be non-endemic with frequent outbreaks. The boundary is observed to be approximately coincident to where the shape of the probability distribution at $I = 0$ transitions away from a decaying form. We use the derivative of the distribution at $I = 0$ as a basic indicator of that shape.

The first derivative of the distribution is given by

$$\frac{dP}{dI} = P(I)\left(\ln\left(\frac{\beta}{(\gamma+\mu)N}\right) + \psi\left(\frac{I\beta + N\kappa}{\beta}\right) + \psi(N - I + 1) - \psi(I + 1)\right), \tag{3.8}$$

where $\psi(x) = \Gamma'(x)/\Gamma(x)$ is the Digamma function. When $I = 0$, equation (3.8) simplifies to

$$\left.\frac{dP}{dI}\right|_{I=0} = \pi_0\left(\ln\left(\frac{\beta}{(\gamma+\mu)N}\right) + \psi\left(\frac{N\kappa}{\beta}\right) + \psi(N+1) - \psi(1)\right). \tag{3.9}$$

In equation (3.9), note that $\psi(1) = -\hat{\gamma}$, where $\hat{\gamma} \approx 0.5772$ is the Euler–Mascheroni constant and the hat is used to differentiate the Euler–Mascheroni constant from the disease recovery rate γ.

The asymptotic expansion of the Digamma function [24] is given as follows:

$$\psi(x) = \ln(x) - \frac{1}{2x} - \frac{1}{12x^2} + \frac{1}{120x^4} + \ldots. \tag{3.10}$$

In the case of sufficiently large N and $N\kappa$, if this expansion is used in equation (3.9), then the terms $\psi(N\kappa/\beta)$ and $\psi(N+1)$ can be simplified to $\ln(N\kappa/\beta)$ and $\ln(N+1)$, respectively. This simplification reveals the following closed-form approximation for the first derivative of the distribution:

$$\left.\frac{dP}{dI}\right|_{I=0} = \pi_0\left(\ln\left(\frac{N\kappa}{\mu+\gamma}\right) - \hat{\gamma}\right). \tag{3.11}$$

Note that the leading order approximation contains the factors which contribute most strongly to the shape of the distribution at $I = 0$. Notably, the contact rate, β, does not appear at leading order. Rather, we see that the invasion pressure, κN, primarily governs the shape of the distribution at $I = 0$. This result is consistent with the results of the sensitivity analysis shown in figure 2. Specifically, in equation (3.11), we see that a change in the contact rate is not expected to significantly affect the shape of the stationary distribution at $I = 0$, and in fact, the sensitivity analysis (figure 2) shows that the disease severity metric is insensitive to changes in the contact rate. For small invasion pressure, the slope of the probability distribution will be negative; for intermediate invasion pressure, the slope will be zero, which corresponds to a distribution peaked about $I = 0$; and for larger invasion pressure, the distribution becomes peaked away from $I = 0$ and the slope becomes positive again. This leading order approximation agrees with our observations, but loses important information. Although equation (3.11) is useful for building intuition, it is most useful to use numerical methods to analyse equation (3.9).

We would like to understand how changes in the population structure contribute to a transition from very few or zero cases to a disease endemic state. This can be accomplished by considering how the shape of the distribution reacts to changes in the parameter values. Particularly, we will consider the transition from a sharply peaked distribution at $I = 0$ to a flat distribution at $I = 0$. This is accomplished by numerically finding where equation (3.9) has a slope consistent with exponential decay, and where it has a slope of zero. In figures 5 and 6, these slopes are labelled as 'decay' and 'flat', respectively. In practice, this is equivalent to solving $\mathrm{d}P/\mathrm{d}I\,|_{I=0} = -\pi_0$ (since if $P(x) = \pi_0 \exp(-x)$, then $\mathrm{d}P/\mathrm{d}x = -\pi_0 \exp(-x)$, and thus $\mathrm{d}P/\mathrm{d}x\,|_{x=0} = -\pi_0$) and $\mathrm{d}P/\mathrm{d}I\,|_{I=0} = 0$ (since if $P(x) = C$, then $\mathrm{d}P/\mathrm{d}x = 0$, and thus $\mathrm{d}P/\mathrm{d}x\,|_{x=0} = 0$), respectively.

These two equations can be arranged to implicitly define $N\kappa$. We then use the Broyden–Fletcher–Goldfarb–Shanno algorithm [25] to numerically solve the equations. The relationship between N and κ is shown in figures 5 and 6 as the dashed downward sloping lines. The coincidence between the shape of the distribution and consistent disease presence is shown in figure 6. Note that as the distribution becomes 'flat' at $I = 0$, we see that the population is expected to have diseased individuals for greater than half of the time.

3.4. Applicability to more complex disease systems

The results of this work are predicated on a simplification of a two-dimensional *SIS* system with external disease exposure. Results from previous work [11] has shown that similar simplifications can be used to give insight into the fundamental behaviour of more complex models and into real-world disease dynamics. Since analogous analytical results are intractable for more complex models, we have used numerical simulation to compare our results with these more complex models.

To determine how well the analytical distribution given by equation (3.4) can explain the outbreak behaviour of the more complex models, we use the Kullback–Leibler (KL) divergence [26] to compare simulated results from four models (one- and two-dimensional *SISκ*, *SIRκ* and *SEIRκ*) with the analytical results. The KL divergence is a distance measure between two probability distributions; small KL divergence values indicate good agreement between the two distributions, and large KL divergence values indicate poor agreement. For each model, we used simulated frequency of infectious individuals to approximate a probability distribution. Parameters were chosen to be constant across the four models, as much as possible.

Good agreement can be seen in figure 7 between our analytical results and the *SISκ* simulated results in both one and two dimensions. For the other two models investigated, *SIRκ* and *SEIRκ*, good agreement is seen when the overall invasion pressure is weak. Once either κ or N is increased so that disease introductions are occurring more than an average of once per day, the one-dimensional analytic *SISκ* results make a poor proxy for the more complex models. The precise point of disagreement will depend on many factors, including the disease properties (a 2 day cold versus a 10 day flu, for instance) and demographic factors.

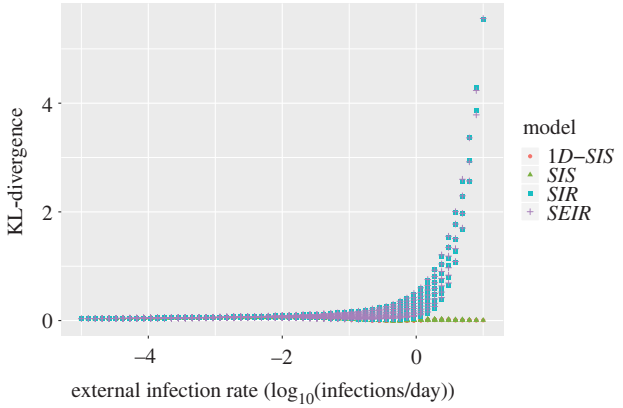

Figure 7. The relationship between the external infection rate and KL divergence for four different models: one- and two-dimensional *SISκ*, *SIRκ* and *SEIRκ*. The divergence is taken between the empirical distribution of infectious individuals (as obtained through stochastic simulation) and the 1D *SISκ* distribution given in equation (3.4). Each point corresponds to a stochastic simulation. Since multiple κ and N values were used, there is a vertical spread in the KL-divergence for any particular external infection rate.

4. Conclusion

In this article, we have presented a novel disease invasion model termed the *SISκ* model which assumes the presence of an inexhaustible disease source which is external to the population of interest. Disease exposure from the external source is proportional to the size of the susceptible class. In this work, the population size is assumed to be constant during the period of investigation, which allows us to find a closed-form equation, up to a normalization constant, for the stationary distribution that describes the state of the infected class. This stationary distribution was used, along with stochastic simulation, to investigate the transition from a practically disease-free state to a disease endemic state. It is important to note that the primary driver of disease is external exposure to the reservoir; community disease spread between population members is possible, but is not common enough to sustain the disease within the population. The *SISκ* model is a simplification of a previously investigated *SEIR*-based Ebola virus disease model [11] whose results were computational only due to the complicated nature of the model. The analytical results presented here give insight into stochastic disease invasion for the simplified model as well as to those previous results pertaining to stochastic Ebola virus disease invasion and outbreak vulnerability. More generally, the *SISκ* analytical results explain stochastic disease invasion for a wide variety of *SIS*-type, *SIR*-type and *SEIR*-type epidemic models that contain a connection to a reservoir.

Here, we have presented a stochastic view of disease invasion and outbreak. Although these phenomena are fundamentally random processes, much of the mathematical research on the topics has been done from a deterministic perspective. In these studies, there is a tendency to make generalizations for a particular collection of sub-populations. Depending on the nature of the scientific inquiry, this approach can be useful, but it commonly overlooks the issue of practical population segmentation that is necessary for the application of a theoretical framework needed to understand disease spread. In this article, we have investigated how a disease can be expected to behave in a non-endemic population over a range of population sizes and invasion pressures. When invasion pressure is low, the non-endemic population is distinct from the

external source of disease. As invasion pressure is increased, the population stops being non-endemic with the disease. In analogy, the external reservoir is not trickling disease into the population of interest, but has turned on a 'fire-hose' of disease. If the connection between the external reservoir and the population of interest is so strong, then a fundamental assumption has been violated; the external reservoir is no longer distinct from the population of interest. When the external reservoir and the population of interest cannot be considered properly distinct from one another, then the disease should be modelled dynamically throughout, and it is an indication that you may have segmented your population incorrectly.

If properly segmented, then external disease sources may not need to be considered explicitly. This means that, rather than dynamically modelling the world's population, a generic invasion term can offer sufficient insight into more local disease dynamics. Given the relatively local way that disease prevention and eradication strategies are devised and implemented, this could lead to better informed policy-makers and higher quality information being used for decision-making. The approach presented here only assumes that there will be a disease reservoir external to the system of interest and that there is some rate of introduction from that external source. This is a more practical level of investigation for informing policy, as compared with the holistic approach (i.e. trying to understand the intricacies of the entire population network, which may be composed of multiple cities, states or countries).

In this article, we have investigated the behaviour and sensitivity of our population system under different population sizes and coupling strengths, but we have not addressed how a coupling strength should be independently determined. This is an open question, and we view it as an interesting topic for future work. Intuitively one would assume that the geographic distance between two populations is likely to be important; the closer two populations are to one another, the easier it is to intermingle. There are a myriad of other potential factors, including possible government restrictions to travel, customs and culture of a population, and other factors that can affect the physical contact between the populations. Quantifying coupling strength is an interesting and complex topic. To address such a difficult question, we expect that mathematical modelling techniques will be most effective in coordination with data-driven studies. Moreover, such a data-driven scientific approach would enable a model validation study with global sensitivity and uncertainty analysis [27,28] to compare with the theoretical results presented in this article.

Ethics. Not applicable.

Data accessibility. The code used to perform the computations is available at https://github.com/GNieddu/SIS_externalRes.

Authors' contributions. G.T.N.: conceptualization, formal analysis, investigation, methodology, validation, visualization, writing—original draft, writing—review and editing; E.F.: conceptualization, formal analysis, funding acquisition, investigation, methodology, project administration, resources, supervision, writing—original draft, writing—review and editing; L.B.: conceptualization, investigation, writing—review and editing.

All authors gave final approval for publication and agreed to be held accountable for the work performed therein.

Conflict of interest declaration. We declare we have no competing interest.

Funding. G.T.N. and E.F. gratefully acknowledge support from the National Science Foundation award CMMI-1233397.

Disclaimer. The contents of this manuscript represent the views of the authors and not Merck & Co. or Montclair State University.

Appendix A. Derivation of equation (2.6)

The two-dimensional mean-field $SIS\kappa$ model is provided in the main text (equations (2.1)–(2.2)) as follows:

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

and

$$\frac{\mathrm{d}I}{\mathrm{d}t} = \beta\frac{IS}{N} + \kappa S - (\mu+\gamma)I.$$

Assuming a constant population size so that $N = S + I$, one can reduce the two-dimensional system to a one-dimensional governing equation provided in the main text (equation (2.3)) as follows:

$$\frac{\mathrm{d}I}{\mathrm{d}t} = \beta I\left(1 - \frac{I}{N}\right) + \kappa(N - I) - (\mu+\gamma)I.$$

This one-dimensional equation can be rewritten as follows:

$$\frac{\mathrm{d}I}{\mathrm{d}t} = (\beta I + \kappa N) - \left(\frac{\beta I^2}{N} + (\kappa+\mu+\gamma)I\right), \tag{A 1}$$

$$= \lambda(I) - \delta(I), \tag{A 2}$$

where $\lambda(I)$ and $\delta(I)$ terms are the growth and decay terms, respectively. For a one-step process with growth rate equal to $\lambda(I)$ and decay rate equal to $\delta(I)$, then the transition rates for the one-dimensional $SIS\kappa$ system are as follows:

$$\left.\begin{aligned} &W(I;\ +1) = \lambda(I), \quad W(I;\ -1) = \delta(I) \\ \text{and}\ \ &W(I-1;\ +1) = \lambda(I-1), \quad W(I+1,\ -1) = \delta(I+1). \end{aligned}\right\} \tag{A 3}$$

By substituting these transition rates into the master equation provided in the main text (equation (2.5)) as follows:

$$\frac{\mathrm{d}\mathbf{P}(\mathbf{X},\mathbf{t})}{\mathrm{d}t} = \sum_r [W(\mathbf{X}-\mathbf{r};\mathbf{r})\mathbf{P}(\mathbf{X}-\mathbf{r},t) - W(\mathbf{X};\mathbf{r})\mathbf{P}(\mathbf{X},t)],$$

one finds that

$$\frac{\mathrm{d}P(I)}{\mathrm{d}t} = \lambda(I-1)P(I-1) - \lambda(I)P(I) + \delta(I+1)P(I+1) - \delta(I)P(I)$$

$$= (\beta(I-1)+\kappa N)P(I-1) - (\beta I + \kappa N)P(I) \tag{A 4}$$

$$+ \left(\frac{\beta(I+1)^2}{N} + (\kappa+\mu+\gamma)(I+1)\right)P(I+1)$$

$$- \left(\frac{\beta I^2}{N} + (\kappa+\mu+\gamma)I\right)P(I), \tag{A 5}$$

which can be rearranged into the master equation for the one-dimensional system provided in the main text (equation 2.6) as follows

$$\frac{\mathrm{d}P_I}{\mathrm{d}t} = [\beta(I-1)+\kappa N]P_{I-1} - \left[\beta I + \kappa N + \frac{\beta I^2}{N} + (\kappa+\mu+\gamma)I\right]P_I$$

$$+ \left[\frac{\beta(I+1)^2}{N} + (\kappa+\mu+\gamma)(I+1)\right]P_{I+1}. \tag{A 6}$$

Appendix B. Numerical simulations

Stochastic realizations, representing possible disease trajectories, of the $SIS\kappa$ model were produced using a type of

Monte Carlo method. The method was originally proposed by Kendall [29] for simulating birth-death processes and was popularized by Gillespie [30] as a useful method for simulating chemical reactions based on molecular collisions. The results of a Gillespie simulation are a stochastic trajectory that represents an exact sample from the probability function that solves the master equation. Therefore, the method can be used to simulate population dynamics where molecular collisions are replaced by individual events and interactions including birth, death and infection [31].

Let $\mathbf{x} = (x_1, \ldots, x_n)^T$ denote the state variables of a system, where x_j provides the number of individuals in state x_j at time t. The first step of the algorithm is to initialize the number of individuals in the population compartments $\mathbf{x}_0$. For a given state $\mathbf{x}$ of the system, one calculates the transition rates denoted as $a_i(\mathbf{x})$ for $i = 1 \ldots l$, where l is the number of transitions. Thus, the sum of all transition rates is given by $a_0 = \sum_{i=1}^{l} a_i(\mathbf{x})$.

Random numbers are generated to determine both the next event to occur as well as the time at which the next event will occur. One simulates time τ until the next transition by drawing from an exponential distribution with mean $1/a_0$. This is equivalent to drawing a random number r_1 uniformly on (0, 1) and computing $\tau = (1/a_0)\ln(1/r_1)$. During each random time step, exactly one event occurs. The probability of any particular event taking place is equal to its own transition rate divided by the sum of all transition rates $a_i(\mathbf{x})/a_0$. A second random number r_2 is drawn uniformly on (0, 1), and it is used to determine the transition event that occurs. If $0 < r_2 < a_1(\mathbf{x})/a_0$, then the first transition occurs; if $a_1(\mathbf{x})/a_0 < r_2 < (a_1(\mathbf{x}) + a_2(\mathbf{x}))/a_0$, then the second transition occurs, and so on.

Finally, both the time step and the number of individuals in each compartment are updated, and the process is iterated until the disease goes extinct or until the simulation time has been exceeded [31].

Appendix C. Derivation of π_0 for the *SISκ* model

Using the online tool made by Wolfram Alpha LLC, a closed-form solution to equation (3.1) can be found. By entering

$$\text{Prod}[(\text{beta}*(\text{n}-1)*(1-(\text{n}-1)/\text{N}) + \text{kappa}*(\text{N}-\text{n}+1))/((\text{mu}+\text{gamma})*\text{n}), [\text{n}, 1, \text{x}]]$$

as input at https://www.wolframalpha.com, one finds that

$$P(x) = \pi_0 \frac{(-N)^{(x)}(-\beta/N(\gamma+\mu))^x (N\kappa/\beta)^{(x)}}{x!}, \qquad (C\,1)$$

where π_0 is a normalization constant, x is the state variable representing the number of infectious individuals, $(\cdot)^{(x)}$ is the rising factorial and $(\cdot)^x$ is the usual exponentiation notation. With $(\cdot)_{(x)}$ the falling factorial, equation (C1) can be simplified using $(-N)^{(x)} = (N)_{(x)}/(-1)^x$ so that

$$P(x) = \pi_0 \frac{(N)_{(x)}(\beta/N(\gamma+\mu))^x (N\kappa/\beta)^{(x)}}{x!} = \pi_0 \frac{N!\ (\beta/N(\gamma+\mu))^x (N\kappa/\beta)^{(x)}}{(N-x)!\ x!}.$$

Although we are primarily considering x as a discrete variable that takes on whole-number values, this equation can be simplified into 'gamma' notation by using the approximation $\Gamma(n+1) = n!$ so that

$$P(x) = \pi_0 \frac{\Gamma(N+1)\Gamma((N\kappa/\beta)+x)}{\Gamma(N-x+1)\Gamma(x+1)\Gamma(N\kappa/\beta)} \left(\frac{\beta}{N(\gamma+\mu)}\right)^x.$$

Alternatively, by recognizing the binomial coefficient, the expression can be written as a modified binomial distribution

$$P(x) = \pi_0 \binom{N}{x} \left(\frac{\beta}{N(\gamma+\mu)}\right)^x \left(1-\frac{\beta}{N(\gamma+\mu)}\right)^{N-x} \left(1-\frac{\beta}{N(\gamma+\mu)}\right)^{x-N} \left(\frac{N\kappa}{\beta}\right)^{(x)}.$$

By substituting $R_0 = \frac{\beta}{\gamma+\mu}$, one obtains

$$P(x) = \pi_0 \binom{N}{x} \left(\frac{R_0}{N}\right)^x \left(1-\frac{R_0}{N}\right)^{N-x} \left(1-\frac{R_0}{N}\right)^{x-N} \left(\frac{N\kappa}{\beta}\right)^{(x)} = \pi_0\, B\left(\frac{x; N, R_0}{N}\right) \left(1-\frac{R_0}{N}\right)^{x-N} \left(\frac{N\kappa}{\beta}\right)^{(x)},$$

where $B(\cdot)$ is the binomial distribution. Summing both sides gives

$$1 = \pi_0 \sum_{-\infty}^{\infty} B\left(\frac{x; N, R_0}{N}\right) \left(1-\frac{R_0}{N}\right)^{x-N} \left(\frac{N\kappa}{\beta}\right)^{(x)},$$

which allows one to find the normalization constant as follows:

$$\pi_0^{-1} = \left\langle \left(1-\frac{R_0}{N}\right)^{x-N} \left(\frac{N\kappa}{\beta}\right)^{(x)} \right\rangle,$$

where the mean is taken with respect to the binomial distribution, $B(x; N, R_0/N)$.

References

1. Wolfe ND, Dunavan CP, Diamond J. 2007 Origins of major human infectious diseases. *Nature* **447**, 279–283. (doi:10.1038/nature05775)
2. Bertuzzo E, Finger F, Mari L, Gatto M, Rinaldo A. 2016 On the probability of extinction of the Haiti cholera epidemic. *Stoch. Environ. Res. Risk Assess.* **30**, 2043–2055. (doi:10.1007/s00477-014-0906-3)
3. Turner WC *et al.* 2016 Lethal exposure: an integrated approach to pathogen transmission via environmental reservoirs. *Sci. Rep.* **6**, 1–13. (doi:10.1038/s41598-016-0001-8)
4. Campbell GL, Marfin AA, Lanciotti RS, Gubler DJ. 2002 West Nile virus. *Lancet Infect. Dis.* **2**, 519–529. (doi:10.1016/S1473-3099(02)00368-7)
5. Shi Z, Hu Z. 2008 A review of studies on animal reservoirs of the SARS coronavirus. *Virus Res.* **133**, 74–87. (doi:10.1016/j.virusres.2007.03.012)
6. Leroy EM *et al.* 2005 Fruit bats as reservoirs of Ebola virus. *Nature* **438**, 575–576. (doi:10.1038/438575a)
7. Lam TT-Y *et al.* 2020 Identifying SARS-CoV-2-related coronaviruses in Malayan pangolins. *Nature* **583**, 282–285. (doi:10.1038/s41586-020-2169-0)
8. Anderson RM, May RM, Anderson B. 1992 *Infectious diseases of humans: dynamics and control*, vol. 28. Kettering, UK: Oxford University Press.
9. Bailey NT. 1975 *The mathematical theory of infectious diseases and its applications*, 2nd edn. High Wycombe, UK: Charles Griffin & Company Ltd.

10. Bartlett MS. 1978 *An introduction to stochastic processes: with special reference to methods and applications*. New York, NY: CUP Archive.
11. Nieddu GT, Billings L, Kaufman JH, Forgoston E, Bianco S. 2017 Extinction pathways and outbreak vulnerability in a stochastic Ebola model. *J. R. Soc. Interface* **14**, 20160847. (doi:10.1098/rsif.2016.0847)
12. Malthus TR. 1888 *An essay on the principle of population*. London, UK: Reeves & Turner.
13. May RM. 2001 *Stability and complexity in model ecosystems*, vol. 6. Princeton, NJ: Princeton University Press.
14. Verhulst P-F. 1838 Notice sur la loi que la population suit dans son accroissement. *Correspondance Mathématique et Physique Publiée par A. Quetelet* **10**, 113–121.
15. Burton J, Billings L, Cummings DA, Schwartz IB. 2012 Disease persistence in epidemiological models: the interplay between vaccination and migration. *Math. Biosci.* **239**, 91–96. (doi:10.1016/j.mbs.2012.05.003)
16. Arino J. 2006 Disease spread in metapopulations. *Nonlinear Dyn. Evol. Equ.* **48**, 1–13.
17. Khasin M, Meerson B, Khain E, Sander LM. 2012 Minimizing the population extinction risk by migration. *Phys. Rev. Lett.* **109**, 138104. (doi:10.1103/PhysRevLett.109.138104)
18. Rohani P, Earn DJ, Grenfell BT. 1999 Opposite patterns of synchrony in sympatric disease metapopulations. *Science* **286**, 968–971. (doi:10.1126/science.286.5441.968)
19. Leirs H, Mills JN, Krebs JW, Childs JE, Akaibe D, Woollen N, Ludwig G, Peters CJ, Ksiazek TG. 1999 Search for the Ebola virus reservoir in Kikwit, Democratic Republic of the Congo: reflections on a vertebrate collection. *J. Infect. Dis.* **179**(Supplement_1), S155–S163. (doi:10.1086/514299)
20. Pourrut X *et al.* 2005 The natural history of Ebola virus in Africa. *Microbes Infect.* **7**, 1005–1014. (doi:10.1016/j.micinf.2005.04.006)
21. Gardiner C. 2009 *Stochastic methods*, vol. 4. Berlin, Germany: Springer Berlin.
22. Bartlett M. 1960 The critical community size for measles in the United States. *J. R. Stat. Soc. Ser. A (General)* **123**, 37–44. (doi:10.2307/2343186)
23. Bartlett M. 1956 Deterministic and stochastic models for recurrent epidemics. In *Proc. of the 3rd Berkeley Symp. on Mathematical Statistics and Probability*, vol. 4, p. 109.
24. Abramowitz M, Stegun IA. 1967 *Handbook of mathematical functions: with formulas, graphs, and mathematical tables*. Cambridge, UK: Cambridge University Press.
25. Fletcher R. 2013 *Practical methods of optimization*. Chichester, UK: John Wiley & Sons.
26. Kullback S, Leibler RA. 1951 On information and sufficiency. *Ann. Math. Stat.* **22**, 79–86. (doi:10.1214/aoms/1177729694)
27. Pianosi F, Beven K, Freer J, Hall JW, Rougier J, Stephenson DB, Wagener T. 2016 Sensitivity analysis of environmental models: a systematic review with practical workflow. *Environ. Model. Softw.* **79**, 214–232. (doi:10.1016/j.envsoft.2016.02.008)
28. Servadio JL, Convertino M. 2018 Optimal information networks: application for data-driven integrated health in populations. *Sci. Adv.* **4**, e1701088. (doi:10.1126/sciadv.1701088)
29. Kendall DG. 1950 An artificial realization of a simple 'birth-and-death' process. *J. R. Stat. Soc. Ser. B (Methodological)* **12**, 116–119.
30. Gillespie DT. 1977 Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.* **81**, 2340–2361. (doi:10.1021/j100540a008)
31. Forgoston E, Moore RO. 2018 A primer on noise-induced transitions in applied dynamical systems. *SIAM Rev.* **60**, 969–1009. (doi:10.1137/17M1142028)