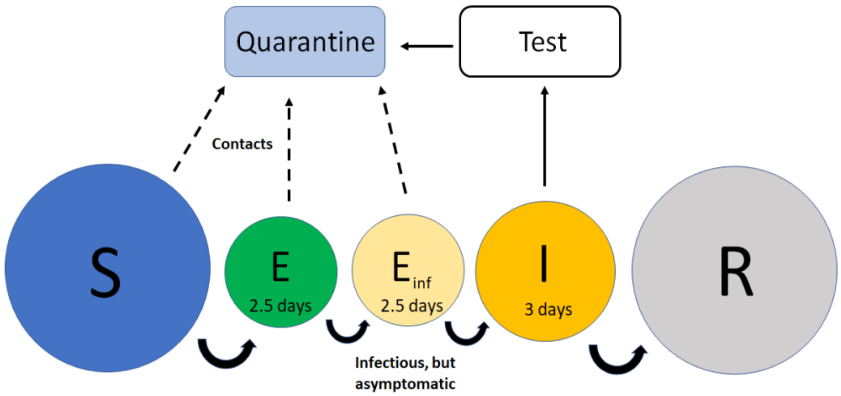
KØBENHAVNS UNIVERSITET



PhD thesis

Population dynamics and diseases

Andreas Eilersen

Advisor: Kim Sneppen

Submitted: December 28, 2021

This thesis has been submitted to the PhD school of The Faculty of Science, University of Copenhagen

Abstract

Wherever multiple organisms coexist, there are population dynamics. This thesis set out to explore the previously less well-described dynamics that arise when a population of pathogenic organisms interact with populations of prey animals and their predators. The main question to be answered was whether prey species can use the pathogens they carry as weapons against their predators, making susceptibility to diseases not exclusively a problem from an evolutionary standpoint. It turned out that this was indeed the case. We subsequently quantified when this effect would be important. Finally, we discovered that ecosystems with multiple prey species exhibit chaotic behaviour when affected by an enzootic disease.

The mathematical study of disease dynamics turned out to be highly relevant for the real world after the outbreak of the COVID-19 pandemic. Some of the complexities of disease spread are missed by traditional differential equation models of epidemics. Therefore, we produced several individual-based models for testing and quarantine strategies to mitigate COVID-19. Our models allowed us to investigate the impact of individual differences on epidemics. Under this headline, we examined the interaction between inhomogeneous disease spreading - or superspreading - and population density. We found that individual differences such as a tendency towards superspreading or inhomogeneous networks may explain some unforeseen dynamics of COVID-19. Finally, we developed evolutionary models that enabled us to guess at how SARS-CoV-2 may evolve in the future. The results of this second section point to a new understanding of epidemic diseases and particularly how to describe them mathematically.

Resumé på dansk

Populationsdynamik findes alle steder, hvor flere levende organismer sameksisterer. Til at starte med var denne afhandlings formål at udforske de dynamikker, der opstår, når en population af patogener interagerer med populationer af byttedyr og de rovdyr, der lever af dem. Hovedspørgsmålet var, om byttedyrearter kan bruge de patogener, de bærer, som våben mod rovdyr. Dette ville i givet fald gøre modtagelighed over for en sygdom til en potentiel evolutionær fordel for byttedyrene. Vores modeller viste, at dette var tilfældet, og vi forsøgte herefter kvantitativt at anslå, hvornår denne effekt ville have mest betydning. Vi opdagede ved samme lejlighed, at økosystemer med flere byttedyrearter opfører sig kaotisk under påvirkning af en enzootisk sygdom.

Matematisk sygdomsmodellering viste sig at være yderst relevant for den virkelige verden, da coronapandemien brød ud. Traditionelle differentialligningsbaserede epidemimodeller overser dog nogle af de finere detaljer ved sygdomsspredning. Derfor opstillede vi flere individbaserede modeller for test- og karantænestrategier til inddæmning af COVID-19. Disse modeller tillod os at undersøge, hvilken betydning, forskelle mellem individer har for epidemier. Herunder undersøgte vi samspillet mellem inhomogen sygdomsspredning - dvs. superspredning - og befolkningstæthed. Vi

konkluderede, at individuelle forskelle såsom varierende tendens til superspredning eller forskelle i socialt netværk kan forklare visse uventede aspekter af COVID-19's opførsel. Endelig udviklede vi evolutionære modeller, der muligvis kan bruges til at forudsige, hvordan SARS-CoV-2 vil udvikle sig i fremtiden. Resultaterne fra anden del af projektet antyder muligheden for en ny forståelse af epidemiske sygdomme og deres matematiske beskrivelse.

Acknowledgements

I would like to thank my supervisor Kim Sneppen for many inspiring ideas and continued support during all the projects that have gone into this thesis. In addition, I wish to show gratitude to my collaborators Lone Simonsen, Bjarke Frost Nielsen, Mogens Høgh Jensen, Nils Christian Stenseth, and Ruiyun Li, as well as to Jan Fabio Nickels, Anastasios Marantos, Julius Bier Kirkegaard, and Bente Markussen of the Biocomplexity group at the Niels Bohr Institute.

This thesis would not have been possible without the help of my family. I thank my mother Birgitte Eilersen, my father Christian Eilersen, and my sister Julie Eilersen for many interesting discussions and much encouragement.

Contents

Introduction

The world of living organisms is made up of countless individuals. Each individual may have specific characteristics, behaviours, and idiosyncrasies which can be unpredictable and thus hard to describe systematically. However, when individuals group together, the group will often show emergent dynamics which are not directly predictable from individual properties. That could be the cyclical fluctuations of some populations of predator and prey animals, or the exponential spread of epidemic diseases and information in human societies. Finding an accurate way to describe how populations of organisms behave and interact is a task relevant for fields as diverse as ecology, epidemiology, and social science. It is also the overarching theme of this thesis. More specifically, the main topic of this thesis is how infectious diseases affect and are affected by population dynamics. The thesis is subdivided into two parts, one dealing with the first example of predator-prey dynamics and the other dealing with disease spread in human societies. The transition reflects the outbreak of the COVID-19 pandemic and the subsequent change in focus of my own research, as well as that of many other researchers at the time.

In the first part, we investigate the interaction between predator-prey communities and pathogens. The fundamental question of this section is whether carrying a pathogen is always a disadvantage to its host species. Given the observation that many diseases are shared between species, is it possible that a prey species could transmit a disease to its predator, thereby causing it more damage than eating the prey is worth? If this is the case, one can imagine prey species using diseases as a weapon against predators, thus gaining an evolutionary advantage. Along the way, we make a variety of related discoveries. To figure out which ranges of parameter values for the proposed models are reasonable, we make use of the fact that many ecological and physiological parameters scale with animal body mass through the so-called allometric scaling. We use this to show that the dynamics of predator-prey systems can be roughly estimated from knowing only the average masses of the species involved. In order to investigate how much easier it is for a predator species to survive if it is a generalist subsisting partially on an immune prey species, we introduce a two-prey one-predator model. It unexpectedly turns out that such a model frequently gives rise to chaotic fluctuations in animal and pathogen populations.

The work on disease modelling in animals leads up to the second half of this project, which concerns the COVID-19 pandemic. Early in the pandemic, it became clear to us that traditional

SIR-type models based on differential equations were inadequate in describing the dynamics of a disease in a human society. Differential equation models are all mean-field, meaning that they are based on the assumption that every member of a society interacts at a constant rate with every other member, regardless of individual properties. One way to put it is that in a mean-field model, one models family life by having every member of society go home to a different family every day. The way we make up for this is by creating agent- or individual-based models of disease dynamics. These models allow us to take into account individual differences in network, social behaviour, and disease spread.

Our work in this field includes a model of disease spread in a society with social clusters around families and workplaces, a model of inhomogeneous infectivity ("superspreading") and how it interacts with a geographically inhomogeneous population, and a model of the effect of mass testing using an imperfect test. Finally, a group of models of disease evolution are also created using both the individual- and differential equation-based framework. These include one SIR-based model of disease evolution in an age-structured population, and individual-based models of the evolution of superspreading and disease incubation and infectious periods.

The work outlined above has helped advance our understanding of disease dynamics in the animal kingdom as well as in human society. The ecological models have shown that diseases may play an important role in evolution and species extinction. The discovery of chaotic behaviour in predator-prey-pathogen systems further suggests that disease dynamics in animals may be more complex than previously thought. In a time when focus on zoonoses is greater than before due to the pandemic, this knowledge may prove to have more than a theoretical interest.

During the course of the COVID-19-related part of the project, we have become increasingly aware of the importance of individual differences in determining the course of an epidemic. Making individual-based models of disease spread is not in itself new, but so far, more understanding of the important ways in which they differ from SIR models has been needed. This is what our work seeks to provide. Research carried out by Sneppen *et al.* [58] shows that the importance of superspreading in COVID-19 may have been the reason why lockdowns were so effective against the pandemic, since preventing mass gatherings disproportionately targets likely superspreader events. The various COVID-19-related subprojects have made use of and built upon this understanding.

In the following, I will present an outline of the work done for my PhD in the form of a synopsis of each major topic, followed by the published articles and unpublished manuscripts resulting from this work.

Chapter 1

Predators, prey, and pathogens

1.1 Problem statement and objectives

Although mathematical modelling of population dynamics in ecosystems has been an area of considerable interest for decades, much is still unknown about why ecological communities behave the way they do. One major question is why mass extinctions occur, where large ecosystems appear to collapse after periods of relative calm [57, 46, 47]. The role of infectious disease in the animal kingdom is another question which has been opened up to studies with the development of more sophisticated ecological and disease models [3, 49, 44, 34]. The fact that many infectious diseases are shared between multiple host species has also received more attention in recent years [8, 41]. In the present chapter, we want to connect the dynamics of infectious diseases in ecological communities with the larger evolutionary picture. The work described here was published in the articles [10, 15, 12], which are all appended after this chapter. As the topic of my master's thesis was disease in predator-prey systems and allometric scaling relations, there is also some overlap between the present chapter and the thesis [9].

Suppose a prey species carries a disease which is enzootic to it, i.e., the prey species coexists with the disease, with the number of infected having reached some steady state. Since we know that many diseases can infect multiple host species, it is not implausible to imagine that such a disease can be transmitted when a predator encounters or eats prey. Since a predator will need to eat prey regularly to survive and reproduce, having some fraction of the prey be able to infect it with a deadly or severe disease would be a major disadvantage for it. Carrying an enzootic disease could therefore possibly be evolutionarily advantageous for a prey species, even if the disease is harmful or deadly to the individual animals [9, 12]. This might in turn explain why some animal species such as bats seem to act as reservoirs for a wide array of diseases which are transmissible to other species [6, 9]. It may also shed some light on a possible reason for sudden extinction of predator species, pointing back to the question of mass extinctions via the breakdown of food webs.

From this main question, a variety of related questions arise. A very practical matter is how to

reduce the often large number of parameters arising when constructing eco-epidemiological models, and how to determine realistic ranges for remaining parameters. Here, we ask whether there might be some biological quantity which if not determines, then at least correlates with many other parameters. As it turns out, animal body mass is such a quantity. We will explore how this can be of practical use in constructing population dynamical models [9, 10].

Another question coming to mind is how a predator-prey-pathogen (PPP) system will behave over time. Will it reach a steady state, exhibit cyclical dynamics, or something else entirely? Since there are many possible configurations of ecological communities with both predators, prey, and pathogens, this is a fairly broad question. We will here focus on the most interesting and, for our analysis, most ecologically relevant examples.

Summary

The questions we want to answer are the following:

- When will carrying a disease which is transmissible to a predator be an evolutionary advantage for a prey species?
- How can we parameterise our eco-epidemiological model in a way that allows us to estimate realistic parameter ranges and avoid excessive unknown parameters?
- How will a model of a predator-prey-pathogen ecosystem behave over time?

1.2 Theory

Since it will be fundamental to answering the other questions, I will here begin by describing the ideas behind our method for parameterisation. This method was developed as a part of my master's thesis [9], and a draft version of the resulting paper [10] was included in the thesis.

It has long been known that several biological parameters of a given species, including metabolic rate, reproduction rate, and lifetime, scale with the average mass of the species. This is the so-called allometric scaling. It gets this name because most of the scaling laws are power laws with quarter-power exponents, whereas from the square-cube law we would expect the various metabolism-related quantities to scale with animal mass to a power of some number of thirds. Since all of the quantities involved in the Lotka-Volterra predator-prey equations are thus related to animal mass, one way of reparameterising these equations could be to rewrite them in terms of predator and prey mass. Such an idea has already been proposed before by Yodzis & Innes [67] and further elaborated by Weitz & Levin [65]. First and foremost, we here wish to investigate how well this method works as an order-of-magnitude estimate of predator-prey dynamics when applied to the simplest possible model. We examine how well the method works by testing it on a few real predator-prey systems. Once we know that the parameterisation yields at least a useable order-of-magnitude estimate of the dynamics of a simple predator-prey system, we will use it to parameterise a more complex PPP model [9].

To describe a PPP system, we will build a model based on two well-known sets of dynamical equations: The Lotka-Volterra (LV) model for predator-prey systems [40, 63] and the SIR model for infectious disease [29]. Previous models along the same lines as ours do exist [27, 28], but mostly either focus on a single-host disease or take a more purely mathematical approach than we do [9, 12]. Similarities between SIR, predator-prey, and parasitism models have also been noted [34]. What mainly sets the model we will propose apart from a model of parasitism is primarily the absence of complex life history dynamics which are often present in parasites. In addition, the prey is here assumed to be the primary host of the disease, whereas for parasites it is often the predator [7].

Furthermore, our approach will focus on the broader evolutionary perspectives of diseases shared between a prey and a predator. More specifically, we will look at the long-term survival of the species involved. This is different from most other studies in the field which mainly focus on linear stability of different equilibria. When attempting to transfer our conclusions to real-world ecosystems our approach has certain advantages. For example, a predator-prey equilibrium might be linearly stable, but if population sizes fluctuate too much, one or both populations might still go extinct during large fluctuations. Our focus on extinction events will allow us to see if existence of predators or disease is truly feasible in an ecosystem of a realistic size at a given set of parameter values [9, 12]. Given the scarcity of data on epizootics in wild animal populations, it should nonetheless be noted that our studies as presented here will examine predator-prey dynamics from a purely theoretical angle.

Since the disease and the predator both subsist on the same resource - the susceptible prey - we have some expectation that there will be competition between the two and thus possibly competitive exclusion. The position of the predator is particularly precarious since it is assumed to be a specialist in the basic version of our model, where there is only one prey and one predator species [9, 12]. We therefore set up a version of the model with two prey species, of which one is immune to the disease. We expect this to change the picture drastically by allowing the predator to survive more often and possibly also to coexist with the disease [12].

While we do find this to be the case as we will see below, we also serendipitously discover that such a system is chaotic for a wide range of parameter values [15]. This leads us to investigate the nature and behaviour of chaos in a PPP system with two prey species. While the role of chaos in ecosystems has been disputed [24], there has been a growing recognition that it might be inherent in nature [26]. An ecological model was one of the first instances of chaotic behaviour to be studied [42], and chaos has also been found in some Lotka-Volterra-derived models [23, 62, 31]. We here examine chaos specifically in a predator-prey system following a disease outbreak. Epidemics are known to frequently be unpredictable in human societies as well as animal ecosystems [45]. The chaotic dynamics that arise in such a system might provide some of the explanation for the apparently random occurence of epidemics and epizootics [15].

Summary

- Allometric scaling of animal metabolism and many related quantities with body mass is well known.
- This scaling has previously been used to parameterise ecological models. However, attempts at quantifying how realistic the results of this are, and to use the method for modelling purposes more generally have not been made. We will here begin by doing this.
- We will build our eco-epidemiological models on the LV model for predator-prey dynamics and the SIR model for disease spread.
- One main novelty of our approach is the focus on the broad evolutionary perspectives of multi-host epidemics in predator-prey systems.
- Finally, we hope that our work on chaos in PPP systems may provide an explanation for the apparent randomness of epizootic outbreaks.

1.3 Model description

The Lotka-Volterra model for predator-prey dynamics gives the following equations for prey population x and predator population y:

$$\frac{dx}{dt} = \alpha x - \phi xy \tag{1.3.1}$$

$$\frac{dy}{dt} = \nu xy - \delta y, \tag{1.3.2}$$

where α is the prey reproduction rate, ϕ is the rate at which prey is eaten per predator, ν is the predator reproduction rate per prey, and δ is the rate of predator starvation in the absence of prey. In practice, we will here take all population quantities x, y, etc. to be densities so as to avoid dealing with the absolute size of the ecosystem. The derivation shown here will follow that of Refs. [9, 10]. Initially, we parameterise these equations using the allometric relations for mammals compiled by Peters [52]. This gives us straightforwardly for the reproduction and starvation rates α and δ:

$$\alpha = \frac{1}{400} m_x^{-1/4} \quad [1/day] \tag{1.3.3}$$

$$\delta = \frac{1}{19} m_y^{-1/4} \quad [1/day]. \tag{1.3.4}$$

Here, m_x is the average body mass of a prey animal in kilograms, and m_y is that of a predator. For the predator eating and reproduction rates ϕ and ν, more careful analysis is necessary. ϕ must be

proportional to the number of prey a predator needs to eat to stay alive. We expect this to be somehow related to predator metabolism, which scales as $m_y^{3/4}$. It must also be inversely related to prey size, giving us a scaling relation of

$$\phi = k\frac{m_y^{3/4}}{m_x} \quad [1/(predator \cdot day)], \tag{1.3.5}$$

where k is some scaling constant. The ratio between ϕ and ν must be equal to the number of prey it takes to "make" one predator. This must in turn be given by the ratio between the masses divided by the *ecological efficiency*: how much of the energy gained from consuming prey will be converted into predator mass. The ecological efficiency varies between a few percent and around 20 % depending on trophic level and the specifics of the species in question [38]. In our calculations, we use an estimate of this efficiency of $f = 10$ %, the well-known "ten-percent law". The ten-percent law is wrong, but given the various difficulties of estimating an average ecological efficiency, it is likely the closest we can get to such an average without further complicating the analysis. We then have for the ratio between ϕ and ν

$$\phi/\nu = \frac{1}{f}\frac{m_y}{m_x}. \tag{1.3.6}$$

From this, we can derive a scaling relation for ν:

$$\nu = kf\frac{m_y^{3/4}}{m_x}\frac{m_x}{m_y} = kfm_y^{-1/4} \quad [1/(prey \cdot day)] \tag{1.3.7}$$

[9, 10]. All this provides us with a good order-of-magnitude estimate of predator-prey dynamics. However, the Lotka-Volterra model is in many ways too simple to accurately capture the dynamics of real-world ecosystems. It has two main issues: first, it lacks a stable fixed point or limit cycle, since the coexistence equilibrium in the classical LV model is an only marginally stable cycle. Secondly, the fact that predator population drops very far whenenever prey population is low poses a problem, since it would lead to extinction in any real ecosystem. At a later point, we will therefore introduce a modified Rosenzweig-MacArthur model which resolves some of these issues. To begin with, we will use the classical LV model to gain an understanding of the dynamics of the system before and after a disease outbreak.

We arrive at the PPP model by adding infected categories and infection terms to the above dynamical system.

$$\frac{dS_x}{dt} = \alpha S_x - \beta_{xx} S_x I_x - \phi S_x (S_y + I_y) \tag{1.3.8}$$

$$\frac{dI_x}{dt} = \beta_{xx} S_x I_x - \phi (S_y + I_y) I_x - \gamma_x I_x \tag{1.3.9}$$

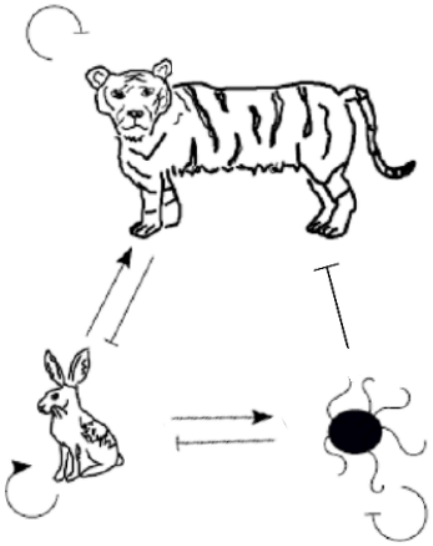


FIGURE 1.1: A diagram of the interactions of the predator-prey-pathogen model. Pointy arrows indicate positive relationships, while blunt arrows indicate negative ones. Figure from Ref. [12].

$$\frac{dS_y}{dt} = \nu(S_x + I_x)S_y - \beta_{yy}S_yI_y - \beta_{yx}S_yI_x - \delta S_y \tag{1.3.10}$$

$$\frac{dI_y}{dt} = \beta_{yx}S_yI_x - \beta_{yy}S_yI_y - (\gamma_y + \delta)I_y. \tag{1.3.11}$$

These equations as well as the mass scaling relations were first proposed in my master's thesis [9], and the mass scaling relations formed the basis of the article [10]. β_{xx} is the infection rate from prey to prey, β_{yy} is the predator to predator infection rate, β_{yx} is the prey to predator infection rate, and γ_x, γ_y are the disease death rates in prey and predator. S_x, I_x and S_y, I_y denote susceptible and infected populations of predators and prey, respectively. To achieve these simple equations, a number of assumptions had to be made. First of all, the disease is assumed to be a hundred percent fatal. This is not a good assumption for most diseases, but it allows us to drop the immune populations which would otherwise vastly complicate the dynamics of the system. Furthermore, it is assumed that infected individuals do not contribute to reproduction. Finally, all cross-species infection is assumed to be via predation. This means that the prey-to-predator infection rate β_{yx} could be written as $p_I\phi$, where p_I is the probability of infection for each infected prey eaten. Relatedly, we do not include a predator-to-prey infection term, since encounter with a predator close enough to lead to disease transmission would also result in the death of the prey. These assumptions lead to disease dynamics very similar to predator-prey or consumer-resource dynamics, meaning that to some extent the disease behaves as a second predator [9, 12]. These equations are relatively easy to modify to include a second prey species which is immune to the disease. This is further explored in the last part of this chapter.

We further parameterise the equations by noting that disease duration also scales with mass according to Cable *et al.* [5], giving us

$$\gamma_i = cm_i^{-1/4} \quad [1/day], \tag{1.3.12}$$

where c is a constant that has been determined experimentally for a variety of multi-host diseases. For our investigation, we use the c-value appropriate for rabies, but this is not crucial for our results. To get an intuition for the magnitude of β_{ij} we conduct the investigation of parameter space which will be presented below in terms of the *basic reproduction number* of the disease. The basic reproduction number is a measure of how many individuals an infectious animal in a totally susceptible population will transmit the disease to. We will here use the definition

$$R_{ij} \equiv \frac{\beta_{ij} S_{i,eq}}{\gamma_j}, \tag{1.3.13}$$

where $S_{i,eq}$ is the susceptible population of species i at the Lotka-Volterra equilibrium with no disease. R_{ij} now measures how many individuals of species i an infected member of species j will infect at the LV equilibrium. It should be noted that this is not quite the mathematically rigorous definition of R_{ij}. This would instead be defined from the *next generation matrix* of the disease (see the attatched manuscript, Ref. [16]), and would include corrections relative to the usual SIR definition in order to take the demographic terms into account. However, as these corrections are small and including them would again obscure our intuition for the magnitude of the infection term, we decide to use the informal quantities R_{ij} [9, 12].

We define an extinction threshold of 10^{-5} individuals per square kilometre. When the population of predators, prey, or infected drops below this threshold, we count the given population as extinct. A fluctuation leading to a period with such low population density would cause an extinction in most conceivable real ecosystems.

When studying extinction events, we would like to avoid counting extinctions due to artificial fluctuations caused by the unstable nature of the LV model. We therefore find it necessary to modify the above equations to make the system more stable. The assumption that predators die at a constant rate regardless of the availability of prey is also somewhat unrealistic. We therefore propose a modified Rosenzweig-MacArthur model [55], where we let predation rates saturate and predator death rates decline with increasing prey population.

$$\frac{dS_x}{dt} = \alpha S_x(1 - (S_x + I_x)/K) - \beta_{xx} S_x I_x - \phi \frac{S_y + I_y}{S_x + I_x + \epsilon} S_x \tag{1.3.14}$$

$$\frac{dI_x}{dt} = \beta_{xx} S_x I_x - \phi \frac{S_y + I_y}{S_x + I_x + \epsilon} I_x - \gamma_x I_x \tag{1.3.15}$$

$$\frac{dS_y}{dt} = \nu \frac{S_x + I_x}{S_x + I_x + \epsilon} S_y - \beta_{yx} S_y I_x - \beta_{yy} S_y I_x - \delta \frac{\epsilon}{S_x + I_x + \epsilon} S_y \tag{1.3.16}$$

$$\frac{dI_y}{dt} = \beta_{yx} S_y I_x + \beta_{yy} S_y I_x - \gamma_y I_y - \delta \frac{\epsilon}{S_x + I_x + \epsilon} I_y \tag{1.3.17}$$

[12]. Here, K is the prey carrying capacity and ϵ is the saturation constant for predation, signifying that even when prey is plentiful, the predator population is not going to grow infinitely fast.

Obviously, we will now have to reparameterise the equations. The constants α and ν are now maximal reproduction rates when food is abundant. We believe that these can be estimated based on the inverse of the gestation period for the animals in question, which sets a definite upper limit on reproduction. This analysis is again restricted to the case of mammals, but similar relations can be found for egg incubation times in other animal classes. We now have

$$\alpha \approx \frac{1}{50} m_x^{-1/4} \quad [1/day] \tag{1.3.18}$$

$$\nu \approx \frac{1}{50} m_y^{-1/4} \quad [1/(prey \cdot day)]. \tag{1.3.19}$$

The carrying capacity can be found directly from Ref. [52]:

$$K \approx \frac{1}{200} m_x^{-3/4} \quad [prey/km^2]. \tag{1.3.20}$$

Finally, the predator starvation rate δ is now determined by an assumption that when the prey population is of the same order as the saturation constant ϵ the predator should be "full" and thus able to reproduce at least at a rate large enough to keep the population at a steady state. This again gives us

$$\delta \approx \frac{1}{50} m_y^{-1/4} \quad [1/day] \tag{1.3.21}$$

[12]. The saturation constant ϵ is now the only free parameter left. It turns out that any value of ϵ between $0.3K$ and $0.9K$ gives fairly similar results, as shown in the supplement of Ref. [12]. Therefore, we simply choose $\epsilon = \frac{1}{2}K$ as the standard value in our further treatment of this model.

We are now left with a PPP model with the free parameters R_{xx}, R_{yx}, R_{yy}, and the masses of prey and predator. It turns out that in the single-prey model, R_{xx} is by far the most important of the infection parameters in determining species survival. This is reasonable, as given the limitation imposed by the ecological efficiency, the number of predators will often be far smaller than the number of prey. Predator-predator infection R_{yy} is therefore also neglected in the following, as we see in Ref. [9] that it makes little difference for survival. In the two-prey model, prey-to-predator infection will again play a key role. In the following section, we will show how well our

mass-parameterisation of the simple Lotka-Volterra model holds up against reality. We will also present the results of our examination of the parameter space of the above-mentioned free parameters. This will show us for which parameter ranges we expect the predator and the pathogen to survive. Finally, we will demonstrate in which regions of parameter space we expect the system to behave chaotically.

Summary

- Most of the parameters in our predator-prey-disease model can be quantified via mass scaling relations that are easily available from literature.
- In this way, we greatly reduce the number of free parameters.
- In addition, mass is a much more easily observable quantity than, e.g. carrying capacity or predation rate.
- We present a simple PPP model based on the classical LV equations and a more complex but hopefully also more stable model based on the modified Rosenzweig-MacArthur equations.

1.4 Results

From the mass-parameterised classical Lotka-Volterra model we are able to derive an estimate for the ratio of prey to predator population, which has the advantage of being independent of the unknown scaling factor k. We find that the equilibrium population ratio at coexistence will be

$$\frac{x}{y} = \frac{\phi\delta}{\alpha\nu} \approx \frac{21}{f}\left(\frac{m_y}{m_x}\right)^{3/4} \tag{1.4.1}$$

[9, 10]. Furthermore, since this is a centre equilibrium we expect the populations to perform cyclical fluctuations around it. From linear stability analysis, we find the expected period of the cycle to be $T = \frac{2\pi}{\sqrt{\alpha\delta}} = 2\pi(m_x m_y)^{1/8}$. This is interesting since many animal populations are seen to go through cyclical fluctuations in nature. These fluctuations are believed to be due to trophic interactions and show their own allometric relationships [53, 60]. However, the scaling observed in nature is closer to $T \propto m_x^{1/4}$ [53]. Various explanations for this have been offered. We here propose that since populations spend most of their time recovering, growing exponentially from a very low level, the duration of cycles should be found by calculating how long the population of prey will take to recover by exponential growth. If we know the peak and trough prey populations x_{max} and x_{min}, we can therefore write up the equation

$$x_{max} = x_{min}\, \mathrm{e}^{\alpha T} \quad prey/km^2, \tag{1.4.2}$$

giving us for the cycle period T

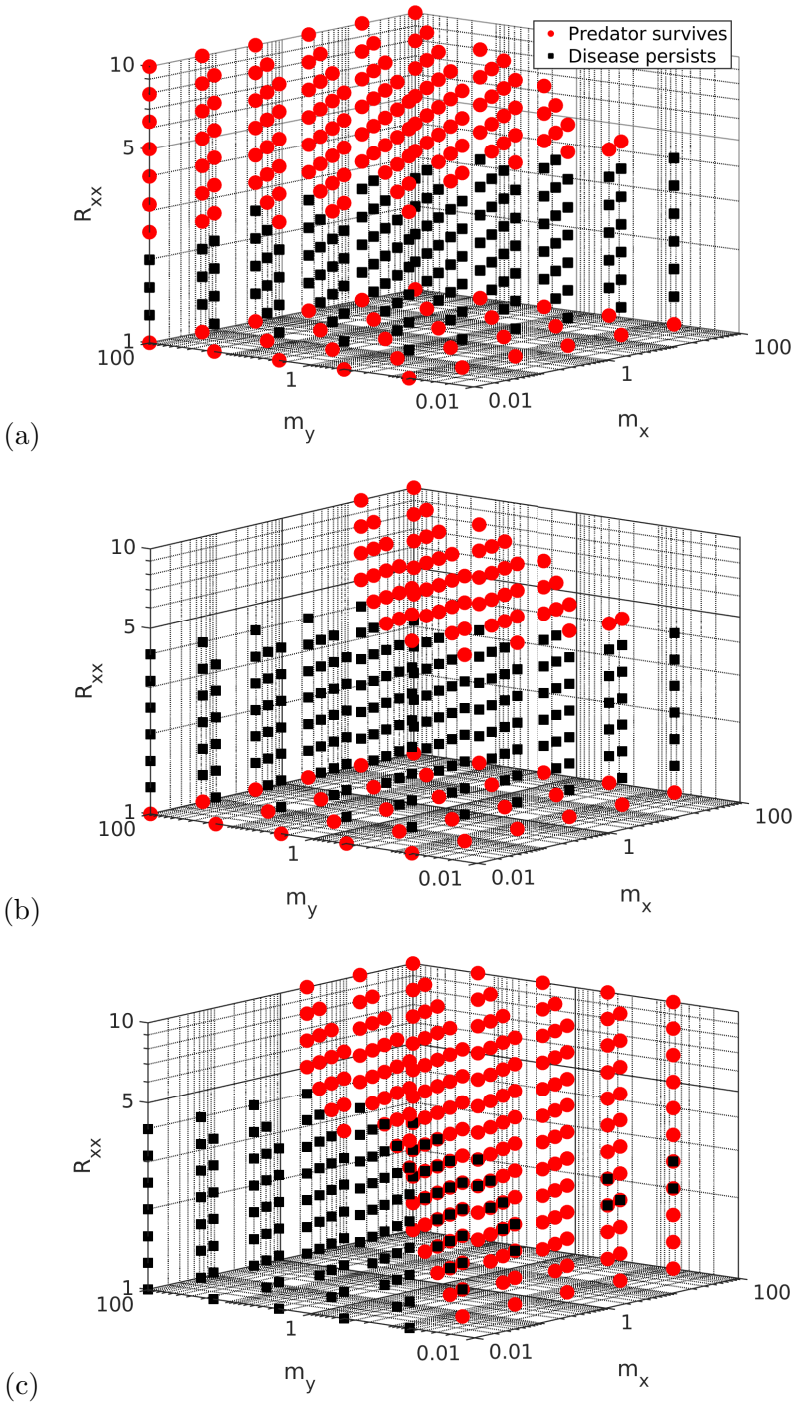


FIGURE 1.2: An overview of predator and pathogen survival in $m_x - m_y - R_{xx}$ parameter space. A red circle indicates a set of parameters for which predators survive, while a black square indicates a set of parameters where the pathogen does. In (a), predators are fully immune to the disease. Note that they nonetheless are often driven to extinction due to prey depletion. In (b) predators can be infected by eating infected prey. (c) shows the system with two prey species, of which one carries a disease which can infect the predator while the other is immune. Figure from Ref. [12].

$$T = \frac{1}{\alpha} \ln \left(\frac{x_{max}}{x_{min}} \right) \quad days \tag{1.4.3}$$

[9, 10]. This exhibits the quarter-power allometric scaling seen in nature which is independent of predator mass. The only issue is that it can be hard to find a definite value for the ratio of maximal and minimal prey populations.

When parameterising the classical LV model in the way shown above, we find that our results agree much better with observations than the order-of-magnitude estimate we were hoping for. Nonetheless, the predicted predator-prey population ratios are still off by a factor of about two, much more than experimental error would account for. This is indicative of the fact that the LV model is oversimplified, which is unsurprising. The results show that our mass parameterisation is at least good enough to gain an intuition for realistic parameter ranges and correspondences between different parameters [9, 10].

Having tested our parameterisation, we turn our attention to the predator-prey-pathogen system. Here, we observe that as long as the predator is a specialist, i.e., there is only one prey species, there is a zone of exclusion where the predator becomes unable to survive at intermediate disease infectivities. At higher infectivities, it appears that the epidemic "burns itself out" and goes extinct after an initial intense epidemic leading to depletion of susceptibles. When the prey species is small relative to the predator, the disease also causes the predator to go extinct even at higher infectivities. This is due to the fact that a large predator has to eat many small prey to survive, and in a situation where some fraction of the prey is infectious this is likely to lead to extinction of the predator, even if the epizootic quickly burns itself out [9, 12]. An attempt at analytically deriving the threshold predator-prey mass ratio for predator survival is made in our paper on the topic [12]. A three-dimensional map of predator and disease survival in parameter space can be seen in Fig. 1.2.

Remarkably, the predator is even excluded by the disease when the predators are completely immune, leading us to believe that the periodic drops in prey population is enough to cause extinction of the predator by starvation. When the predator and the pathogen subsist on the same resource - the prey - there is a form of competitive exclusion. This was in turn what led us to look at the case with one immune prey species [9, 12].

When adding a second, immune prey species with the same mass and basic characteristics as the infected prey, the predator unsurprisingly survives in all cases when there is no interspecies transmission of the disease. When there is transmission of infection from the infected prey to the predator (Fig. 1.2(c)), we again see a zone of exclusion at intermediate infectivity. As opposed to the one-prey case, there here is some coexistence of predator and disease, as well as some cases where only predators survive at intermediate infectivities. Furthermore, the threshold predator-prey mass ratio for predator survival appears to jump to a much higher value at higher

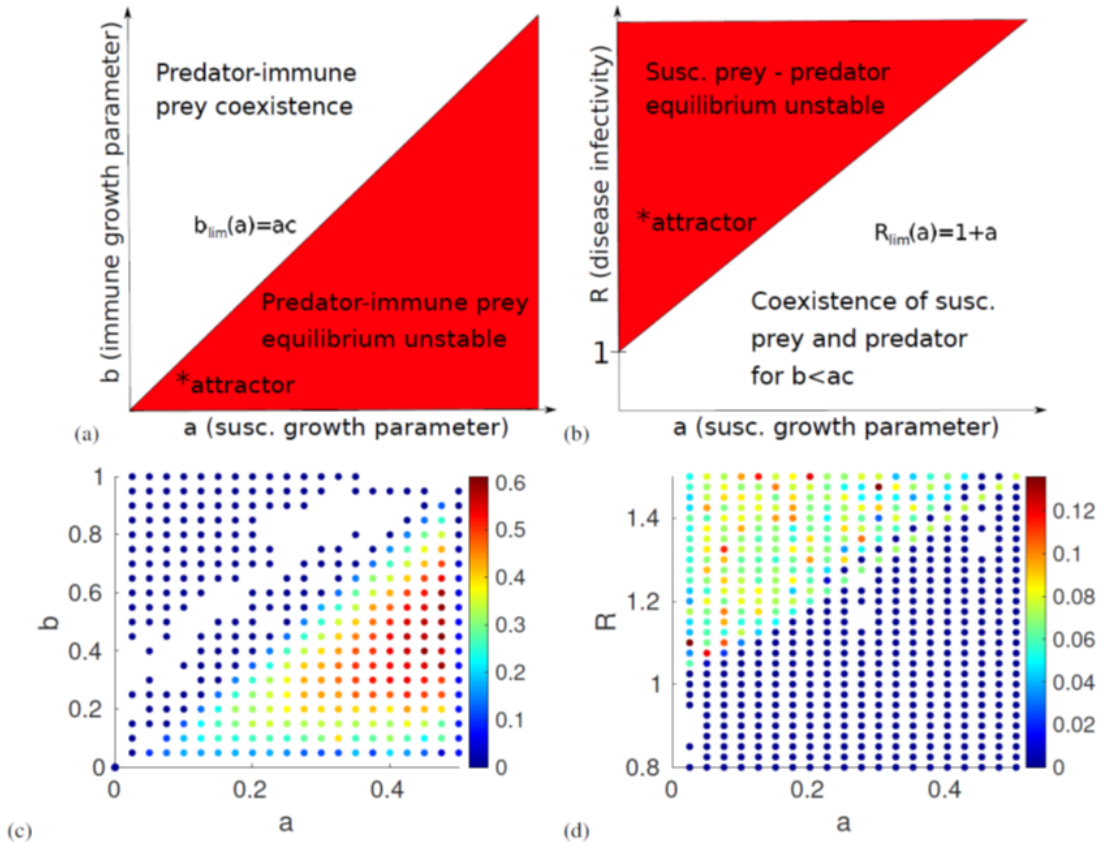


FIGURE 1.3: Stability of the two-prey, one predator system with a disease. (a) and (b) show the theoretically expected linear stability of the system as a function of a, b, and R. Suspected chaotic parameter regions are shown in red, and the location of the chaotic attractor from Fig. 1.4 is shown with an asterisk. (c) and (d) show comparable plots of our numerical measurements of the Lyapunov exponent, and they correspond well with the theoretical expectations. Due to numerical error, stable systems sometimes get a very small positive Lyapunov exponent. Parameter combinations with negative Lyapunov exponents are left out. Figure from Ref. [15].

infectivities, meaning that a species of larger predators will more often be able to survive a short epizootic since much of the prey they eat is immune [12].

At this point, we realised that the behaviour of the time series of the various populations near the parameter space region with predator-pathogen coexistence was irregular and unpredictable. Therefore, we decided to investigate if this irregularity was actually chaos and if so, whether conditions could be derived for when it arises [15].

In true chaos, the trajectory of the system depends sensitively on initial conditions, so much that even a tiny difference between two systems will make their trajectories diverge exponentially. The Lyapunov exponent is a measure of the growth rate of this exponential divergence. We therefore set out to numerically estimate the Lyapunov exponent of the two prey, one predator system for different parameter values, and to calculate analytically where we would expect chaos to occur. As a consequence, we therefore return to the basic, classical Lotka-Volterra model with a disease of eqs. (1.3.8)-(1.3.11).

Linear stability analysis of this system shows that it has four biologically feasible fixed points. Only two of these are potentially stable: One where the immune prey coexists with the predator, reaching a classical LV equilibrium, and one where the susceptible prey coexists with the predator. None of the equilibria where the disease becomes endemic is stable. We would therefore expect that chaos can only occur when both of the predator-prey equilibria are unstable. We now derive the conditions for this. First, we define the dimensionless equivalents of the reproduction/death rates and predation coupling constants: $a \equiv \alpha_{susc}/\gamma$, $b \equiv \alpha_{imm}/\gamma$, $c \equiv \phi_{imm}/\phi_{susc}$, and $d \equiv \delta/\gamma$, where $\alpha_{susc}, \phi_{susc}$ are the reproduction and predation rates of the susceptible prey species and α_{imm}, ϕ_{imm} are the equivalents for the immune prey species. In terms of these dimensionless variables, the conditions for linear stability of the two fixed points become

$$b > ac \tag{1.4.4}$$

for the immune prey and predator equilibrium, and

$$b < ac \text{ and } R < 1 + a \tag{1.4.5}$$

for the susceptible prey-predator equilibrium. Chaos should thus only be able to occur when these conditions are not fulfilled, for $b < ac$ and $R > 1 + a$. When numerically determining the Lyapunov exponents (see Fig. 1.3), we see that they are positive and greater than $5 \cdot 10^{-3}$ for all parameter sets that fulfill these conditions, meaning that the system is at least somewhat chaotic for any of these parameter sets. This shows that we have correctly determined the conditions for chaos from linear stability analysis. When analysing the trajectory of the system in the four-dimensional space spanned by the four populations, we find that in the chaotic parameter regions, an attractor is formed with a fractal dimension of 3.8 ± 0.1. This further demonstrates that the behaviour of the system is genuinely chaotic [15]. The attractor can be seen in Fig. 1.4.

In terms of animal mass, if we assume the same equivalences as above the condition $b < ac$ translates to $m_{susc}/m_{imm} > 1$. If our theoretical framework holds, we should therefore mainly expect to see chaotic behaviour when the immune prey species is smaller than the susceptible one. Since the derived conditions are not very restrictive, we would expect chaotic behaviour to be very common in nature. This might in turn explain the apparent randomness of epizootics. Furthermore, in time series of population sizes, we observe that the various populations of susceptible and infected prey and predators undergo enormous fluctuations over many orders of magnitude. This is in part due to the simplifications of our model, but if similar dynamics occur in nature, it may also be part of the explanation for why we see sudden population collapses and extinctions [15].

Summary

- By relating parameters of the Lotka-Volterra and Rosenzweig-MacArthur equations to allometric mass scaling laws, we obtain realistic descriptions of the dynamics of predator-prey

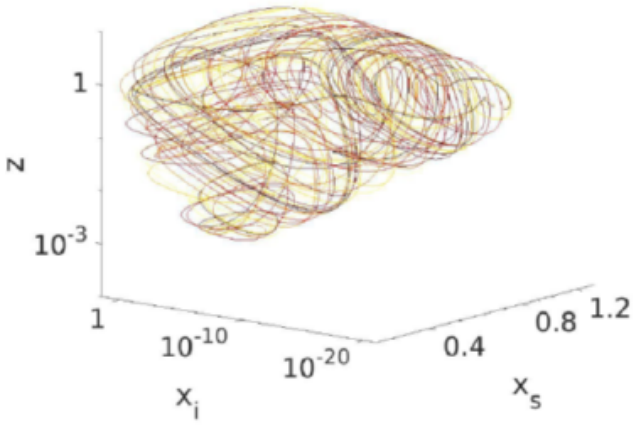


FIGURE 1.4: When the trajectory of the system in the chaotic region is plotted in susceptible prey (x_s) - infected prey (x_i) - predator (z) space, a croissant-shaped chaotic attractor with a dimension of approximately 3.8 emerges. Figure from Ref. [15].

systems, including the ratio of predator and prey population sizes and the duration of population cycles.

- We use this to describe such a system under the influence of a pathogen. Here, we show that predator and pathogen are likely to exclude one another through competition, since they are both dependent on the same resource.
- In particular, predators have trouble surviving at intermediate disease infectivities where the disease becomes endemic. They also often go extinct when the predator is so large relative to the prey that it needs to eat many potentially infectious prey to survive.
- The pathogen on the other hand runs a high risk of using up its supply of susceptibles all at once if it is very infectious.
- When including a second, immune prey, this makes it easier for the predator to survive and coexist with the disease, but does not ensure predator survival unless the predator is itself immune to the disease.
- The two-prey one-predator model with a disease in only one of the prey species shows rich, chaotic dynamics. We are able to derive conditions for when chaos will occur.

1.5 Discussion

With the above results, we make some fairly sweeping conclusions. Given the many simplifying assumptions that have to be made for us to reach these conclusions, we naturally have to insert a number of caveats. First of all, as previously noted the Lotka-Volterra model is at best an oversimplification of real predator-prey dynamics. Our paper on mass parameterisation gives an idea of the accuracy of the estimates derived from the LV model [10]. Furthermore, the mass

parameterisation of the modified Rosenzweig-MacArthur model is derived somewhat more loosely from the data in Ref. [52] than the model presented in our paper [10]. The disease dynamics of Ref. [12] are also simplified quite a bit. Particularly the assumption that the disease is a hundred percent deadly is somewhat unrealistic. Highly deadly diseases are not unusual in nature; a prominent example of such a disease is rabies, which also has many potential host species and as such was part of the inspiration for this analysis. However, including a possibility of recovery with immunity would make the model more general, if also much more complicated [9].

Our addition of a second prey species to the system is an attempt at making the model more generally applicable. The majority of predators have more than one potential prey, with even relatively specialised ones often being able to switch prey in times when their preferred prey is less available [61, 9, 12]. We demonstrate that a specialist predator will often be driven to extinction by a disease outbreak in its prey species, even when it is not itself susceptible to the disease. This is one reason why being a specialist predator is a highly precarious position. It has previously been described that competition for resources may drive species towards specialisation [22]. Our results might suggest an opposite drive towards generalisation, since subsisting on a single resource leaves a species highly vulnerable if this resource should become depleted, for example due to an epidemic [9, 12].

When examining the literature, we find that fairly few known predator-prey-pathogen systems exist where the predator is susceptible to the disease. There are a few such systems among fish [25], as well as one among chimpanzees and colobus monkeys [43] and many predator-prey-parasite systems [33]. This does however not necessarily speak against our analysis. Our analysis shows that predator-prey-pathogen systems with predator infection are relatively unstable, particularly if the predator is a specialist. Based on this, we would expect such systems to be rare. Although our inability to find them in literature are not proof that they do not exist, it is not surprising that there are few studies of PPP systems with a predator that is susceptible to the pathogen [9, 12].

The recent pandemic has highlighted the need for a better understanding of epizootic diseases among wild animals. Most human epidemic diseases start their life as zoonoses, animal diseases that gradually develop the ability to spread among humans [66]. Looking into the potential reasons for why some animal species develop into reservoirs for certain multi-host diseases and why epidemics apparently randomly flare up in animal populations as well as in human society is highly relevant in the context of understanding pandemics. This in turn points towards the next topic to be examined in this thesis, namely modelling and understanding the COVID-19 pandemic.

In this project, we have shown that even a highly simplified predator-prey model yields useable predictions about population dynamics when parameterised using allometric mass scaling relations. As an additional benefit, our work on this parameterisation has suggested an alternative explanation for the puzzling scaling relation of prey population cycle duration with prey mass, which is seemingly completely independent of the predator. This method has allowed us to take a

look at the bigger picture of disease and evolution. We find that particularly for small prey species, carrying a disease can be a sort of evolutionary weapon against its predator. This is doubly true if the disease is transmissible to the predator, but might even hold if it is not, since a specialist predator in many cases will not be able to survive the regular depletions in prey population caused by an enzootic disease. Specialisation therefore carries some risk to the predator species. We suggest that this might drive a predator species towards greater generalisation.

Finally, we show that there is a possibility of chaos in predator-prey-pathogen systems with two prey species, of which one is immune to the disease. We derive mathematical conditions for chaos and find that they are fairly easily fulfilled. Chaos should therefore be common in predator-prey systems following disease outbreaks. It is possible that this might explain the unpredictability of epizootics. All in all, this work has presented some new ways of analysing and quantifying the interaction between predation and disease. The knowledge we acquired through this project turned out to be highly useful when we moved on to describing epidemics in human society, even though the modelling methods used for this were different.

Summary

- The main caveats to note in this analysis are that the Lotka-Volterra model is of limited applicability to the real world, and that few real diseases are fatal in all cases. Instead, they often end with recovery with some degree of immunity.
- In general, few predator species are entirely specialist. The results from our model with a single prey species might provide an evolutionary reason for this, suggesting a drive towards generalisation. The model with two prey species is an attempt at making our work more generally applicable, seeing as total specialisation is rare.
- We know of few specialist predator-prey-pathogen systems with pathogen transmission from prey to predator. Since we find that such systems should be relatively unstable, this is hardly surprising.
- Most human epidemics start out as zoonoses. A better understanding of epizootics would therefore be immediately relevant, also from a human health perspective. Furthermore, the knowledge gained from studying and modelling epizootics can be transferred to human epidemics.

Chapter 2

Modelling a pandemic

2.1 Problem statement and objectives

The outbreak of the COVID-19 pandemic suddenly highlighted the need to understand epidemic dynamics. Models could give an idea of the course of the pandemic and suggest possible routes of mitigation. While the fundamentals of modelling diseases in animal populations are the same as for modelling a pandemic, the intricacies of human societies complicate matters. Humans interact in complex social networks and are able to respond to information about a spreading disease in a way that animals are not. This is why new types of models are necessary to correctly understand the spread of the pandemic.

So far, the predominant modelling framework for epidemics has been differential equation-based, like the SIR model described in the previous part. However, we realised that such models have an important limitation: They do not sufficiently take individual differences and social heterogeneities into account. Previously, attempts have been made to model this by including more compartments for each type of socialisation or age class in SIR-type models. However, since the assumption still is that all interactions within these groups are mean-field and homogeneous, important structures are still missed.

For example, network clusters form around families and other social groups. These clusters remain roughly stable over time and may be isolated from each other, geographically or otherwise. This is not accurately captured by including a compartment for each age class and having disease transmission between compartments defined by an interaction matrix of mean-field transmission rates. This limitation is particularly problematic when discussing lockdowns, which as we will see exploit the fact that infection does not spread as easily when people spend most of their time at home with the same few people. In the following, we will show that biological differences in infectivity between individuals may have further amplified this effect.

Keeping this in mind, our goal here is to create new models for the spread of epidemic diseases which are able to take social network and infection heterogeneities into account. A good modelling

framework for this is agent-based or individual-based modelling. Using this, we attempt to create models for the effectivity of lockdowns, testing and contact tracing. As we realise the importance of infection heterogeneity - "superspreading" - for the dynamics of the epidemic we also create a model for the interaction between this heterogeneity and the geographical variations in population distribution. As testing becomes more widespread, we model the potential effect of mass testing at regular intervals as a tool for epidemic management. Finally, we try to shed light on the future evolutionary direction of SARS-CoV-2 (and potentially other pathogens) through a set of evolutionary models.

The work described in this chapter has resulted in the publication of two articles [11, 13], with three more in preparation [16, 48, 14]. These articles are included after this chapter.

Summary

- We want to use individual-based models to more accurately model disease dynamics taking into account individual differences. These are especially important when dealing with a superspreader-dominated disease in a complex human society.
- Using these models, we will make an educated guess at the best strategies for lockdowns, testing and contact tracing.
- We will attempt to shed light on the future evolution of the pathogen.

2.2 Theory

Using individual-based models to describe disease spread is not itself a new idea [51, 21], and these models were also quickly put to use in the early days of the pandemic [19]. However, we set out to use them in a new way, specifically where individual differences matter. As far as we are aware, our model of testing, tracing, and quarantine, as presented in Ref. [11], was the first to make use of individual-based modelling to show the efficacy of such measures.

Previously, studies have been made which estimate the number of social interactions each individual has during the day, and which investigate in what social arenas these contacts take place. This was greatly beneficial for our work and formed the basis for many of the assumptions made in our first individual-based model. The data we needed to estimate how much of each individual's social interaction takes place in each social arena - the family, the workplace, among friends, and in public - was taken from the BBC pandemic experiment [32].

Fairly quickly, researchers realised that COVID-19 exhibited some unexpected properties. For one, infected individuals turned out to be able to transmit the disease before the onset of symptoms [64, 36]. This made the disease harder to contain, especially when tests were scarce. Furthermore, Endo *et al.* [18] showed that there was considerable individual variation in infectivity. Most people did not infect very much, whereas a few individuals, dubbed superspreaders, caused the majority of infections. This phenomenon was already known from the previous SARS coronavirus [39]. We

realised that this would have a significant effect under the constraints of lockdown and social network clustering [58].

As tests for SARS-CoV-2 infection became widely available, an idea was proposed by among others Larremore *et al.* [35] that rapid tests could be used at regular intervals to quickly quarantine those infected and thus mitigate the epidemic. This idea had already been implemented on a smaller scale at various institutions [17]. We further investigated its expected effectivity, and this mitigation strategy was later employed on a national scale in Denmark [4].

Finally, we have attempted to answer the still urgent question of which direction the evolution of the pathogen will take. To do this we have made use of the theory of evolutionarily stable strategies, or evolutionarily stable states (ESS). An ESS describes a situation where an organism can outcompete any new mutant of its species with slightly different parameters and is thus resistant to invasion by new mutants. This theory was first developed by Smith & Price [56] and was subsequently further developed by among others Reed & Stenseth [54]. A theory of ESS for parasites and pathogens has also been developed by Anderson & May [2], and we attempt to expand this further by finding the ESS of a pathogen in an age-structured population.

The previous research presented here points towards some possibilities for new theoretical investigations. Using the method of individual-based modelling, as well as the mathematical and theoretical framework that we started developing in the first chapter, we try to create models to answer the questions that are still unanswered in each area.

Summary

- Our work makes use of agent- or individual-based models to describe epidemic dynamics. Several previous models of this kind do exist, but we use them specifically to investigate the effect of individual differences on epidemics.
- It has been known for some time that diseases do not spread completely homogeneously. We realise that this may be unexpectedly significant.
- During the first year of the pandemic, regular mass testing was proposed as a mitigation strategy. We work further on assessing the effectivity of this strategy.
- Much of our work on evolution is based on the theory of evolutionarily stable states from population ecology.

2.3 Models and results

2.3.1 Network model for testing and quarantine

For the first individual-based model for testing and quarantine, we use an SEIR model of the disease progression, adding an exposed but presymptomatic state E to the basic SIR framework. The incubation times for COVID-19, at least as caused by the wildtype SARS-CoV-2 virus, are

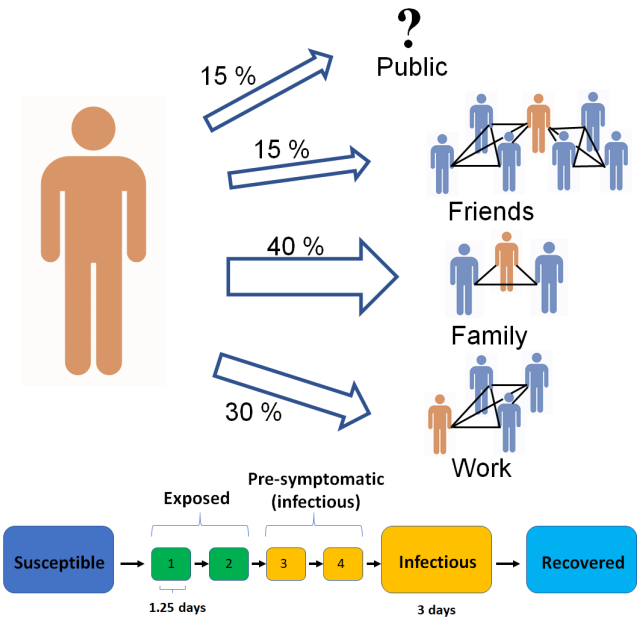


FIGURE 2.1: An illustration of the network clusters, interaction rates, and disease progression of the testing and quarantine model. Figure from Ref. [11].

roughly Gamma distributed with a mean duration of about five days [36]. This we model by subdividing the E-state into four states with exponentially distributed durations with a mean of 1.25 days each. The two last of these stages are infectious to take into account presymptomatic transmission. After the presymptomatic E-state is over, the infected then become symptomatic and move to the I-state [11].

The main advantage of this individual-based model over differential equation models is the ability to include social networks. In our model, we assume a simplified social network with society divided into clusters around families, friend groups, and workplaces. Families sizes are distributed as in data from Statistics Denmark [59], with a mean size of two persons. Workplaces contain ten people each, while each friend group contains five. Each person is assumed to have two different friend groups. We assume that everyone within each of these groups knows each other. People spend 40 % of their time with their family, 30 % at work, and 15 % with friends. The remaining 15 % is spent with randomly selected contacts from the rest of society [11]. These numbers are roughly based on Ref. [32]. See Fig. 2.1 for an illustration of the model.

Before implementing testing and quarantine, we first investigate the effect of placing restrictions on each of the social arenas outside of the family. We see some effect of this on the peak number of infected, although it is not overwhelming. Likewise, we see a small effect of reducing workplace sizes from ten to five people. This makes it clear that we need more effective measures to control the epidemic.

Such a measure could be mass testing and home quarantine of close contacts. To model testing and quarantine, we allow everyone to get tested with some probability for each day of symptomatic

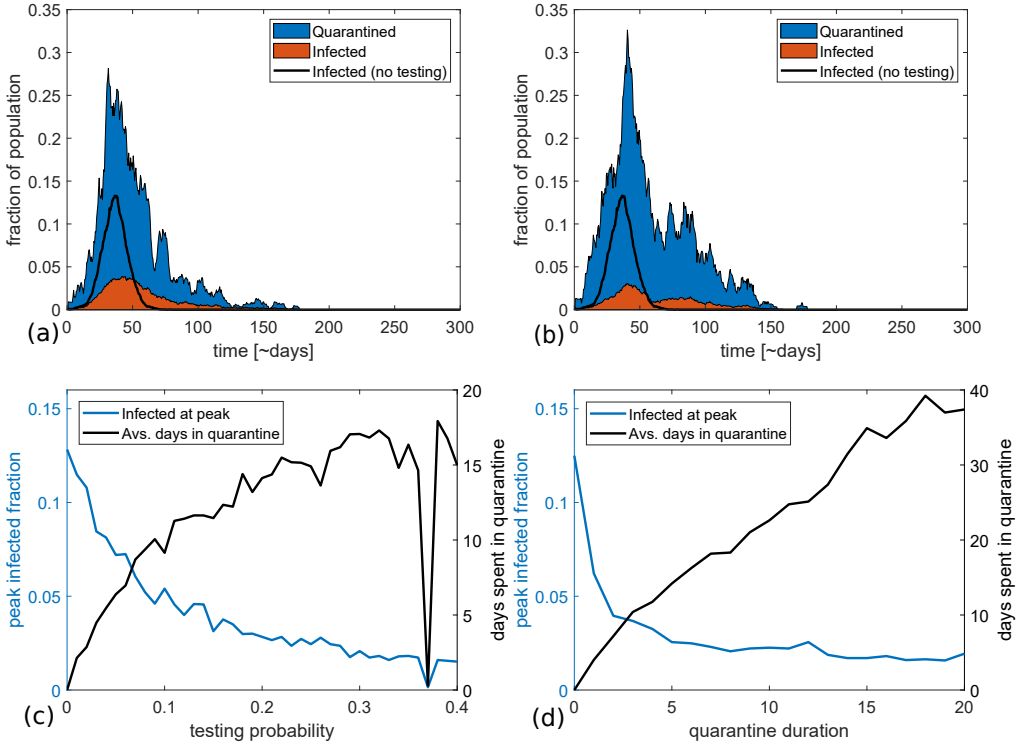


FIGURE 2.2: The effect of testing and quarantine. (a) and (b) show the trajectory of the epidemic if quarantine lasts five days and there is a probability of 10 and 20 percent respectively of being tested per day of symptomatic illness. The shaded orange area shows the number of infected and the blue area shows the number of people in quarantine. The black curve shows the epidemic trajectory without mitigation. (c,d) show the peak infected fraction and average number of days spent in quarantine as functions of testing probability per day of symptomatic illness and quarantine duration, respectively. Figure from Ref. [11].

illness. When someone is found to be infected, we quarantine "all the usual suspects": their family, workplace, and friends. They are quarantined for a set number of days and are then let out, unless they have become symptomatic in the meantime. People are assumed to quarantine together with their families. We measure the cost of this strategy by the number of days each person has to spend in quarantine and its effectivity by the peak number of infected. This is a reasonable measure in a situation where the healthcare system might be strained during the epidemic peak [11].

We find testing and quarantine to be much more effective than a simple reduction of the contact rates. It reduces the peak number of infected by two-thirds, even if the probability of getting tested is only ten percent per day of symptomatic illness and the duration of the quarantine is five days. However, we find that the cost of this is relatively high, with each person on average getting quarantined twice during a well-controlled epidemic. We try to mitigate this cost by varying the duration of the quarantine. This shows that a quarantine duration of approximately five days is optimal for reducing the peak number of infected. At longer quarantine durations, the effect plateaus [11].

To make this model more applicable to the pandemic situation, we try including a "lockdown" near

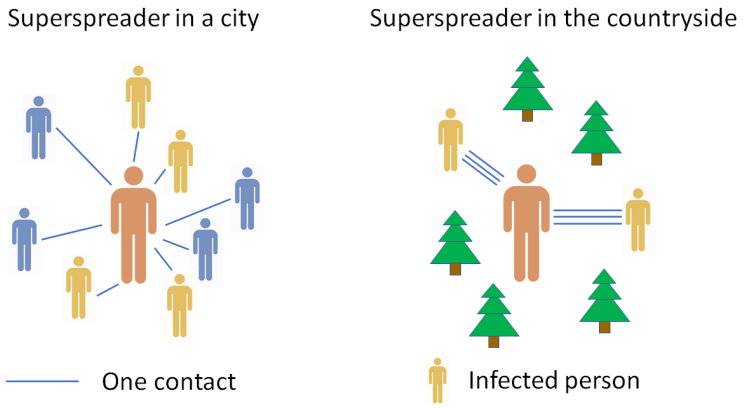


FIGURE 2.3: A diagram of the dynamics of the model for population density and superspreading. A superspreader in the countryside may have many interactions with the same few people and thus almost certainly infects them. However, as there is a limited number of neighbours, this superspreader is not nearly as effective as the one in the city who spends little time with a diverse group of people. The figure was adapted from Ref. [13] and was also used in our poster at the Epidemics8 conference.

the beginning of the simulation where all non-family interactions are reduced by 75 % . Here, a possible fault in the model should be noted, as people are not assumed to then spend more time with their families during lockdown, as would otherwise be realistic. After 30 days, the lockdown is relaxed and replaced by a testing and quarantine system. We find that replacing a lockdown with such a system is a viable strategy, preventing a resurgence of the epidemic after reopening [11].

2.3.2 Superspreading and population density

After Sneppen *et al.* [58] realised the possibly crucial role of superspreading, we wanted to create a model taking this into account. We also wanted to explore the interaction between varying population density and superspreading.

To do this, we model society as consisting of agents on a lattice, instead of the network described above. The behaviour of the system then automatically becomes dependent upon population density. Here, we use a simpler model of the disease progression, sticking to the SIR framework with an exponentially distributed infectious period lasting five days on average. Although agents are stationary, we allow each agent to interact with a number of others for each timestep, drawing the distance to the other agent from a distribution $p(r) = r_0 \mathrm{e}^{-r/r_0}$. This distribution is consistent with data for short distance travel [50]. If the interacting agent encounters an empty site instead of another agent, the interaction attempt is counted as failed and a new attempt is made until the required number of daily interactions is achieved. This is done to make sure that everyone is equally social. People in the countryside spend the same amount of time around others as people in the city, there are just fewer neighbours to choose from [13]. Fig. 2.3 shows a diagram of the action of superspreaders in cities and the countryside in this model.

Most importantly, we draw the infection probability of each infectious agent from a Gamma distribution with some overdispersion parameter k. A small k means that a small number of

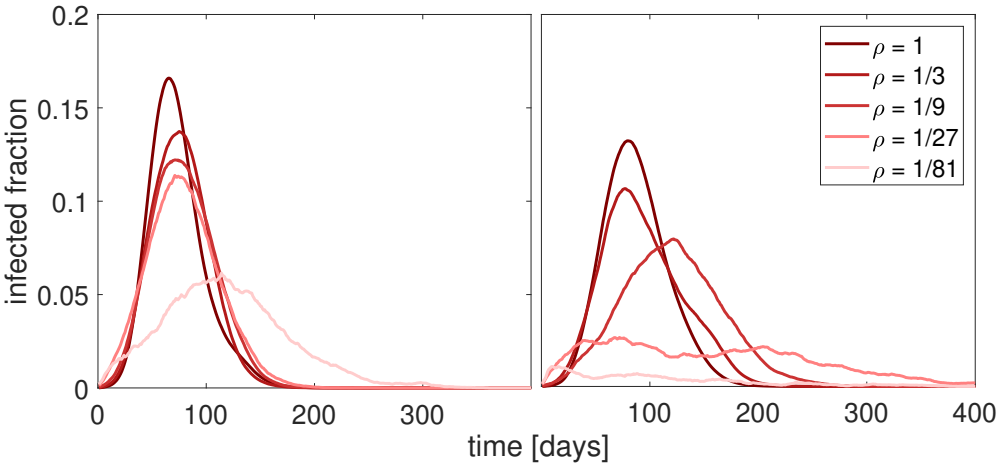


FIGURE 2.4: Incidence of disease over time in connected regions of different population density. (a) shows the case without superspreading, while (b) shows the case with superspreading. We see that when disease spread is dominated by superspreading, the epidemic never really gets started in areas with low population density. Figure from our poster at the Epidemics8 conference. A similar figure is found in Ref. [13].

superspreaders cause most of the infections. For wildtype SARS-CoV-2, k is estimated to be around 0.1, meaning that 10 % of the infected cause approximately 80 % of new infections.

In some iterations of the model, we include a large "city" with a population density of $\rho = 1$ in the middle of the lattice to highlight the role of the city in driving the epidemic. When trying to quantify the difference between areas of different population densities and the importance of disease spillover, we instead split the lattice into five sections of highly different density [13].

Simulating disease spread in an inhomogeneous population on a lattice, we observe that the superspreader-dominated disease will spread very rapidly in densely populated "cities" but much more slowly in the surrounding, sparsely populated "countryside". This leads to an irregular epidemic curve with an early peak. We hypothesise that this might be part of the explanation for the unexpected appearance of COVID-19 epidemic curves, which were also characterised by peaks that long preceded herd immunity. Furthermore, we see that the threshold for disease percolation, that is, when the attack rate suddenly jumps from a few percent to nearly 100 %, depends strongly on the overdispersion parameter k. At $k = 0.1$, the percolation threshold is about five times higher than for $k = 1$ [13], corresponding to the overdispersion of influenza [20].

Plotting the disease incidences for areas of different population densities, we see a much more sensitive dependence on population density for superspreader-dominated diseases than for more homogeneously spreading diseases (see Fig. 2.4). We thus expect COVID-19 and other heavily overdispersed diseases to be mainly city phenomena, at least barring other social factors [13].

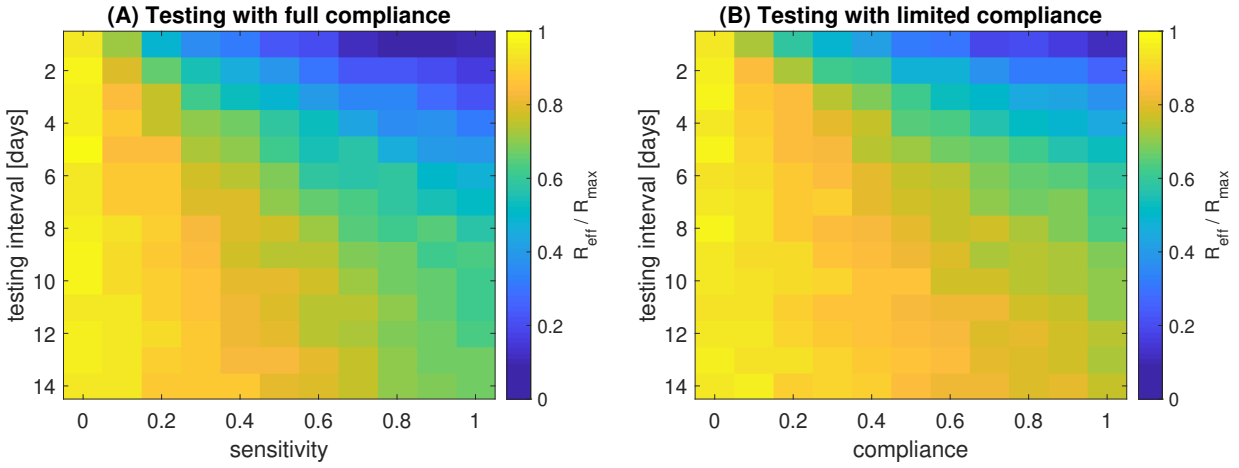


FIGURE 2.5: Heat map of the effectivity of a mass testing strategy. The effectivity is quantified as the ratio between effective reproductive number with testing and maximal reproductive number achieved when the outbreak is unmitigated. (a) shows the value of this ratio given test sensitivity (x-axis) and interval between tests in days (y-axis) assuming full compliance with the mass testing strategy. (b) shows a similar plot, but here the compliance is varied instead of sensitivity. Sensitivity is here assumed to be 70 %. Figure from Ref. [4].

2.3.3 Repeated mass testing model

The repeated mass testing model is a very simple individual-based model created for a report [4] produced for the State Serum Institute. It simply consists of a small lattice of agents interacting with, and possibly infecting, a random neighbour at every timestep. Once every T timesteps, every agent is tested and anyone who is infected is quarantined with a probability p, depending on the sensitivity of the test. We then measure how many secondary infections arise from one index case and use this to predict the effectivity of a repeated mass testing strategy using an imperfect test.

As can be seen in Fig. 2.5, repeated mass testing can reduce the effective reproductive number of COVID-19 quite drastically if done frequently enough. Here, the frequency of the testing matters as much if not more than the sensitivity of the test. We also see that a limited compliance has a similar effect as a reduced test sensitivity. These results point the way to a testing and quarantine strategy where society can be kept open while everyone is tested regularly. A similar type of strategy was implemented in Denmark from spring 2021 [4].

2.3.4 Evolutionary models

To study the evolutionary future of SARS-CoV-2, we create a somewhat different model. This model is not individual-based, but based on a set of age-structured differential equations modified from Li *et al.* [37]:

$$\frac{dS_i}{dt} = a_{i-1}S_{i-1} - a_iS_i - \delta_iS_i + \omega(R_i + \tilde{R}_i) - \sum_j \beta_{ij}S_iI_j - \sum_j \tilde{\beta}_{ij}S_i\tilde{I}_j \tag{2.3.1}$$

$$\frac{dI_i}{dt} = a_{i-1}I_{i-1} + \sum_j \beta_{ij} S_i I_j - (a_i + d_i + \delta_i + \gamma_i) I_i \tag{2.3.2}$$

$$\frac{dR_i}{dt} = a_{i-1}R_{i-1} + \gamma_i I_i - (a_i + d_i + \omega) R_i \tag{2.3.3}$$

$$\frac{d\tilde{I}_i}{dt} = a_{i-1}\tilde{I}_{i-1} + \sum_j \tilde{\beta}_{ij} S_i \tilde{I}_j - (a_i + d_i + \delta_i + \tilde{\gamma}_i) \tilde{I}_i \tag{2.3.4}$$

$$\frac{d\tilde{R}_i}{dt} = a_{i-1}\tilde{R}_{i-1} + \tilde{\gamma}_i \tilde{I}_i - (a_i + d_i + \omega) \tilde{R}_i \tag{2.3.5}$$

[16], where the indices represent the number of the age class, and S, I, and R are susceptible, infected, and recovered populations as usual. a is the aging rate, ω represents the immunity waning rate, β_{ij} is the infection rate from class j to class i, δ_i is the natural death rate in age class i, d_i is the corresponding death rate due to disease, and γ_i is the recovery rate. The variables and parameters with tilde are those of the mutant strain.

If we are to be able to find an ESS, we need to make the fundamental assumption that there is a relationship between disease duration γ and infection rate β. That is, the age-dependent recovery rate γ is treated as an evolutionary parameter such that $\beta = \beta(\gamma)$. We find this to be a reasonable assumption as it should not be possible to have a disease which is both infinitely infectious and very long-lasting, even though this would be ideal from the perspective of the pathogen. There must therefore be some kind of tradeoff between the two variables, forcing the pathogen to balance longevity with infectivity [16]. This is the so-called tradeoff hypothesis [2, 1].

In order to determine the ESS of the pathogen in this model, we find the conditions necessary for a mutant to be unable to invade. This condition is given by the gradient equation

$$\nabla_{\tilde{\gamma}} R_{inv}(\gamma, \tilde{\gamma})|_{\tilde{\gamma}=\gamma} = 0, \tag{2.3.6}$$

where R_{inv} is the effective reproductive number of the invading mutant at the existing equilibrium. That is, the ESS must be a local maximum of the effective reproductive number as a function of γ, or else an invader with a higher reproduction number upon invasion would be able to take over in the future.

Starting from eq. (2.3.6), if we assume that disease dynamics happen at a much faster timescale than aging and natural death, we may reduce the equation to

$$\frac{\partial}{\partial \tilde{\gamma}_l} \frac{\beta_{ll} \hat{S}_l(\gamma_l)}{(a_l + \tilde{\gamma}_l + d_l + \delta_l)}|_{\tilde{\gamma}_l=\gamma_l} = 0, \tag{2.3.7}$$

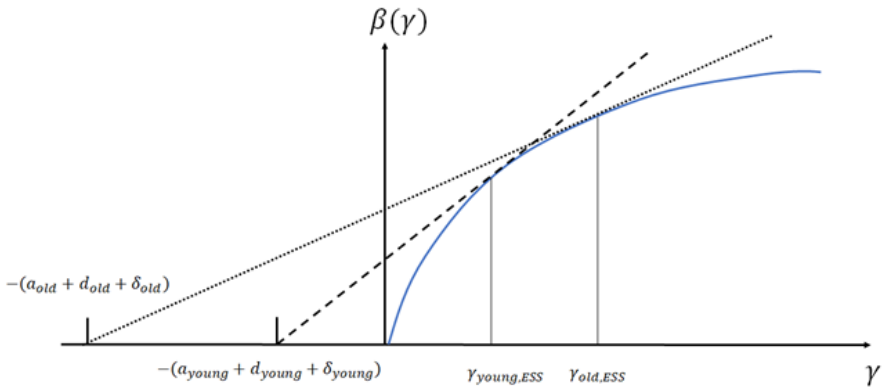


FIGURE 2.6: Finding the ESS when infectivity β depends on recovery rate γ is equivalent to finding the point where a straight line through $\gamma = -(non - disease\ loss\ rate)$ touches the graph of $\beta(\gamma)$ [1]. Our work is a generalisation of this result to multiple age classes. Figure from Ref. [16].

where $\hat{S}_l(\gamma_l)$ is given by

$$\hat{S}_l(\gamma_l) = \frac{\alpha_l S_0}{1 + \sum_{k=2}^{n} \alpha_k}, \tag{2.3.8}$$

and

$$\alpha_i = \frac{\Pi_{k=2}^{i} a_{k-1}}{a_k + \delta_k}. \tag{2.3.9}$$

Carrying out the differential from before, we are left with the equation

$$\beta'_{ll}(\gamma_l)\hat{S}_l(\gamma_l)(a_l + \gamma_l + d_l + \delta_l) = \beta_{ll}(\gamma_l) \tag{2.3.10}$$

[16]. This is equivalent to generalising a well-known result found when studying ESS for parasites [2]. It also shows that the lower the natural death rate is in an age group, the longer the disease may be expected to last in this age group at the ESS (see Fig. 2.6). Further details of the derivation can be found in Ref. [16].

We make use of a similar idea to determine an evolutionarily stable state for the disease duration. The idea here is that a pathogen causing a very long-lasting illness may have time to cause a lot of new infections during the infectious period, i.e., a longer disease duration may lead to a higher reproductive number R_{eff}. However, this slow disease may be outcompeted by a faster variant with a lower R_{eff}, since the faster variant can go through more disease generations in shorter time, growing exponentially with each generation. Furthermore, in a human society a long-lasting disease also increases the risk of the infected person getting quarantined or becoming bedridden at some point during the infectious period, limiting the utility of a long disease duration [14].

The way we approach this problem is, however, very different from the above ESS problem. We construct an agent-based model letting the simulated virus freely mutate and observing which

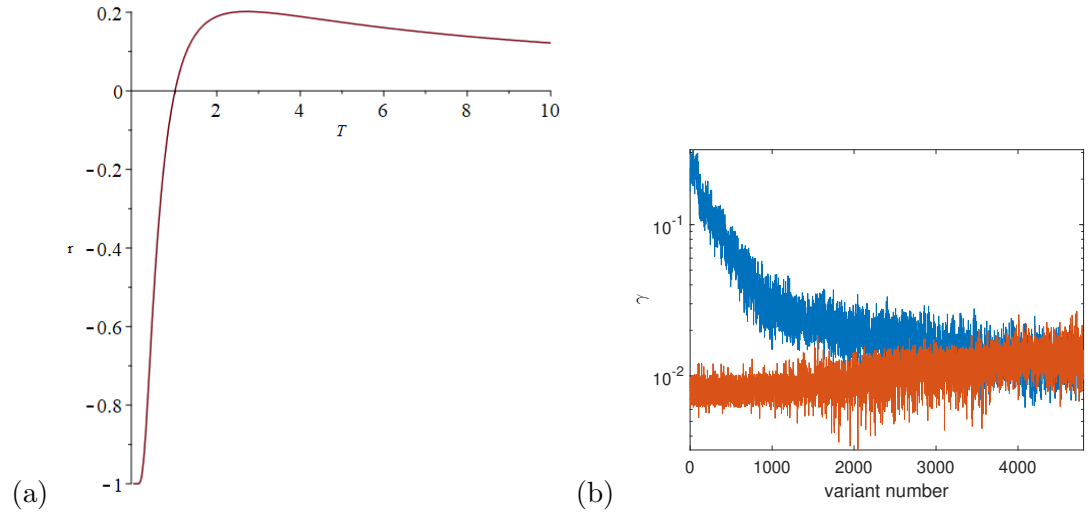


FIGURE 2.7: (a) The exponential growth rate of the disease as a function of its duration for $\beta = 1, c = 1$. We see that this function has an optimum for $T = \mathrm{e}$. (b) The recovery rate γ of successive variants in a simulation where we let symptomatic individuals be quarantined with some probability. We see that whether we start from a high or a low value, they approach the same optimum. $\beta = 0.5$, $p_{quarantine} = 0.1$. Figures from Ref. [13].

variant(s) end up dominating.

From our simulations of the evolution of disease duration, we find that at the endemic state, a disease which infects at a constant rate per day will always gain an advantage by evolving a longer-lasting infectious period. This changes if there is some probability of being quarantined or beginning self-isolation for each day of symptomatic illness. In this case, there will be an optimum disease duration [14].

If instead of the endemic state we consider a state where a disease is spreading exponentially, such as early in an epidemic or when containment measures fail, the picture is different. Here, we see that shorter-lasting diseases will sometimes win out over longer-lasting ones, simply because their exponential growth rate in number of cases is higher. We derive the following expression for the exponential growth rate of a disease in terms of the duration of the infectious period:

$$r = (\beta T)^{\frac{1}{(1+c)T}} - 1, \tag{2.3.11}$$

where r is the daily exponential growth rate of the number of cases, β is the daily infection rate, T is the duration of the infectious period, and c is the proportionality between this and the duration of the incubation period, assumed to be constant. This function has a local maximum for $T = \mathrm{e}/\beta$. Thus, we conclude that, although there is no optimal disease duration at an endemic equilibrium of an unmitigated disease, the duration of the infectious period will have an optimum in the cases where (a) symptomatic individuals can be quarantined or (b) the epidemic is exponentially growing [14].

Sometimes, studying stable states is not necessary or desirable if we are to gain an overview of the

"direction" of evolution. In the final model presented here, as shown in the attached paper [48], we again let new variants arise randomly in agents on a network. Here, the variable evolutionary parameter is the overdispersion k. We then simply quantify how k affects the chance of survival of new variants, and how k is expected to evolve over time under given constraints, such as a lockdown which reduces the number of contacts of each agent. By using this method, we do not have to look for a final, stable state. Instead, we can examine how k evolves on its own. This method is somewhat more mathematically simple and still gives an idea of where new variants of SARS-CoV-2 are heading.

Simulating disease spread on an Erdős-Renyi network where overdispersion is the evolutionary parameter, we find that over time, there will be a tendency for variants to evolve towards a higher k, i.e., less prominent superspreading. This is most notable in environments where the diversity of contacts is limited, i.e., during a lockdown that restricts large gatherings. Without a lockdown, there is limited benefit to spreading more homogeneously, leading to a slower evolution towards homogeneous spreading. If we reduce disease infectivity without limiting the number of individual contacts per person, we also see a much slower evolution towards homogeneity, demonstrating that it is the reduction in connectivity that makes homogeneity so evolutionarily advantageous for new variants.

At the same time, we find that an emerging variant with a small k will be more vulnerable to extinction due to stochastic fluctuations than a more homogeneously spreading one. This makes sense when one considers that when a highly overdispersed new variant emerges, the first few that are infected with it are unlikely to be superspreaders. If none of them spreads the new variant effectively, it will go extinct. Thus, a high overdispersion is evolutionarily disadvantageous for a new variant in two ways [48].

Summary

To describe the dynamics of COVID-19 we design and explore the dynamics of the following models:

- An individual-based model incorporating a network with clusters around families, workplaces, and friend groups. This model allows for testing and contact tracing. Using this model, we show that mass testing and quarantine can be effective at mitigating the pandemic, even at low testing rates and for relatively short quarantine durations.
- A geographical model of population density and superspreading. We see that if a disease is dominated by superspreading, it means that this disease becomes very effective at spreading in densely populated areas, but much less effective at spreading in sparsely populated areas.
- A mean field, but still individual-based model of a mass-testing strategy with an imperfect test. Testing everyone with a test of low sensitivity is a viable strategy for controlling a COVID-19 outbreak, as long as people are tested frequently enough.
- An ODE-based model of the evolution of infectivity and disease duration in an age-structured

population. At the evolutionarily stable state, we expect a pathogen like SARS-CoV-2 spreading in an age-structured population to evolve a disease duration that depends on age, such that the disease lasts a long time, but has a low infection rate per day in groups with a low natural death rate. On the other hand, it should evolve towards a high daily infection rate and a short duration in groups with a high natural death rate.

- Individual-based models of the evolution of disease overdispersion and the duration of the infectious and incubation periods. These models show that
  - Evolution will generally favour a less overdispersed disease with a higher k-value if the diversity of contacts between individuals is limited, for example by a lockdown.
  - At the endemic state, there is no optimal duration of the infectious period. However, if there is a probability of being quarantined for each day of symptomatic illness or if the epidemic is in an exponentially growing phase, there is an optimum.

2.4 Discussion

In this part of the thesis, we have shown a few examples of cases where individual differences in network and propensity for superspreading matters. These cases are best modelled using individual-based models. We have also shown that in order to accurately capture the long-term dynamics of a pandemic, evolution has to be taken into account. Both the infectivity, recovery rate, and overdispersion of the disease are subject to evolutionary pressures. Furthermore, the structure of human society and the restrictions imposed by pandemic mitigation efforts shape these evolutionary pressures.

The models presented here are highly idealised and thus only provide a rough image of the magnitude of the various effects. The network used in our testing and contact tracing model assumes a highly simplified network with a small number of very regular clusters centered around families, workplaces etc. We also assume a relatively small family size, as is found in Danish society. We tried to run the model with larger family sizes, which appeared to not make much of a difference [11]. Furthermore, the fact that lockdown and quarantine in this model are assumed to simply reduce the overall number of social interactions people have in a day is somewhat unrealistic, as less time spent in public should result in more time spent with the family. A final caveat to note is that we, crucially, neglect overdispersion in this model, as its importance had not yet come to our attention when the model was devised. Had we included it, the relative effect of lockdown as compared with testing and quarantine might have been greater.

With respect to the second model, it was created at a time in the pandemic when only one epidemic wave had occurred in the west, and COVID-19 was still largely a city phenomenon. This turned out to not be a firm rule for the disease, with large waves being seen in rural parts of the US in autumn 2020. This illustrates that social and climatic effects cannot be ignored [13]. The

superspreader phenomenon may also have changed with the emergence of new variants such as Alpha (B.1.1.7) and Delta (B.1.617.2). These variants are suspected of having more homogeneous infectivity profiles, as the distribution of viral loads in those infected is less skewed than for wildtype SARS-CoV-2 [30, 48]. Nonetheless, overdispersion is not unique to SARS-CoV-2 [39], so the results achieved here should still have wide applicability.

There is some doubt about the tradeoff hypothesis on which we build our ESS model in an age-structured population, and particularly whether it may be expressed as a simple dependence $\beta = \beta(\gamma)$ [1]. Despite this, we chose to make this assumption anyway, as some form of tradeoff is necessary for an ESS to exist. In the case of disease duration, such a tradeoff arises naturally. The model of overdispersion evolution here has the advantage that its purpose is not to look for an ESS, and therefore, there is no need to assume a tradeoff. Instead, we can simply observe the direction of evolution of k.

Based on this work, we have gained a significantly better understanding of the dynamics of COVID-19 and epidemic diseases in general. We have realised the great potential effect of both superspreading and evolution during a long-lasting pandemic. In response to this, we have created models to examine these effects. Especially the topic of evolution is quickly gaining in importance as population immunity grows and the COVID-19 pandemic thus transitions from an initial epidemic state to a longer-term endemic state. During the pandemic many models have been created trying to describe the physics of disease spread in a social system. This type of modelling will still be highly important as we enter the pandemic endgame.

Summary

- One of the main points of this part of the thesis has been the development of disease models that are able to take individual differences and social structures into account. We have shown how both these factors may shape the course of a pandemic.
- It is important to keep in mind the simplifications made here. For example, the network of our testing and quarantine model is highly simplified, the geographical model neglects social effects, and the tradeoff hypothesis of our age-structured evolutionary model is controversial.
- The pandemic has highlighted the utility of physics-inspired modelling of disease spread. This will remain important in understanding the behaviour of COVID-19 on a longer timescale.

Conclusion and perspectives

In this thesis, we have conducted an investigation into the population dynamics of a variety of different systems. In particular, we have focused on the action of diseases on animal and human populations. Infectious diseases change the population dynamics of different animal species, and they affect and are themselves affected by the social structures of human societies.

We have seen that infectious diseases may shape the evolution of animals by providing prey species with a potential weapon against their predators. Even if a predator species is not directly affected by a disease, epizootics in its preferred prey may still cause its population to collapse, and potentially even extinction. Incidentally, we discovered the utility of even very simple mass-parameterisation methods in estimating the parameter values of population dynamic equations. As a similar spinoff from our main work on diseases and evolution, we discovered that a disease outbreak in one prey species may cause chaotic behaviour in ecosystems with multiple prey species.

In response to the COVID-19 pandemic we have shown that to properly model a disease spreading in a complex human society, it is not enough to know the number of infected and their interaction rate with the rest of society. Instead, the most effective models will also take the structure of networks and the range of individual biological characteristics and behaviours into account. This has the added benefit of allowing for easy modelling of network-specific mitigation strategies such as contact tracing.

Building on this, our work on superspreading has shown that to limit the spread of a superspreader-dominated disease such as COVID-19 (at least as caused by the wildtype SARS-CoV-2 variant), it is not as important to reduce the total amount of social interaction as it is to reduce the diversity of contacts.

Finally, our studies on pathogen evolution show us that the structure of society and the mitigation measures we impose to control the pandemic are important in shaping the future of SARS-CoV-2 and thus of the pandemic. Specifically, we show that a pathogen in an age-structured population may evolve an age-dependent infectivity profile and duration. Furthermore, pathogens are likely to generally evolve towards a more homogeneous infectivity with a smaller tendency towards superspreading. The latter is particularly true when lockdowns restrict the diversity of social contacts. Quarantine measures may also affect disease duration by limiting the evolutionary utility

of a very long infectious period for a pathogen.

The two topics studied under this project are connected. First of all they both fall under the broad headline of disease dynamics. Perhaps more importantly, since pathogens as mentioned above are frequently shared between species, epizootics in animals may result in a pandemic among humans. As we have seen during the pandemic, a disease is capable of jumping back and forth between humans and other animal hosts. During this process, it may pick up new evolutionary traits by adapting first to one host and then the next. Such multi-host evolutionary dynamics could be an interesting and highly relevant topic for future research. Generally, the role of pathogens in species and ecosystem evolution still contains many unanswered questions which deserve to be explored further.

The individual-based approach that turned out to be enlightening for the study of human epidemics might also prove useful when studying animal communities, where our ecological models are based on the assumption of mean-field dynamics. Individual animals may be subject to many of the same biological and behavioural variations as humans, meaning that the individual-based approach could shed further light on the dynamics of trophic interactions, evolution, and disease dynamics. One obstacle to this is the fact that limited data exist on epizootics in wild animals, which has also been one reason why the present ecological analysis remains largely theoretical. With the current focus on epidemics and consequently on zoonoses, it is to be hoped that more research into this field will be done in the future.

In conclusion, our work has expanded our knowledge of the wide range of dynamics that arise in the interplay between human and animal populations and a disease. The pandemic has made it very clear that the mathematical study of population and disease dynamics is very much relevant for human society. Understanding how animals and pathogens interact with each other and with humans, and how these populations evolve over time, is therefore vital if we are to understand current and future epidemics.

Articles for chapter 1

The following articles form the basis of the chapter on ecological disease models. Articles 1 and 2 are based partly on preliminary work done during my master's project. Therefore, there may be some overlap between chapter 1 and my master's thesis, Ref. [9]. An earlier draft of article 1 was appended to the master's thesis. Article 3 is based entirely on work done during the PhD, though the PPP equations of this article are derived from the same SIR/Lotka-Volterra framework as those of the master's thesis. All articles were completed and published during the PhD.

1. Eilersen, A. and Sneppen, K. (2019). Applying allometric scaling to predator-prey systems, *Physical Review E* **99** (2): 022405.
2. Eilersen, A. and Sneppen, K. (2020). The uneasy coexistence of predators and pathogens, *The European Physical Journal E* **43** (7): 1–7.
3. Eilersen, A., Jensen, M. H. and Sneppen, K. (2020). Chaos in disease outbreaks among prey, *Scientific Reports* **10** (1): 1–7.

Applying allometric scaling to predator-prey systems

Authors: Andreas Eilersen* and Kim Sneppen*

*The Niels Bohr Institute, University of Copenhagen, Copenhagen, Denmark

My contributions: I contributed to the development of the model, to the associated computational work, figure production, and to the writing of the article.

Publication status: Published in *Physical Review E* (2019).

doi: 10.1103/PhysRevE.99.022405

Notes on relation to master's project: The mass parameterisation model and associated equations were developed during my master's project [9], and as previously mentioned a draft version of this article was appended to the master's thesis. Nonetheless the article is included here, as it was completed during the PhD with significant additions to the data section being made after the completion of the master's degree. Furthermore, the work done here forms part of the basis of the rest of the ecological work presented in chapter 1. Therefore, it is my impression that the chapter would be less complete without this article.

Applying allometric scaling to predator-prey systems

Andreas Eilersen* and Kim Sneppen†
Niels Bohr Institute, University of Copenhagen, Blegdamsvej 17, 2100 København Ø, Denmark

(Received 6 October 2018; revised manuscript received 10 January 2019; published 8 February 2019)

In population dynamics, mathematical models often contain too many parameters to be easily testable. A way to reliably estimate parameters for a broad range of systems would help us obtain clearer predictions from theory. In this paper, we examine how the allometric scaling of a number of biological quantities with animal mass may be useful to parameterize population dynamical models. Using this allometric scaling, we make predictions about the ratio of prey to predators in real ecosystems, and we attempt to estimate the length of animal population cycles as a function of mass. Our analytical and numerical results turn out to compare reasonably to data from a number of ecosystems. This paves the way for a wider usage of allometric scaling to simplify mathematical models in population dynamics and make testable predictions.

DOI: 10.1103/PhysRevE.99.022405

I. INTRODUCTION

When modeling the dynamics of ecological communities, a recurring problem is the difficulty of estimating model parameters. If we desire to develop a model that can describe real ecosystems, a common approach is to add terms and parameters to account for as many real-world complications as possible. The result is, unfortunately, that many of the models end up being too complicated to actually make any definitive predictions due to uncertainties about the often large number of parameters. A model that requires precise measurements of parameters for every individual system one wishes to study will of course be interesting for the isolated case, but it will be difficult to derive more general principles from it. We believe that a simplified model that makes approximate but clear predictions might be a more useful approach. In this paper, we will argue that, by using allometric mass scaling, it is possible to estimate the parameters of the classic Lotka-Volterra predator-prey equations in such a way that this highly idealized model can be used to predict the behavior of actual populations. It is our hope that we will be able to rewrite all parameters of the equations in terms of only two quantities: prey mass and predator mass. We will also look at the implications of body size for the period of animal population cycles. By doing so, we wish to conclusively demonstrate the usefulness of allometric mass scaling relations in population modeling.

The fact that many ecological variables scale allometrically with animal body mass has attracted increasing attention in recent years. Ginzburg and Colyvan [1] go as far as to call the allometries fundamental laws of ecology, comparing them to Kepler's laws in physics. Peters [2] compiled a list of variables exhibiting allometric scaling, which we will make use of in this paper. For example, generation time and metabolic rate correlate with mass to powers of (approximately) 1/4 and 3/4, respectively. It is these relationships that we will exploit to write the Lotka-Volterra equations in terms of animal body mass. For a compelling attempt at finding a theoretical foundation for these quarter-power scaling laws, see the work of West *et al.* [3].

On the larger ecosystem scale, there are also examples of allometric scaling. In particular, many animals—most prominently rodents such as lemmings—exhibit a regular population cycle. The time elapsed between peaks in abundance of such animals tends to scale with the average mass of the animal. Empirically, the scaling relation is found to be $T \propto m^{0.26}$ [4]. We wish to argue for a theoretical basis of this relationship.

Yodzis and Innes [5] use mass to parameterize a system of equations similar to generalized Lotka-Volterra equations, with consumer and resource (whether plant or animal) substituted for predator and prey. Their model assumes that the predator reproduction will saturate with increasing prey population, giving the predator a Holling type II or type III functional response. Also, they argue that the strength of the predator-prey interaction should scale with the ratio of prey mass to predator mass to some power and that it should be possible to determine the coefficients of this scaling law from measurable biological quantities. With the model in place, they analyze the linear stability of the dynamical system and find that, for certain predator-prey mass ratios, it will have a limit cycle with a period $T \propto m_C^{1/8} m_R^{1/8}$, where m_C is the consumer (predator) mass and m_R is the resource (prey) mass.

We will here proceed down a similar path, though our model will be notably simplified and our approach to the predator-prey functional response will be different. The original Lotka-Volterra equations on which we will be basing our model assume that the predation and predator reproduction rates increase in proportion with prey population density, a so-called Holling type I functional response. We here assume that prey population is always far from the carrying capacity of the ecosystem, resulting in a prey reproduction rate that is also proportional to prey population. Instead of trying to determine a biologically reasonable coefficient for the scaling

*andreaseilersen@nbi.ku.dk
†sneppen@nbi.ku.dk

of interaction strength with the predator-prey mass ratio, we will let the coefficient remain unknown. We will determine the equilibrium populations in terms of this unknown coefficient. Luckily, it turns out that when we look at the ratio of the populations, this coefficient cancels out. Thus, our method still yields useful information.

Finally, we will look at the period of population cycles. The simplest version of the Lotka-Volterra equations has a nontrivial equilibrium which is a center rather than a limit cycle. Here we likewise find a period of $T \propto m_C^{1/8} m_R^{1/8}$ as mentioned above. In order to obtain the empirically determined $m^{1/4}$ relationship with population cycle length, Yodzis and Innes point out that one can assume a direct proportionality between predator (consumer) size and prey (resource) size. While this relationship may hold in many systems (see, e.g., Ref. [6]), it certainly does not in such cases as the wolf-moose system studied by Peterson *et al.* [4], and the relationship is hardly well defined in systems where the resource is a plant. Ginzburg and Colyvan [1] even present a critique of the whole idea of using only the linearization of the Lotka-Volterra equations to predict the length of population cycles. It would therefore be preferable if we could derive a relationship between prey mass and cycle period that is independent of predator mass. This is what we will attempt to do in the following.

The model put forward here is thus an application of the basic idea of Yodzis and Innes to a heavily simplified system of equations, without making attempts at determining the exact interaction strength between predator and prey directly. It is our hypothesis that even such a simplified model will still give reasonable order-of-magnitude predictions about real ecosystems.

II. PARAMETERIZING THE LOTKA-VOLTERRA EQUATIONS

The original Lotka-Volterra predator-prey equations read as follows [7]:

$$\frac{dx}{dt} = \alpha x - \beta xy, \tag{1}$$

$$\frac{dy}{dt} = \gamma xy - \delta y. \tag{2}$$

Here x denotes prey, y predator, α the per-capita reproduction rate of prey, and δ the per-capita death rate of predators in the absence of prey. The interaction strengths β and γ are slightly harder to define. β denotes the risk of each prey being eaten per predator, and γ represents the increase in predator reproduction rate per prey. These latter two parameters are of course more difficult to estimate than the first two, and we will therefore need to find a way around this obstacle.

As opposed to Yodzis and Innes, we choose to work with animal abundances rather than biomass densities. We do this because it is conceptually easier and data are more readily available for abundances than for biomass densities for the systems that we wish to study. A complication arising from this is that when working with abundances, there is a distinction between somatic growth (individuals growing larger) and reproductive growth, which would be unimportant if we were to work with biomass densities. We shall therefore ignore the finer details of animal reproduction and growth and simply assume that all growth results in the production of new individuals. Furthermore, we assume that the populations are large enough and reproductive events evenly distributed enough in time that population growth can be modelled as continuous rather than discrete.

According to Peters [2], we then have the following empirical relation for reproduction rate:

$$\alpha = \tfrac{1}{400} m_x^{-1/4} \quad (\text{day}^{-1}). \tag{3}$$

In the cited mass scaling relations, all masses are in kilograms. As the predator-prey pairs we will be examining here are all mammals, we shall be using the mass scaling relations that apply to mammals. For cold-blooded animals such as reptiles the relations will be different, though not radically so.

It should be possible to calculate the death rate of predators in the absence of prey from the so-called turnover time. This is defined as the average time it will take an animal to metabolize its entire energy reserves. In turn, this can be calculated from the metabolic effect. Again from [2]

$$t_{\text{turnover}} = 19 m_y^{1/4} \quad (\text{day}). \tag{4}$$

This implies

$$\delta = t_{\text{turnover}}^{-1} = \tfrac{1}{19} m_y^{-1/4} \quad (\text{day}^{-1}). \tag{5}$$

The coupling coefficient β we assume to be proportional to predator ingestion rate. We believe this to be justified, since the more a given predator consumes, the higher the per-capita risk of being eaten by it should be for the prey. The predator ingestion rate in terms of energy scales with mass as [2]

$$I \propto m_y^{3/4} \quad [\text{J (day predator)}^{-1}]. \tag{6}$$

The number of individual prey that a predator needs to eat to satisfy this energetic demand is inversely proportional to prey mass, and we therefore write β as

$$\beta = k \frac{m_y^{3/4}}{m_x} \quad [(\text{day predator})^{-1}], \tag{7}$$

where k is an unknown proportionality constant. Knowing the equilibrium population of prey or predator should make it possible to determine k if this is desired.

Our parametrization thus deviates notably from that of Yodzis and Innes, since they assume that the predator death rate and the interaction strength scale with the ratio of prey mass to predator mass to the power of 3/4 (here converted to abundance rather than biomass, as was originally used). Strictly speaking, the ingestion rate of y predators reflects some kind of average prey consumption rate at average prey abundance. What we really need here is the slope of predator kill rate as a function of prey abundance. Furthermore, the units of the ingestion rate is $[\text{J (predator day)}^{-1}] \propto [\text{prey (predator day)}^{-1}]$ and not $[(\text{predator day})^{-1}]$ as we need it to be for our units to match. Despite all this, we still believe that the allometric scaling of the ingestion rate is a reasonable approximate measure of the predator's ability to consume and therefore of the dependence of consumption rate on prey abundance. We now only need to find a way around not knowing the exact proportionality.

The slope of predator kill rate with prey abundance that really constitutes β depends on a number of factors (temperature, prey population density, predator satiation, etc. [2]), and it is probably not possible to make a universal estimate of it. Instead, we let k embody all these complications and tune it to fit the systems that we will study. As mentioned above, it fortunately cancels out in the final calculation of the prey to predator population ratio anyway.

The relation between the number of prey eaten and the number of predators produced can be derived approximately if we know the ecological efficiency η of the predator-prey interaction. The ecological efficiency here refers to the percentage of prey biomass that is converted into predator biomass. Ecological efficiencies vary considerably depending on the nature of the interaction [8], and it is therefore difficult to find an estimate that is both precise and general. For systems with a low predator-to-prey mass ratio and positive correlation of biomass density with body mass, ecological efficiency should be high ($\eta \approx 35\%$) according to a review by Trebilco *et al.* [9], which, however, deals with aquatic ecosystems. Lindeman's original paper similarly shows an efficiency that rises with trophic level [8]. On the other hand, a case study of the Isle Royale wolf-moose system that we will discuss below suggests that the wolves have a much lower efficiency than we would expect based on the above ($\eta \approx 2\%$) [10]. In laboratory experiments, a figure of about $\eta = 10\%$ is observed [10], and for lack of a better estimate, we shall use this so-called 10% law in our calculations. Given that we are not going for an exact description of any one particular interaction, we believe that it is justified to use this rough estimate.

The relation between mass of consumed prey ($m_{x,c}$) and mass of produced predator ($m_{y,p}$) is now

$$m_{y,p} = \eta m_{x,c}, \tag{8}$$

assuming that prey and predator have similar energy content per unit mass. Rewriting this in terms of numbers of individual predators produced ($N_{y,p}$) and prey consumed ($N_{x,c}$), we get

$$N_{y,p} = m_{y,p}/m_y = \frac{m_{x,c}}{m_y}\eta = \frac{m_x N_{x,c}}{m_y}\eta. \tag{9}$$

In the Lotka-Volterra equations, the number of predators produced per unit time is given by the term

$$N_{y,p} = \gamma xy \tag{10}$$

and the number of prey consumed by the term

$$N_{x,c} = \beta xy. \tag{11}$$

Thus, we get the following relation between β and γ:

$$\gamma = \frac{m_x}{m_y}\eta\beta = k m_y^{-1/4}\eta \quad [(\text{day prey})^{-1}]. \tag{12}$$

We have now written all the parameters of the equations in terms of the animal body masses alone, with k from Eq. (7) being the only parameter that remains to be determined. However, we can get around this by focusing our attention on the equilibrium predator-to-prey population ratio instead of the absolute populations.

The Lotka-Volterra equations have the nontrivial equilibrium

$$(x, y) = \left(\frac{\delta}{\gamma}\ ,\ \frac{\alpha}{\beta}\right), \tag{13}$$

which is neutrally stable. The equilibrium ratio between prey and predator populations is therefore

$$x/y = \frac{\beta\delta}{\alpha\gamma} = \frac{21}{\eta}\left(\frac{m_y}{m_x}\right)^{3/4}. \tag{14}$$

This number depends only on the masses. We see that due to the factor $1/\gamma$ this ratio is inversely proportional to ecological efficiency, so that if our estimated 10% efficiency is a factor 2 too great, then we will estimate a ratio that is half the "correct" value.

III. THE PERIOD OF POPULATION CYCLES

The Jacobian matrix of the Lotka-Volterra equations at the nontrivial steady state has the eigenvalues $(i\sqrt{\alpha\delta}, -i\sqrt{\alpha\delta})$, meaning that for small perturbations away from equilibrium, the system will oscillate over time with a period of $T = \frac{2\pi}{\sqrt{\alpha\delta}}$. This leads to the aforementioned scaling of population cycle period with mass $T \propto m_x^{1/8} m_y^{1/8}$, contrary to the observed $T \propto m_x^{1/4}$. A problem with using linearization in this case is that the period thus obtained only applies when oscillations are relatively small. Population cycles in actual predator-prey pairs, such as the vole-weasel pair in northern Scandinavia, can involve fluctuations over two orders of magnitude [11]. When solving the equations numerically, we see that much of the time, the population of prey will be in a state of slow, exponential recovery, while the predator population slowly approaches zero. When the prey population recovers, the predator population quickly explodes, initiating a swift collapse of the prey population. The collapse phase observed in real rodent cycles does indeed appear to be notably shorter than the growth and peak phases, and the corresponding predator cycles are similarly observed to be very sharply peaked [11,12]. We therefore believe that the dynamics can be realistically modelled as consisting of a slow exponential growth phase and a fast collapse phase. Using this two-timescale assumption, we will try to derive an expression for the period T of population cycles. Splitting more complex predator-prey models into slow and fast phases has previously been done by Rinaldi and Muratori [13]. In the following, we shall use a similar basic idea but a different mathematical approach and solve for the period T rather than maximal abundance as they did. An illustration of the cycle and its fast and slow segments can be seen in Fig. 1. For our derivation, we will use the maximum and minimum prey density of a cycle, which should be easily obtainable from observations and available in the literature.

The slow approach to and subsequent drifting away from the saddle point at (0,0) is what takes up the majority of the orbital period of the system. For this reason, we will here attempt to derive an approximate relation for the cycle length by looking at the behavior around the saddle point at (0,0) instead of the center at $(\frac{\delta}{\gamma}\ ,\ \frac{\alpha}{\beta})$. Although the period of the cycle is mainly determined by the hyperbolic approach to the

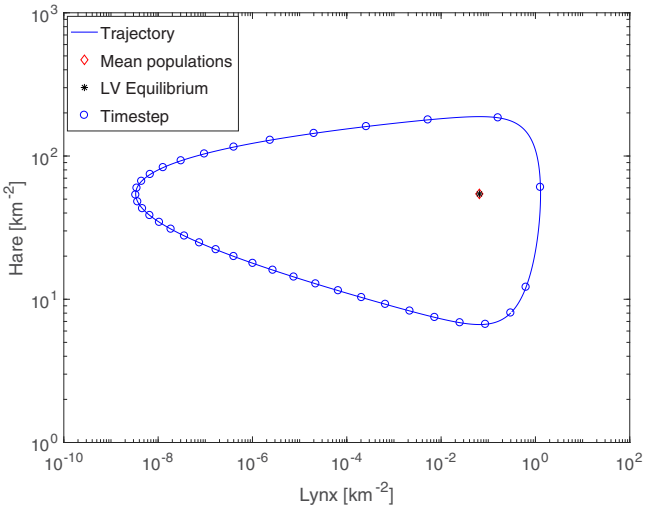


FIG. 1. An illustration of the dynamics of the predator-prey system. This numerical solution is based on parameters appropriate for the lynx-hare system discussed below. The line shows the trajectory of the system in predator-prey space, and the circles are all spaced evenly in time at a separation of 50 days. The distinction between a fast and a slow segment of the trajectory can be clearly seen from the spacing of the circles. Note also that equilibrium abundances are practically identical to mean abundances, meaning that we can use the two interchangeably.

saddle point, the oscillation still happens around the center equilibrium at $(\frac{\delta}{\gamma}, \frac{\alpha}{\beta})$. As can be seen in Fig. 1, the time average populations are very close to the equilibrium populations at the center. We therefore do not believe that there is a contradiction between using the saddle point linearization to determine the oscillation period but determining population ratios based on center equilibrium values.

Using the linearization around the center equilibrium, we obtain a period that is independent of initial conditions but which does not match observations, as the assumption that initial conditions are close to the equilibrium breaks down in the real systems studied here. Instead, we assume that the initial conditions are far from the center equilibrium. For this asymptotic approximation, the period will depend on initial conditions and the calculated period matches observations better.

Starting from a population $x_{\min}$, the prey population should grow as follows:

$$x(t) = x_{\min} e^{\alpha t}. \tag{15}$$

When predator population is low, prey population grows unobstructed. After one period of length T, we should have the maximal population density

$$x_{\max} = x(T) = x_{\min} e^{\alpha T}. \tag{16}$$

The time it will take the population to recover to a density of $x_{\max}$ now becomes

$$T = \frac{1}{\alpha} \ln\left(\frac{x_{\max}}{x_{\min}}\right) = 400 \ln\left(\frac{x_{\max}}{x_{\min}}\right) m_x^{1/4} \quad \text{(day)}. \tag{17}$$

We thus get the $m_x^{1/4}$ relation found empirically. The above expression should be valid when the amplitude of oscillations is very large, so that the period of the predator-prey cycle is dominated by the slow growth phase which in the Lotka-Volterra model occurs at low predator abundances. Note, however, that we at no point have assumed that the population crash should be due to the influence of a predator. We just assumed that the crash was fast and did not extend the period length or influence the exponential growth phase significantly. The derivation here should therefore be equally valid if a population crash is caused by, e.g., a shortage of food or an epidemic. Given that in the case of many rodents it is unclear if it is actually predation that drives the cycle [15], this is a significant advantage.

Another interesting feature of this expression is the logarithmic scaling with population maximum-minimum ratio. Hanski has already hinted at such a scaling relation for the vole-weasel system [16]. In his 1991 paper, he shows that $\ln(\frac{x_{\max}}{x_{\min}})$ correlates with latitude and that oscillation period also correlates with latitude. Oscillation period thus also correlates with the logarithm of the maximum-minimum ratio. It is possible that we have found a theoretical explanation for this correlation.

In the next section, we will demonstrate that Eq. (17) roughly fits the pattern seen in oscillating populations in nature, although there is a significant deviation between predicted and observed numbers. For the prey-predator ratios, on the other hand, the parameters derived above mostly give realistic results.

IV. COMPARING THEORY WITH DATA

The classic example of a system described well by the Lotka-Volterra equations is the interaction between the Canadian lynx (*Lynx canadensis*) and the snowshoe hare (*Lepus americanus*). Although there has been some doubt as to whether the hare population cycle is driven primarily by predation or other factors, there seems to be evidence that changes in hare mortality are mainly due to predation [23]. The population density of hares oscillates from around 8 to just under 200 per square kilometer over the 8- to 10-year-long cycle [24]. The average density of lynx ranges from 0.03 to 0.3 km^{-2} [25].

To see how well our model fits with observations, we plug the average masses of lynx, on average roughly 11 kg [20], and hares, roughly 1.6 kg [26], into the equations and solve them numerically. We choose initial conditions corresponding to the density per square kilometer when hare abundance is lowest ($x_0 \approx x_{\min} = 8$ and $y_0 = 0.3$, due to the phase difference between lynx and hare population oscillations, we let lynx population start out high and hare population start out low). We then tune the parameter k to obtain the correct ratio between cycle highs and lows. The result can be seen in Fig. 2(a). Our simulation predicts an average prey to average predator population ratio that is quite close to the observed values. The period is off by about a third, which, given the simplifications of the model, is not a bad estimate. The fact that the population collapse takes such a short time in our simulation contributes to our underestimating the period. In reality, the collapse takes about 1–2 years [24]. The spiky appearance of the graph is also not very naturalistic. However, taking increasing predation from other predators, increasing

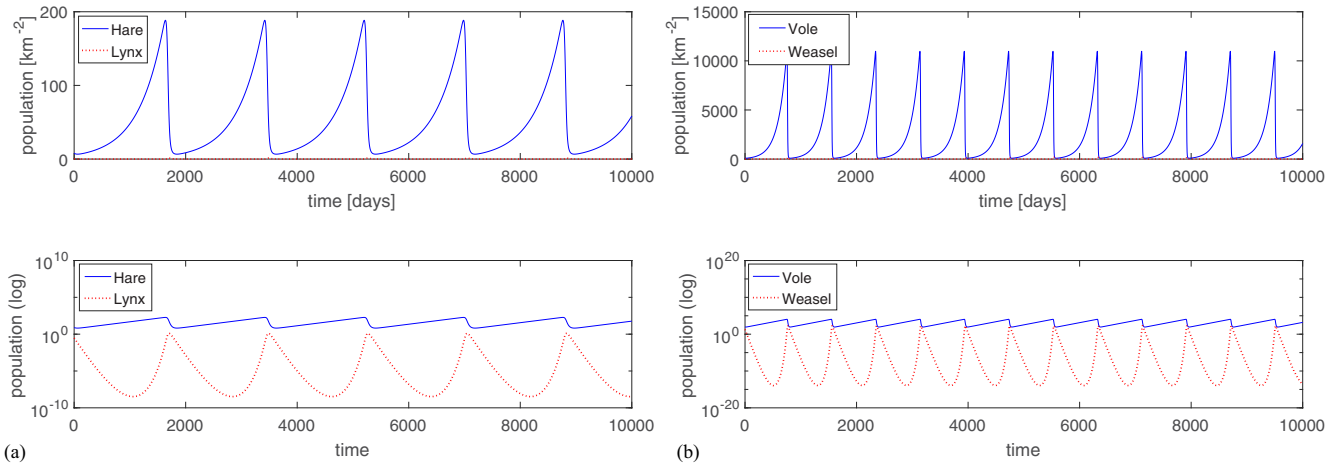


FIG. 2. (a) A numerical simulation of the Lotka-Volterra equations for lynx and hare; $x_{\max} \approx 180$ km^{-2}, $x_0 = x_{\min} \approx 8$ km^{-2}, $y_0 = 0.3$, and $k = 1.05 \times 10^{-2}$. The period is just under 2000 days, or 5.5 years, and the average hare density is 51 km^{-2}. Average lynx density is 0.059 km^{-2}. The ratio of the averages is 860 hares per lynx. As can be seen from the logarithmic plot, the predicted predator oscillations are too violent, with extinction of lynx at the cycle minimum. When this does not actually happen, it may be due to the fact that lynx can survive partially on other prey when hare population is low [14]. (b) The solution obtained using the masses of voles and weasels. $x_{\max} \approx 10^4$ km^{-2}, $x_0 = x_{\min} \approx 10^2$ km^{-2}, $y_0 = 20$, and $k = 2 \times 10^{-4}$. We still see a cycle somewhat shorter than the observed, with an estimated $T \approx 2.3$ years. Again, the predator oscillation is unrealistically violent. Average vole density is 2100 km^{-2} and weasel density is 4.6 km^{-2}, giving 460 voles per weasel.

susceptibility to disease, and other complicating factors that increase with population density into account would most likely lead to a more rounded shape of the peaks, similarly to the one seen in actual observations. It turns out that the time average abundances are fairly close to the predicted equilibrium abundances in all of our numerical solutions. We shall therefore use mean abundances and equilibrium abundances interchangeably when validating our results.

We also plug the masses into Eqs. (14) and (17). The theoretical estimates obtained this way and their uncertainties can be seen in Table I. For this particular system, we estimate a period of $T \approx 1400$ days. Compared to the observed period of around 3000 days, the error is about 50%. As far as order-of-magnitude estimates go, this is still reasonable. Neglecting the duration of the collapse phase is probably part of the reason for this error. For comparison, the cycle period obtained from linearization gives us $\frac{2\pi}{\sqrt{\alpha\delta}} = \frac{550}{(m_x m_y)^{1/8}} = 770$ days, which is far too short. This again underlines the usefulness of approximating the cycle as a series of instantaneous collapse phases followed by exponential growth phases.

Another case where the basic Lotka-Volterra equation might be useful is the interaction between the vole (*Microtus agrestis*) and least weasel (*Mustela nivalis*) in northern Scandinavia, as mentioned above. Although there still is some doubt about the role of predation in the cycle here as well, there is evidence that predation plays at least a significant part. Vole density ranges from 10^2 to 10^4 km^{-2} over a cycle, while weasel density ranges from 1 to 20 km^{-2} and is strongly correlated with vole density at northern latitudes [16]. The cycle is observed to be about 4 years long in the areas we are interested in [11]. A numerical solution of the Lotka-Volterra equations for these parameter values can be seen in Fig. 2(b). This numerical solution gives us an estimate of the period $T \approx 830$ days $= 2.3$ years and of the prey-predator ratio of 460 voles per weasel. A comparison between theoretical results calculated using the derived expressions and observations can again be seen in Table I.

Large population oscillations are observed in some rodent species even when there is no single obvious predator feeding on the rodent. One example of this is the northern collared lemming of Greenland (*Dicrostonyx groenlandicus*) [17]. We, of course, cannot use such an example to test our hypothesis about prey-predator population ratios, but we may still use it to examine the accuracy of the derived period. The results of our examination can be seen in the table, and both the estimated period and the error are similar to those of the vole.

As a final example, we will consider the wolves (*Canis lupus*) and moose (*Alces alces americanus*) of Isle Royale in

TABLE I. Table of the data used and the values calculated, including uncertainties. Numbers are rounded to the highest uncertain digit [16–22].

System	m_y (kg)	m_x	$x_{\max}$	$x_{\min}$	Observed x/y ratio	Theoretical x/y	Obs. T (days)	Theoretical T
Lynx-hare	11 ± 1	1.6 ± 0.1	180 ± 80	8 ± 4	600 ± 400	850 ± 70	3000 ± 200	1400 ± 300
Vole-weasel	0.08 ± 0.01	0.025 ± 0.002	$(1.0 \pm 0.2) \cdot 10^4$	100 ± 50	200 ± 200	510 ± 60	1600 ± 200	730 ± 90
Wolf-moose	33 ± 1	350 ± 10	–	–	40 ± 20	42 ± 2	–	–
Lemming osc.	–	0.064 ± 0.003	1000 ± 200	14 ± 5	–	–	1460 ± 0	860 ± 80

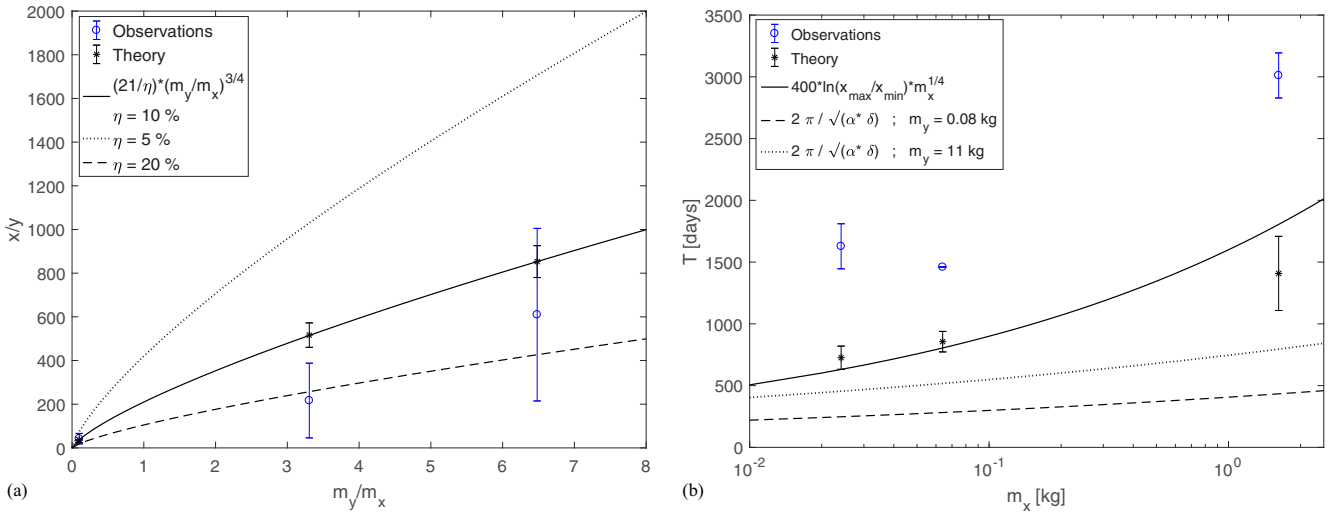


FIG. 3. (a) The observed and theoretically calculated prey-predator population ratios for the wolf-moose, vole-weasel, and hare-lynx systems. Here the full line shows the predicted power law. The dotted line shows the theoretically calculated prey-predator ratio for half the ecological efficiency used in this paper (5%), while the dashed line shows the prey-predator ratio calculated for twice the used ecological efficiency (20%). (b) The observed and calculated periods of population oscillations for voles, lemmings, and hares. The black line shows the corresponding mass power law where we have set $\ln\left(\frac{x_{\max}}{x_{\min}}\right) = 4$. This number is close to the values for the vole and lemming oscillation. The dotted and dashed lines show the $(m_x m_y)^{1/8}$ scaling law predicted from linearization, where m_y is that of lynx and weasel, respectively. We could not calculate an oscillation period for the moose of Isle Royale, and no single predator is known to cause the lemming oscillation, so they each only occur in one of the plots. Data points and error bars show the numbers without any rounding.

Lake Superior, Michigan. On this island, wolves and moose coexist in isolation, with very little interference from other animals. Due to the small size of the island, animal populations are so small that random events (such as the introduction of parvovirus to the wolf population in 1980) will have a large influence on the population, which seems to fluctuate almost erratically [27]. Therefore, we cannot determine an observational population cycle length for this system. However, the average populations should reflect an equilibrium ratio that should be predictable from wolf and moose mass. As can be seen in Table I, we obtain an accurate estimate of this ratio.

Based on these cases we may conclude that our idealized model works as an order-of-magnitude estimate of the behavior of ecosystems. There is a discrepancy between the derived period of population oscillations and what is observed, and this discrepancy cannot be explained entirely by experimental uncertainty. However, our results reproduce two patterns observed empirically, which have not yet been theoretically explained. One is the apparent scaling of oscillation periods with mass to the quarter power. Another is the scaling of period with $\ln\left(\frac{x_{\max}}{x_{\min}}\right)$. We will therefore argue that the derived expression is of interest despite the discrepancy.

V. DISCUSSION

As predicted in the Introduction, we have been able to parameterize the Lotka-Volterra equations using animal body mass in such a way that they provide fairly accurate predictions of the equilibrium predator-prey population ratio. When we also know the amplitude of the fluctuations of prey population, we obtain analytical estimates of the oscillation periods that reproduce the patterns found in nature, albeit with a discrepancy. Notably, our approximate expression for the cycle period exhibits the same allometric mass scaling as the one found empirically. Furthermore, it shows a logarithmic scaling of period with the ratio of maximum to minimum populations, which is also found in data. The ratios of average prey population to average predator population found in our simulations fit relatively well with real-world data. For the population ratios, the uncertainty of population counts and animal weights explain the errors in two of three cases. Our prediction of the amplitude of predator oscillations, however, is unreasonable in comparison with observations, possibly because of the assumption that the predator is entirely dependent on one prey species.

Of course, even though our model was only meant as a crude estimate, we need to address why we see the discrepancy that we do between theory and observations. In the case of the prey-predator population ratios, the uncertain estimate of ecological efficiency is a likely source of error. The range of ecological efficiencies observed in the real world is so large that it poses a challenge to this kind of population dynamical modeling. If our estimated efficiency is a bit too low, then it will explain the discrepancy, as can be seen in Fig. 3.

The period is off by a larger percentage, and it is less clear what might cause the error. One drastic assumption that we have made is that it takes no time for animals to grow to adult size. We have considered whether this delay might explain some of the error. To take the time required to reach full size into account, we have attempted a numerical solution of the equations while including a time delay in predator and prey reproduction. Unfortunately, this does not significantly change the oscillation period. Another possible source of error is the assumption that collapse is instantaneous. In reality, it does take some time, though not as long as the exponential growth phase. If the duration of the collapse phase also

scales with animal mass, then it would help explain why we consistently underestimate the period by about 50%.

Finally, our model is a mean-field theory, whereas, in reality, geographical separation does play a role. Maybe the fact that real predators have to seek out the prey and that prey may survive for longer in some locations than in others may serve to slow the dynamics of real, geographically extended ecosystems. This, however, is a subject that we will leave for future studies.

Despite these discrepancies, our work demonstrates that, by using the many available allometric mass scaling laws, it is possible to obtain reasonable predictions from even very simple population dynamical models. This fact should have wide applications in population dynamics. Another area where this could be applicable is in epidemiology. The incubation and recovery times of a variety of diseases with multiple host species have already been shown to scale with host mass [28], and Dobson [29] has studied a multihost disease model parameterized using mass. A possible further use of the model described here could be to construct an epidemiological model taking predation into account. Models of epidemics in predator-prey systems have been proposed before [30], but they often contain so many unknown parameters that an examination of parameter space becomes difficult. Here a parametrization using mass could significantly reduce the number of free parameters.

In 1992, Yodzis and Innes pointed out that the application of mass scaling relations to population dynamics can potentially make it a lot easier to make realistic estimates of the parameters involved. Still, to our knowledge, only recently have the predictions of a mass-parameterized population dynamical model have been tested against real-world data. The scaling of reproduction rate with animal mass has also provided us with a possible explanation for the relationship between population cycle length and mass, at least in systems where the amplitude is large. This is, for example, very much the case for several rodent and lagomorph species. In conclusion, we find that the allometric mass scaling laws that apply to a variety of biological quantities could potentially prove highly useful in population dynamics.

ACKNOWLEDGMENTS

The authors thank David Vasseur for his detailed and useful comments on an earlier draft of this article. Our research has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP/2007 2013)/ERC Grant Agreement No. 740704.

[1] L. R. Ginzburg and M. Colyvan, *Ecological Orbits: How Planets Move and Populations Grow* (Oxford University Press, Oxford, 2004).

[2] R. H. Peters, *The Ecological Implications of Body Size* (Cambridge University Press, Cambridge, 1983).

[3] G. B. West, J. H. Brown, and B. J. Enquist, A general model for the origin of allometric scaling laws in biology, Science **276**, 122 (1997).

[4] R. O. Peterson, R. E. Page, and K. M. Dodge, Wolves, moose, and the allometery of population cycles, Science **224**, 1350 (1984).

[5] P. Yodzis and S. Innes, Body size and consumer-resource dynamics, Am. Nat. **139**, 1151 (1992).

[6] U. Brose, Body-mass constraints on foraging behavior determine population and food-web dynamics, Funct. Ecol. **24**, 28 (2010).

[7] A. J. Lotka, Analytical note on certain rhythmic relations in organic systems, Proc. Nat. Acad. Sci. U.S.A. **6**, 410 (1920).

[8] R. Lindeman, The trophic-dynamic aspect of ecology, Ecology **23**, 399 (1942).

[9] R. Trebilco, J. K. Baum, A. K. Salomon, and N. K. Dulvy, Ecosystem ecology: Size-based constraints on the pyramids of life, Trends Ecol. Evol. **28**, 423 (2013).

[10] P. A. Colinvaux and B. D. Barnett, Lindeman and the ecological efficiency of wolves, Am. Nat. **114**, 707 (1979).

[11] I. Hanski, H. Henttonen, E. Korpimäki, L. Oksanen, and P. Turchin, Small-rodebt dynamics and predation, Ecology **82**, 1505 (2001).

[12] C. J. Krebs, S. Boutin, R. Boonstra, A. R. Sinclair, J. N. Smith, M. R. Dale, K. Martin, and R. Turkington, Impact of food and predation on the snowshoe hare cycle, Science (N.Y.) **269**, 1112 (1995).

[13] S. Rinaldi and S. Muratori, Slow-fast limit cycles in predator-prey models, Ecol. Model. **61**, 287 (1992).

[14] M. E. Sunquist, *Wild Cats of the World* (University of Chicago Press, Chicago, IL, 2002).

[15] P. Turchin, L. Oksanen, P. Ekerholm, T. Oksanen, and H. Henttonen, Are lemmings prey or predators? Nature **405**, 562 (2000).

[16] I. Hanski, L. Hansson, and H. Henttonen, Specialist predators, generalist predators, and the microtine rodent cycle, J. Anim. Ecol. **60**, 353 (1991).

[17] O. Gilg, I. Hanski, and B. Sittler, Cyclic dynamics in a simple vertebrate predator-prey community, Science (N.Y.) **302**, 866 (2003).

[18] H. Henttonen, T. Oksanen, A. Jortikka, and V. Haukisalmi, How much do weasels shape microtine cycles in the northern fennoscandian taiga? Oikos **50**, 353 (1987).

[19] M. Carlsen, J. Lodal, H. Leirs, and T. S. Jensen, The effect of predation risk on body weight in the field vole, microtus agrestis, Oikos **87**, 277 (1999).

[20] R. Moen, J. M. Rasmussen, C. L. Burdett, and K. M. Pelican, Hematology, serum chemistry, and body mass of free-ranging and captive canada lynx in minnesota, J. Wildlife Dis. **46**, 13 (2010).

[21] B. J. Gillingham, Meal size and feeding rate in the least weasel (mustela nivalis), J. Mammal. **65**, 517 (1984).

[22] U. S. Seal and L. D. Mech, Blood indicators of seasonal metabolic patterns in captive adult gray wolves, J. Wildlife Manage. **47**, 704 (1983).

[23] C. J. Krebs, R. Boonstra, S. Boutin, and A. R. E. Sinclair, What drives the 10-year cycle of snowshoe hares? BioScience **51**, 25 (2001).

[24] C. J. Krebs, Of lemmings and snowshoe hares: The ecology of northern canada, Proc. R. Soc. B **278**, 481 (2011).

[25] G. Mowat, M. O ' Donoghue, and K. Poole, Ecology of lynx in northern canada and alaska, in *Ecology and Conservation of Lynx in the United States*, edited by L. F. Ruggiero, K. B. Aubry, S. W. Buskirk, G. M. Koehler, C. J. Krebs, K. S. McKelvey, and J. R. Squires (University of Colorado Press, Boulder, 2000), chap. 9, pp. 265–306.

[26] R. L. Smith, D. J. Hubartt, and R. L. Shoemaker, Seasonal changes in weight, cecal length, and pancreatic function of snowshoe hares, J. Wildlife Manage. **44**, 719 (1980).

[27] J. A. Vucetich and R. O. Peterson, The influence of prey consumption and demographic stochasticity on population growth rate of isle royale wolves (canis lupus), Oikos **107**, 309 (2004).

[28] J. M. Cable, B. J. Enquist, and M. E. Moses, The allometry of host-pathogen interactions (allometry and disease), PLoS ONE **2**, e1130 (2007).

[29] A. Dobson, Population dynamics of pathogens with multiple host species, Am. Nat. **164**, 64 (2004).

[30] Y.-H. Hsieh and C.-K. Hsiao, Predator-prey model with disease infection in both populations, Math. Med. Biol. **25**, 247 (2008).

The uneasy coexistence of predators and pathogens

Authors: Andreas Eilersen* and Kim Sneppen*

*The Niels Bohr Institute, University of Copenhagen, Copenhagen, Denmark

My contributions: I contributed to the development of the model, to the associated computational work, to the production of figures and data processing, and to the writing of the article.

Publication status: Published in *European Physical Journal E* (2020).

doi: 10.1140/epje/i2020-11966-7

Notes on relation to master's project: The idea of investigating the potential role of disease as an evolutionary weapon for prey species against their predators was developed during my master's project [9], and the method for examining parameter space used here is the same as in the master's thesis. The modified Rosenzweig-MacArthur model investigated in the present article is however developed entirely during my PhD. Thus, all results presented in the paper are produced during the PhD, and the entire article is likewise written as part of the PhD work.

Eur. Phys. J. E (2020) **43**: 42
DOI 10.1140/epje/i2020-11966-7

THE EUROPEAN PHYSICAL JOURNAL E

Regular Article

The uneasy coexistence of predators and pathogens*

Andreas Eilersen[a] and Kim Sneppen[b]

Niels Bohr Institute, University of Copenhagen, Blegdamsvej 17, 2100 København Ø, Denmark

Received 28 February 2020 and Received in final form 19 May 2020
Published online: 2 July 2020

Abstract. Disease and predation are both highly important in ecology, and pathogens with multiple host species have turned out to be common. Nonetheless, the interplay between multi-host epidemics and predation has received relatively little attention. Here, we analyse a model of a predator-prey system with disease in both prey and predator populations and determine reasonable parameter values using allometric mass scaling relations. Our analysis focuses on the possibility of extinction events rather than the linear stability of the model equations, and we derive approximate relations for the parameter values at which we expect these events to occur. We find that if the predator is a specialist, epidemics frequently drive the predator species to extinction. If the predator has an additional, immune prey species, predators will usually survive. Coexistence of predator and disease is impossible in the single-prey model. We conclude that for the prey species, carrying a pathogen can be an effective weapon against predators, and that being a generalist is a major advantage for a predator in the event of an epidemic affecting the prey or both species.

Introduction

Predation is one of the fundamental modes of interaction among living organisms. Mechanisms similar to predation are found in anything from mammals to bacteria. Another equally important factor is epidemic disease, which is also found on all scales in the ecosphere. In recent years it has become clear that many epidemic pathogens are shared between several species [1], of which some presumably prey on each other. If the predator runs a risk of becoming infected when eating infected prey, it is possible that the prey species will be able to use the pathogen as a weapon against the predator. This could even be a very effective evolutionary strategy, given that prey species are often much more numerous than their predators, leading to a high infection pressure against the predator species [2]. On this basis, we propose the hypothesis that a disease shared between a prey species and its predator will turn out to be a major problem for the predator, and thus perhaps a long-term advantage for the prey. However, if the predator has several prey options, epidemics should pose much less of a threat to it, as it can just feed on an immune prey species in the event of an epidemic.

The dynamics of predator-prey-pathogen interactions in general have received some attention in recent decades. Most attention has been given to the interaction between predators and single-host epidemics or parasitism [3–7]. Lafferty *et al.* [8] have attempted to coordinate the variations on Lotka-Volterra and SIR models that have been proposed to deal with predation, disease, and parasitism. In their work, they propose a general framework underlying all consumer-resource models. However, the interplay between predation and multi-host infectious disease specifically has not been as thoroughly studied. Though this is partly justified by the similarity between parasitism and multi-host epidemic disease, the pathogen that we here want to study lacks the life-cycle dynamics associated with trophically transmitted parasites [8].

A few models similar to the one we will put forward in this paper do exist. Hsieh and Hsiao [9] have constructed one such model, and Han *et al.* [10] briefly cover another. These examples focus their analyses on the linear stability of the fixed points of their system, whereas we will focus on extinction events. We choose this focus, since an epidemic outbreak at least initially is a perturbation away from equilibrium which may temporarily drive populations to such low densities that it would lead to extinction in the real world. We will attempt to derive analytical relations for the boundaries between predator-, disease-, and prey-dominated regions of parameter space.

When analysing epidemiological models, it is difficult but crucial to determine what parameter ranges are

* Supplementary material in the form of a .pdf file available from the Journal web page at
https://doi.org/10.1140/epje/i2020-11966-7

[a] e-mail: andreaseilersen@nbi.ku.dk
[b] e-mail: sneppen@nbi.ku.dk

reasonable. A discussion of this problem is often missing from more theoretical treatments [9,10]. Therefore, we will here attempt to use the allometric mass scaling laws for many demographic and epidemiological quantities to estimate the range of parameters.

It has long been known that quantities such as reproduction rate and metabolic effect scale with animal mass to some quarter power [11]. Attempts have been made in ecology to use this to predict the behaviour of predator-prey systems [12–14]. More recently, it has been shown that disease recovery and death rates also scale with animal mass [15], which is useful in epidemiological modelling [16]. The parameterisation that we will use here will be based in part on our previous work on parameterising the Lotka-Volterra predator-prey equations [14]. The mass scaling relations are for the most part fairly general across different classes of animals. We will here be using the mass scaling relations valid for mammals. One could construct similar models for predation among other animals by mainly changing the constants of proportionality [11], and we would therefore expect our model to be relevant even for non-mammals. Only when looking at entirely different organisms such as bacteria do we need to be more careful, as the mechanisms that might be responsible for the scaling are different [17]. Nonetheless, a similar scaling law for metabolic effect exists even for bacteria [18].

In summary, the questions that we will try to answer here will be whether an epidemic affecting a prey species can drive a predator species to extinction, and if so, for what parameter values this will be most likely. We also want to examine the effect of a predator being a generalist, *i.e.* having an alternative prey option that is not affected by the epidemic.

The model

To study this system, we use the Rosenzweig-MacArthur equations for predator-prey interactions [19], but with a modified predator death rate:

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \alpha x(1 - x/K) - \phi x \frac{y}{x+\epsilon}\,, \tag{1}$$

$$\frac{\mathrm{d}y}{\mathrm{d}t} = \nu y \frac{x}{x+\epsilon} - \delta y \frac{\epsilon}{x+\epsilon}\,, \tag{2}$$

where x is prey, y is predator, α is the *per capita* prey reproduction rate and δ is the predator starvation rate in the absence of prey [20]. K is the prey carrying capacity and ϵ is the half-saturation constant for predators. ϕ and ν are coupling constants of the functional and numerical responses, respectively [21]. We choose to modify the predator death rate, since we do not expect this rate to be constant with respect to prey population. Instead, it should approach δ when there is no prey and the predators starve at a constant rate, and zero if there is a lot of prey. The last part holds if starvation is assumed to be the primary cause of death for predators, as in the Lotka-Volterra model [20].

We will combine these equations with the SIR model, which gives the following equations for the changes in population during an epidemic:

$$\frac{\mathrm{d}S}{\mathrm{d}t} = -\beta SI, \tag{3}$$

$$\frac{\mathrm{d}I}{\mathrm{d}t} = \beta SI - \gamma I, \tag{4}$$

$$\frac{\mathrm{d}R}{\mathrm{d}t} = \gamma I. \tag{5}$$

Here, S denotes susceptible individuals, I infected, and R recovered or dead individuals. β gives the rate at which each infected individual infects susceptible individuals, and γ gives the death or recovery rate of the infected [22].

When constructing our model, we shall make the assumption that the disease is always deadly, as the possibility of recovery with immunity will vastly complicate the analysis in a predator-prey system. Furthermore, we assume that infection from predator to prey is impossible, as any close encounters between the two species are likely to cause the immediate death of the prey. When modelling the system below, we find that varying the predator-predator infection rate makes relatively little difference. Figures illustrating this can be found in the supplement. For the sake of simplicity, in the following we will therefore only treat the case where the majority of predator infections stem from prey, and predator-predator infections can be neglected. We also let only healthy animals reproduce, although both healthy and infected predators eat prey. Combining the SIR and Rosenzweig-MacArthur models, we end up with the following equations for the single-prey system:

$$\frac{\mathrm{d}S_x}{\mathrm{d}t} = \alpha S_x(1 - (S_x + I_x)/K) - \beta_{xx} S_x I_x - \phi \frac{S_y + I_y}{S_x + I_x + \epsilon} S_x, \tag{6}$$

$$\frac{\mathrm{d}I_x}{\mathrm{d}t} = \beta_{xx} S_x I_x - \phi \frac{S_y + I_y}{S_x + I_x + \epsilon} I_x - \gamma_x I_x, \tag{7}$$

$$\frac{\mathrm{d}S_y}{\mathrm{d}t} = \nu \frac{S_x + I_x}{S_x + I_x + \epsilon} S_y - \beta_{yx} S_y I_x - \delta \frac{\epsilon}{S_x + I_x + \epsilon} S_y, \tag{8}$$

$$\frac{\mathrm{d}I_y}{\mathrm{d}t} = \beta_{yx} S_y I_x - \gamma_y I_y - \delta \frac{\epsilon}{S_x + I_x + \epsilon} I_y. \tag{9}$$

The equations for the number of dead individuals have been dropped, as they add no information when the disease is universally fatal. Subscripts here denote the species, with β_{ij} being the coefficient for infection from species j to species i. If we set the probability of infection when eating an infected prey equal to 1, the infection coefficient β_{yx} becomes equal to $\frac{\phi}{S_x + I_x + \epsilon}$, as the number of infected prey eaten equals the number of predators infected.

It should be noted that the exact functional form of these equations is not very important for the conclusions of this study. In the Electronic Supplementary Material, we show a parameter sweep analogous to the one found in fig. 1, but using the classical Lotka-Volterra equations

with linear functional and numerical responses. This turns out to not significantly affect the conclusions of this study.

From [11,14,23] we can find relations between predator and prey mass (m_x and m_y) and the parameters α, δ, ϕ, and ν. We want α and ν to represent theoretical maximal reproduction rates for prey and predators respectively. Instead of using the data from growing populations in the wild, where starvation, disease and other complications practically always play a role, we believe that the theoretical cap on reproduction should be set by the gestation period. α and ν should thus be the inverse gestation period [11]:

$$\alpha \approx 1/t_g \approx \frac{1}{50} m_x^{-1/4}, \qquad \nu \approx \frac{1}{50} m_y^{-1/4} \quad [1/\text{days}], \tag{10}$$

with mass in kilograms. A similar mass scaling law can be found for the incubation period of species that lay eggs [24]. We assume that when the predator is satisfied ($S_y \approx \epsilon$), the predator population is constant, giving us $\delta \approx \frac{1}{50} m_y^{-1/4}$ as well. In order to calculate how many prey the predators need to eat to reproduce this much, we must know the ecological efficiency η. The ecological efficiency, defined as the fraction of consumed prey biomass converted into predator biomass, we estimate to be 10% although the quantity varies significantly with trophic level and the specifics of the species [25,26]. Knowing the efficiency, we can calculate the number of prey eaten as $\phi \frac{S_x}{S_x+\epsilon} S_y = \frac{m_y}{\eta m_x} \nu \frac{S_y}{S_x+\epsilon} S_x$, which implies $\phi = \frac{10 m_y}{m_x} \nu = \frac{m_y^{3/4}}{5 m_x}$ Finally, also from Peters [11], we have the following approximate relation for the carrying capacity:

$$K \approx 200 m_x^{-3/4} \quad [\text{prey/km}^2]. \tag{11}$$

This relation is valid if we assume that the prey is a mammal and accept that the metabolic scaling exponent 3/4 is the "true" theoretical value of the empirically estimated scaling exponent (~ 0.61) of the carrying capacity. By using this carrying capacity relation, we decide that the units of the population densities are [km^{-2}]. ϵ is difficult to determine, and we therefore choose to set $\epsilon = K/2$. We believe this to be reasonable, as it allows the predator population growth to saturate before the prey population reaches its carrying capacity. However, as can be seen in the supplement, we can set ϵ to practically any value between $0.3K$ and K and still get similar results.

To extend the predator-prey model to the predator-prey-disease case, we also need to know the scaling relations for disease duration. According to Cable *et al.* [15], both the time until first symptoms and the time until recovery or death scale as $t = cm^{1/4}$, where c is an experimental constant. Here, we shall use the constants appropriate for rabies. We choose to use these constants since we need an estimate of the order of magnitude of the scaling coefficient. It should be stressed, however, that the disease modelled here does not correspond to any single real-world disease, since we also wish to study the effects of varying its infectivity.

According to Cable *et al.* the duration of the period during which the infected individual is symptomatic can be written $t_I \approx t_D - t_S = (c_2 - c_1) m^{1/4}$, where c_1 and c_2 are the scaling coefficients appropriate for the time until first symptoms and death, respectively. We assume that this period is of the same order as the infective period of the disease. The constants have been determined using statistical analysis, and their values are $c_1 = 9$ $(4, 19)$ and $c_2 = 16$ $(7, 32)$, where the numbers in parentheses are the boundaries of the confidence interval from $p = 2.5\%$ to $p = 97.5\%$ [15,27]. γ_i can now be found as $1/t_{I,i}$.

Finally, to make the parameterisation more intuitive, we choose to express infectivity in terms of a quantity R_{xx} related to the basic reproduction number (R_0) of the disease. The basic reproduction number represents the number of secondary infections that occur when exposing an infected individual to a completely susceptible population. The reproduction number is related to the infection coefficient as $R_{ij} = \frac{\beta_{ij} S_{i,0}}{\gamma_j}$ in the SIR model [2], where $S_{i,0}$ is the initial density of susceptible individuals of species i at the onset of the epidemic. R_0 has the important property that if it is less than 1, the disease-free equilibrium is stable in the SIR model. We find that for the prey-prey reproduction number R_{xx} to have this property in the case with immune predators, we cannot simply use the formula given above. Instead, we would have to add a correction, so that $R_{xx} \to R_{xx} - \frac{\alpha}{\gamma_x}(1 - \frac{\delta\epsilon}{\nu K})$. Nonetheless, for simplicity we will here use the formula $R_{xx} = \frac{\beta_{xx} S_{x,0}}{\gamma_x}$, since the correction is quite small (~ 0.1) for most parameter values in our parameterisation. If we choose the starting population $S_{x,0}$ to be the predator-prey equilibrium in the absence of disease, we have $R_{xx} = \frac{\beta_{xx}\delta\epsilon}{\gamma_x \nu}$.

R_0 ranges from 1, where an epidemic is barely able to sustain itself, up to 18 in measles [28]. We here vary R_{xx} from 1 to 10. The cross-species reproduction number R_{yx} will be determined by the number of prey eaten by predators which in turn depends on their mass ratio. As the initial predator population, we similarly choose the predator-prey equilibrium value, $S_{y,0} = \frac{\alpha\epsilon}{\phi} \frac{K\delta\nu + K\nu^2 - \delta^2\epsilon - \delta\epsilon\nu}{\nu^2 K}$, which reduces to $S_{y,0} = \frac{\alpha\epsilon}{\phi}$ given our parameterisation.

By using this parameterisation, we are now left with only five parameters: Prey mass, predator mass, ϵ, prey-prey disease reproduction number, and the infection probability when predators eat infected prey. If we fix this probability at 1, we save another parameter. This is not always a good approximation [29]. However, varying the infection probability has a much smaller effect than varying m_i or R_{xx}, as is demonstrated in the supplement. We therefore choose to fix the probability at 1. The mass parameterisation further ensures that the values of the parameters used are at least biologically plausible.

Examining parameter space

Based on the theoretical setup of our model we can make some predictions about the behaviour of the system in different regions of parameter space. Since the disease coupling constant $\beta_{xx} = \frac{R_{xx}\gamma_x}{\delta\epsilon}\nu = \frac{m_x^{1/2}}{700} R_{xx}$ is larger than the

maximal predator growth rate per prey $\frac{\nu}{\epsilon} = \frac{m_x^{3/4}}{5000 m_y^{1/4}}$ for all parameter values except very high prey mass and very low predator mass, we expect that if the disease survives the initial outbreak and following depletion of susceptible prey, it will competitively exclude the predator. Given that we introduce an extinction threshhold, much of the behaviour of the system will depend on whether any of the populations reach this threshhold in the initial large excursion from equilibrium caused by the epidemic. We can derive an approximate expression determining if the disease or the predator will go extinct when prey population is initially depleted.

When hit by the epidemic, the prey population will be reduced to approximately $S_{surv} \approx S_{x,0} e^{-R_{xx}}$ [2]. We may assume that while the populations of predators and infected prey are also low, the prey will grow approximately exponentially from this low population. If τ_x is the time it takes for the prey to recover, this assumption gives us

$$S_x \approx S_{x,0} e^{-R_{xx}} e^{\alpha \tau_x} = S_{x,0}, \tag{12}$$

$$\tau_x \approx \frac{R_{xx}}{\alpha}. \tag{13}$$

Similarly, the time τ_y it will take for the predator population to reach the extinction threshhold C assuming exponential decay can be found as

$$S_{y,0} e^{-\delta \tau_y} \approx C \implies \tau_y \approx -\frac{1}{\delta} \ln\left(\frac{C}{S_{x,0}}\right). \tag{14}$$

By comparing eqs. (13) and (14), we find that the predator species will be able to survive the initial period of starvation if the inequality

$$R_{xx} < -\frac{\alpha}{\delta} \ln\left(\frac{C}{S_{y,0}}\right) = -\left(\frac{m_y}{m_x}\right)^{1/4} \ln\left(10^{-6} m_y^{3/4}\right) \tag{15}$$

is fulfilled. An analogous expression

$$R_{xx} < -\frac{\alpha}{\gamma} \ln\left(\frac{C}{S_{y,0}}\right) = -\frac{7}{50} \ln\left(10^{-7} m_x^{3/4}\right) \tag{16}$$

can be derived, giving the condition for the disease to survive the initial suppression of the susceptible prey population. However, since the disease growth rate is much higher than the predator growth rate, the assumption that the infected population declines exponentially until the prey population has recovered completely is no longer approximately true. This expression therefore underestimates the value of R_{xx} where the transition from predator extinction to disease extinction occurs, as we shall see in fig. 1.

Finally, since the dependence of S_{surv} on R_{xx} is exponential, we may assume that most prey will contract the disease in the initial oubreak, $I_x \approx \frac{\delta\epsilon}{\nu}$. Using this, we can derive a relation between predator mass and prey mass at the boundary where disease spillover drives the predator to extinction. The number of prey eaten must be roughly equal to the number of predators for disease spillover to become a serious threat. We therefore have that $\phi \frac{I_x S_y}{I_x + \epsilon} \approx S_y$. Given our assumption that $I_x \approx S_{x,0}$ at the epidemic peak, this implies that

$$\frac{\delta\epsilon}{\nu} \approx \frac{\epsilon}{\phi - 1}. \tag{17}$$

By using the mass parameterisation given above, we can finally derive the relation

$$m_y^{3/4} \approx 100 \eta m_x. \tag{18}$$

When $m_y^{3/4} > 100\eta m_x$, the predator is driven to extinction by disease spillover during the initial epidemic. If some portion of the prey population is immune, as we will discuss below, this mass limit becomes higher.

To test the validity of these expressions, we perform a parameter sweep where we let the different masses and the reproduction number vary logarithmically. We scan a region of parameter space large enough that the species falling within this region are interestingly different.

By inspection of numerical solutions to the model equations, we have found that after the disease outbreak, the populations will usually perform damped oscillations of initially large amplitude around some equilibrium. Although the new post-outbreak equilibria might be stable, the initial perturbation may cause the population to temporarily reach such low values that it would lead to extinction in any system with a discrete number of individuals. As initial condition, we choose the nontrivial equilibrium of eqs. (1) and (2) to avoid introducing further, artificial oscillations into the system.

We introduce an extinction threshhold of $C = 10^{-5}$. If a population dips below this value, we consider it extinct. It should be noted that the precise value of the threshhold makes a relatively small difference in the end result, as might be expected due to the logarithmic relations of eqs. (15) and (16). After solving the equations numerically over $T = 20000$ days, we classify the end state of the system into one of four categories: Scenarios with predator survival, disease persistence, disease-predator coexistence, and scenarios where only the healthy prey population survives. To filter out transient predator-pathogen coexistence, we let the simulation run up to 10^5 days if there is still coexistence at the end of the first simulation. Plots of the regions of parameter space with predator survival and disease persistence can be seen in fig. 1.

From the plots, we see that our estimates of the behaviour of the system are approximately correct. When the epidemic does not directly affect the predators (fig. 1(a)), the predators usually survive at high R_{xx}. As predicted, for high prey mass, low predator mass, and high R_{xx}, the predator may go extinct due to starvation during the initial outbreak. We see a "zone of exclusion" at intermediate R_{xx} where the disease persists, even if its upper boundary is higher (around $R_{xx} = 4$) than predicted from eq. (16) ($R_{xx} = 2.26$). Since the disease growth rate is always higher than the predator growth rate for these parameters, the disease always drives the predator to extinction if it becomes endemic. The gap at intermediate R_{xx} is thus evidence of competitive exclusion between predator

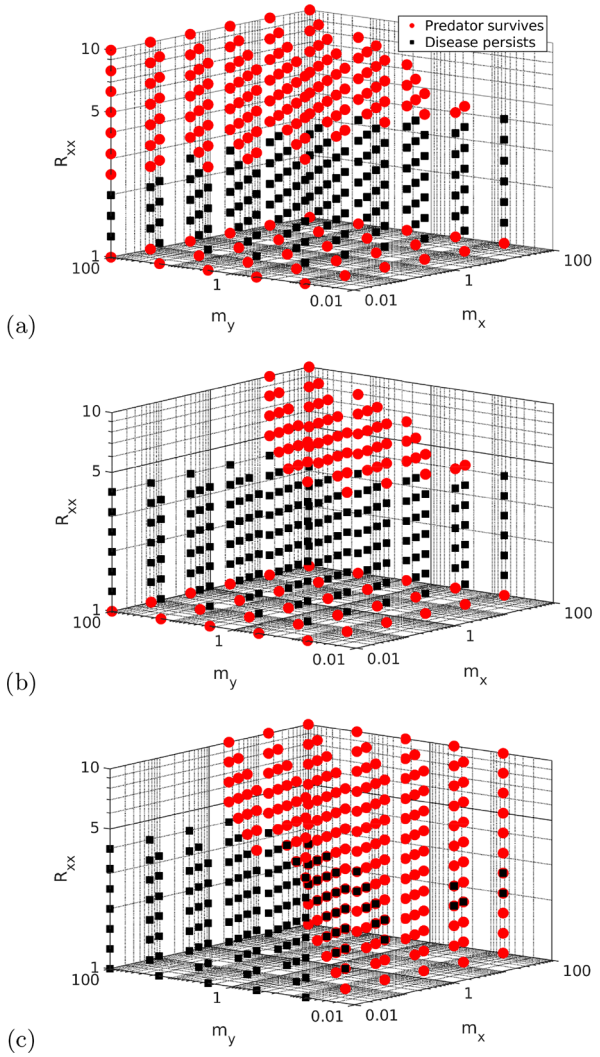


Fig. 1. Parameter space regions where the predator survives or the disease persists, as a function of prey mass m_x, predator mass m_y (both in kg), and disease reproduction number between prey R_{xx}. The coordinates of each red dot indicates a set of parameter values where the predator survives, while the location of the black squares indicate parameter values with disease persistence. If (a) the predators are immune, predators usually survive at high R_{xx}. On the other hand, if (b) the predators are susceptible, they survive at high R_{xx} only if they are not too large compared to the prey. In (c), an immune prey is included alongside the susceptible one. The susceptible predators survive regardless of disease infectivity, as long as the predator is not too large compared to the prey. At low R_{xx} there is predator-disease coexistence. If the predator is not susceptible, it always survives (figure not shown).

and disease, which both subsist on the same resource, the susceptible prey.

In the case where predators are susceptible to the disease (fig. 1(b)), the diagram again confirms our expectation that predators survive at high R_{xx} if they have less than the predicted mass relative to the prey mass. All in all, the diagrams and our approximate calculations show that sharing a pathogen with a prey species will often cause the predator to go extinct. In fact, even an outbreak of a prey-specific epidemic can cause predator extinction, at least if the predator is a specialist.

The effect of an additional, immune prey species could be interesting to study, since this would provide the predator with a resource not shared with the pathogen. We would expect this to soften the effect of competitive exclusion. If there is disease spillover, the predator may still be driven to extinction as predicted by eq. (18), but the threshhold mass will be higher if a part of the prey population is immune. To test this hypothesis, we modify eqs. (6)–(9) to include another prey that is unaffected by the disease. We assume that the immune prey is similar to the susceptible prey and simply set their parameters to be equal. The initial combined prey population is the same as before, and the prey species do not compete. We get the results seen in fig. 1(c).

There is a striking difference compared to the case with only susceptible prey. We here see the same effect as in fig. 1(b), that predators bigger than a certain mass need to eat a lot of prey and that disease spillover therefore leads to predator extinction at high predator masses. If we instead assume that predators are immune, they always survive. At lower R_{xx}, the mass threshhold is not nearly as clear and cannot be easily derived through analytical arguments. The reason for this is that the system becomes chaotic in this region, which also leads to predator-disease coexistence in some cases. This has been shown and examined in more detail in a previous study [23].

To elucidate the nature of the transition between the endemic state at intermediate R_{xx} and the predator-dominated state at higher R_{xx}, we plot the local minima of the infected prey and predator population time series as functions of a control parameter in fig. 2. Figure 2(a) shows that the assumption that disease extinction is caused by a large dip in number of infected after the initial outbreak is correct. We see the global minimum value of the number of infected prey decrease with R_{xx} until it reaches 10^{-5}, at which point the disease dies out. Figure 2(b) concerns the case with two preys and examines the nature of the transition from predator exclusion to predator-disease coexistence. The figure shows two time series at different prey masses. It can be seen here how chaotic behaviour of the system may lead to disease extinction for one set of parameters and a decaying chaotic transient leads to predator-disease coexistence for another.

Discussion

The most striking conclusion to be drawn from this study is that an emerging epidemic in a specialist predator-prey system will tend to drive the predator, *but not the prey*, to extinction. Packer *et al.* have previously concluded that there are many situations in which a predator species might keep prey epidemics and parasites in check [7]. The argument that we will make based on this study is the converse: Given our dynamical model, epidemic pathogens will make life hard for predators. The parameter sweeps

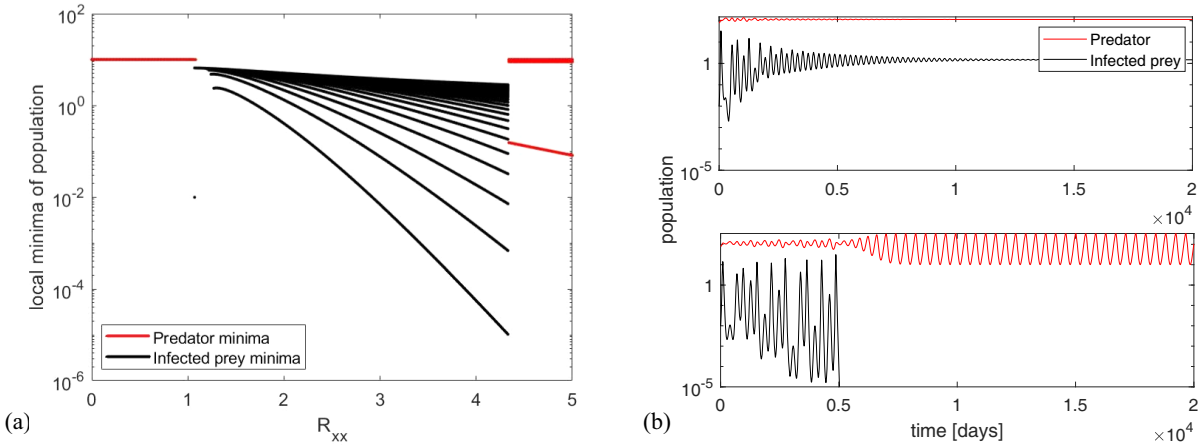


Fig. 2. Graphs illustrating the nature of transitions between different regions of parameter space. In (a) we plot the local minima of the timeseries of the infected prey population (black) and predator population (red) as a function of the control parameter R_{xx}. Like a bifurcation diagram, this provides information about the fixed points of the system, but also makes it possible to observe when the population reaches the extinction threshhold during an initial, large excursion. The diagram shows the transition from a predator-dominated state at low R_{xx} to a disease-dominated state at intermediate R_{xx}, to another predator-dominated state at high R_{xx}. The latter transition happens when the global minimum reaches the extinction threshhold, while the former happens when the predator-prey equilibrium becomes unstable at $R_{xx} \approx 1$. Here, $m_x = m_y = 1$. Panel (b) shows two timeseries illustrating the transition from a state with predator-disease coexistence (upper panel) at $m_x = 0.063$ to a predator-dominated state at $m_x = 0.40$ (lower panel) in the version of the model with two prey species. We see that the system is at least transiently chaotic. $m_y = 0.063$ and $R_{xx} = 1.26$.

show that disease and specialist predators cannot coexist. We believe this to be an example of the disease competitively excluding the predator. Both the pathogen and the predator share a resource —the susceptible prey— and in such cases, long-term coexistence is impossible [30]. As the spread of the disease is not limited by saturation or energetic concerns, it will tend to win over the predator. What saves the predator in our model is the extinction threshhold which means that very infectious diseases do not become endemic.

The implication of this conclusion is that we should see very few ecosystems with specialist predators, prey, and a shared pathogen in the real world, as they are inherently unstable. One potential caveat is that we have here focused on mammals, using the mass scaling relations and assumptions relevant for mammalian predator-prey systems. However, due to the near-universality of mass-scaling relations in animals [18] we expect that most of the relations derived here should be easily transferable to other classes.

In the large region where the predator goes extinct but the pathogen becomes endemic, carrying the pathogen may still turn into an advantage for the prey species. From evolutionary biology, we know that when a pathogen becomes endemic in a given species, there will be a pressure for it to evolve to become less lethal over time [31]. This allows the pathogen to live longer in each host, and possibly to spread more effectively. An initially fatal epidemic can thus end up becoming harmless to its primary host species. If it has wiped out the predator in the process, this will represent a win-win situation for the prey species.

Finally, as an additional result, predators that are much bigger than the size of their prey are a lot more vulnerable to infection with a shared pathogen from their prey, since they need to eat more potentially infected individuals to survive. This is true even for generalist predators and is an obvious consequence if a large percentage of the prey population is infected. What is less obvious is that the upper bound on predator to prey mass ratio drops abruptly when R_{xx} dips below the disease persistance threshhold in the generalist predator case. Above this threshhold, a generalist predator species can be many times the size of its infectious prey species and still not go extinct due to infection. Below the threshhold, a predator species larger than the infected prey will be driven to extinction by cross-species infections. A further complication in this region of parameter space is the emergence of chaotic behaviour, which means that within a relatively small region, there are cases where the pathogen dies out, where the predator dies out and where both coexist.

The physical reason behind the change in threshhold mass is that at high infectivities, the epidemic quickly uses up the supply of susceptibles and dies out. Therefore, a smaller portion of the predator population has time to be infected. This result, in addition to energetic concerns about hunting very small animals, could lead to an evolutionary pressure for predators to not grow too large compared to their prey.

Given all of the above, we conclude that epidemic diseases can serve as an evolutionary weapon against specialist predators. A pathogen infecting a prey species will competitively exclude any specialist predator species, even when the predator is not itself susceptible to the pathogen. Epidemics shared between predator and prey may help impose an upper limit on the predator-prey size ratio, since eating a lot of small prey is dangerous if the

prey is infectious. The negative effect of prey disease on the predator is however weakened a lot when we take into account additional, immune prey species. The uneasy coexistence of predators and pathogens should make specialist predator-prey-disease systems rare in the real world. Our study supports the conclusion that being a specialist predator is a highly vulnerable position when the prey is affected by an epidemic pathogen, and that being a generalist should be evolutionarily favourable for predator species. Normally, one would expect that competitive exclusion presents a drive towards speciation and specialisation [32]. Our model, on the contrary, provides an example of how the inherent vulnerability of specialists will drive species towards generalisation.

In conclusion, our study supports the idea that shared epidemic diseases could be a much more important factor in the coevolution of predator and prey species than they are usually given credit for.

Our research has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme under grant agreement No. 740704.

Author contribution statement

KS proposed the subject of the study and the model equations. AE produced the figures, derived the parameter space equations and wrote the manuscript. KS and AE revised the manuscript.

References

1. Mark E.J. Woolhouse, Sonya Gowtage-Sequeria, Emerg. Infect. Dis. **11**, 1842 (2005).
2. Sergei Maslov, Kim Sneppen, Phys. Rev. E **96**, 022412 (2017).
3. J. Chattopadhyay, O. Arino, Nonlinear Anal. Theory Methods Appl. **36**, 747 (1999).
4. Herbert W. Hethcote, Wendi Wang, Litao Han, Zhien Ma, Theor. Popul. Biol. **66**, 259 (2004).
5. Pierre Auger, Rachid Mchich, Tanmay Chowdhury, Gauthier Sallet, Maurice Tchuente, Joydev Chattopadhyay, J. Theor. Biol. **258**, 344 (2009).
6. H.I. Freedman, Math. Biosci. **99**, 143 (1990).
7. Craig Packer, Robert D. Holt, Peter J. Hudson, Kevin D. Lafferty, Andrew P. Dobson, Ecol. Lett. **6**, 797 (2003).
8. Kevin D. Lafferty, Giulio DeLeo, Cheryl J. Briggs, Andrew P. Dobson, Thilo Gross, Armand M. Kuris, Science **349**, 854 (2015).
9. Ying-Hen Hsieh, Chin-Kuei Hsiao, Math. Med. Biol. **25**, 247 (2008).
10. Litao Han, Zhien Ma, H.W. Hethcote, Math. Comput. Modell. **34**, 849 (2001).
11. Robert Henry Peters, *The Ecological Implications of Body Size* (Cambridge University Press, Cambridge, 1983).
12. P. Yodzis, S. Innes, Am. Nat. **139**, 1151 (1992).
13. Joshua Weitz, Simon Levin, Ecol. Lett. **9**, 548 (2006).
14. Andreas Eilersen, Kim Sneppen, Phys. Rev. E **99**, 022405 (2019).
15. Jessica M. Cable, Brian J. Enquist, Melanie E. Moses, PLoS ONE **2**, e1130 (2007).
16. A. Dobson, Am. Nat. **164**, 64 (2004).
17. Geoffrey B. West, James H. Brown, Brian J. Enquist, Science **276**, 122 (1997).
18. Lev Ginzburg, Mark Colyvan *et al.*, *Ecological Orbits: How Planets Move and Populations Grow* (Oxford University Press on Demand, 2004).
19. Michael L. Rosenzweig, Robert H. MacArthur, Am. Nat. **97**, 209 (1963).
20. Alfred J. Lotka, Proc. Natl. Acad. Sci. U.S.A. **6**, 410 (1920).
21. Crawford S. Holling, Can. Entomol. **91**, 385 (1959).
22. William Ogilvy Kermack, A.G. McKendrick, Gilbert Thomas Walker, Proc. R. Soc. London A **141**, 94 (1933).
23. Eilersen Andreas, Mogens H. Jensen, Kim Sneppen, Sci. Rep. **10**, 3907 (2020).
24. James F. Gillooly, Eric L. Charnov, Geoffrey B. West, Van M. Savage, James H. Brown, Nature **417**, 70 (2002).
25. Raymond Lindeman, Ecology **23**, 399 (1942).
26. P.A. Colinvaux, B.D. Barnett, Am. Nat. **114**, 707 (1979).
27. Alan C. Jackson, *Research advances in rabies*, *Advances in Virus Research*, Vol. **79**, 1st edition (Elsevier/Academic Press, Amsterdam Boston, 2011).
28. William J. Moss, Diane E. Griffin, Nat. Rev. Microbiol. **4**, 900 (2006).
29. R.O. Ramsden, D.H. Johnston, J. Wildl. Dis. **11**, 318 (1975).
30. G.F. Gause, Science **79**, 16 (1934).
31. P.W. Ewald, Emerg. Infect. Dis. **2**, 245 (1996).
32. Robert Macarthur, Richard Levins, Am. Nat. **101**, 377 (1967).

Supplemental figures to the article "The uneasy coexistence of predators and pathogens"

May 19, 2020

Outcomes for different values of the half-saturation constant ϵ

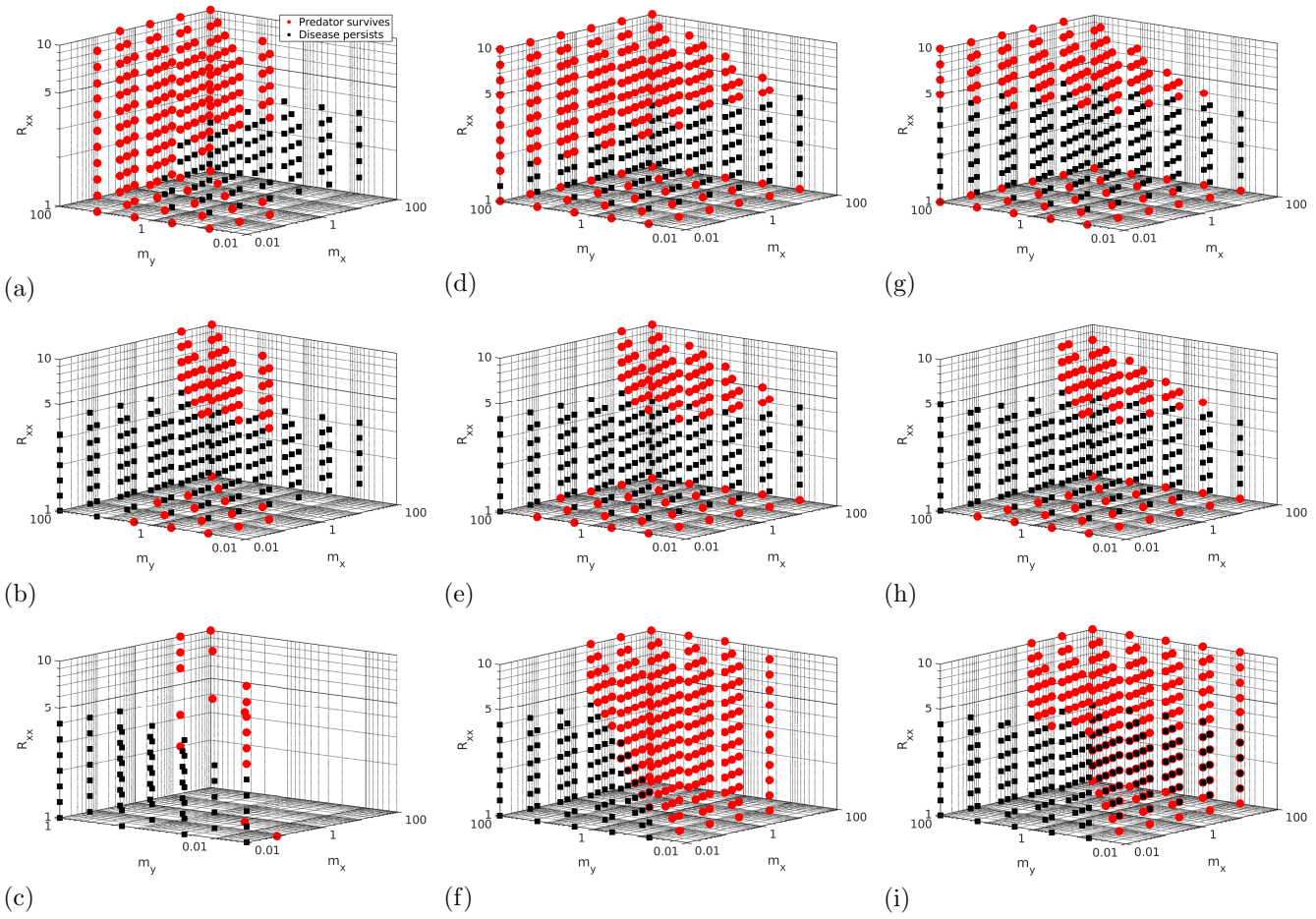


Figure 1: Plots showing the distribution of predator survival (red circles) and disease persistence (black squares) scenarios in parameter space for the three versions of our model. The coordinates of the markers indicate the parameter values leading to the given scenario, with prey mass m_x and predator mass m_y in kg. Row 1 (a,d,g) shows the model with immune predators, row 2 (b,e,h) susceptible predators, and row 3 (c,f,i) the model with two prey species, one immune and one susceptible, as well as susceptible predators. In the first column, figs. (a-c), we have set $\epsilon = 0.1K$. Here, the system appears to be unstable, at least when there are two prey species (fig. (c)). When $\epsilon = 0.3$ (figs. (d-f)), we get a distribution of predator survival and disease persistence similar to the one shown in the main text of this paper. This remains true for $\epsilon = 0.95K$ (figs. (g-i)). When $\epsilon > K$, the predator reproduction rate remains smaller than the death rate for all possible prey densities, and the model is therefore invalid for these ϵ. We see that in the range from $\epsilon = 0.3K$ to $\epsilon \approx K$, the exact value of ϵ makes little difference, and we are therefore justified in taking it to be $K/2$.

Outcomes for different prey-predator and predator-predator infectivities

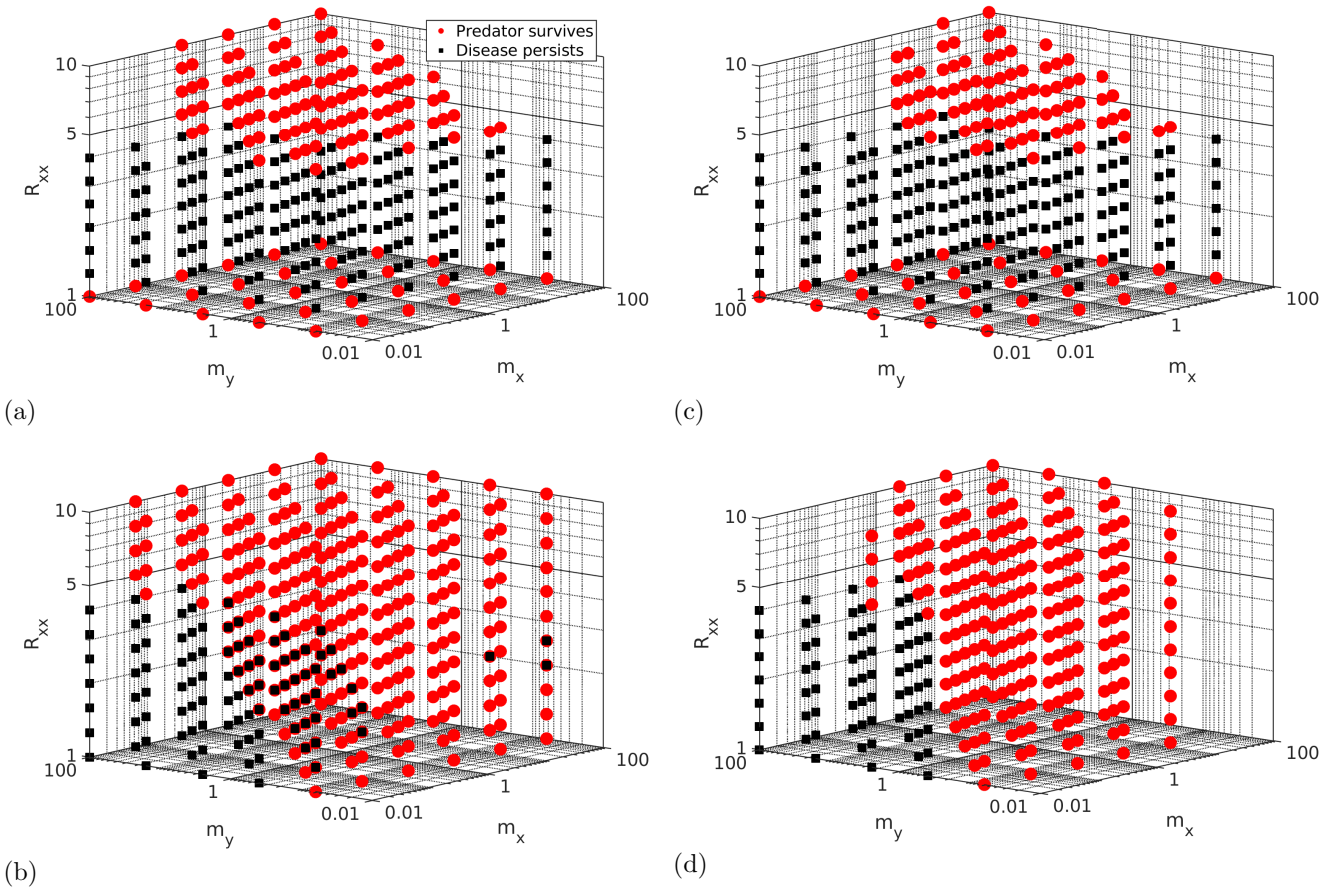


Figure 2: The effects of varying the infection probability when predators eat prey, p_I. The locations of the red dots in parameter space indicate parameter values leading to predator survival, while black squares indicate parameter values leading to disease persistence. (a) shows the one-prey system for $p_I = 0.1$ and (b) shows the two-prey system, also for $p_I = 0.1$. (c)-(d) show the same two systems for $p_I = 0.5$.

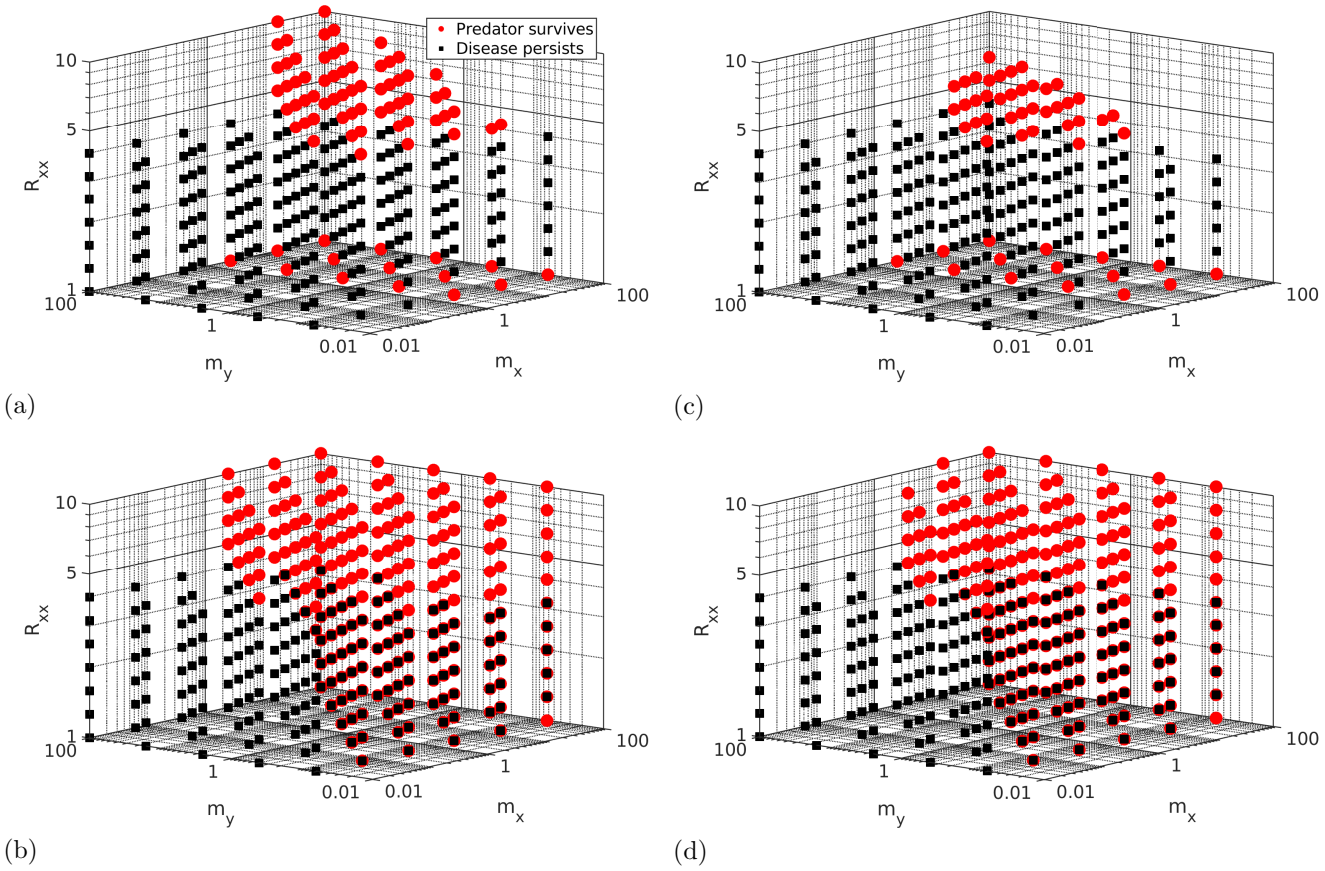


Figure 3: Plots demonstrating the effect of varying the rate of infection from predator to predator, here expressed via the parameter R_{yy} which we have defined as $R_{yy} \equiv \beta_{yy} S_{y,0}/\gamma_y$, where $y_0 = \alpha\epsilon/\phi$ is the initial density of predators at the Lotka-Volterra equilibrium without disease. It is analogous to the basic reproduction number for prey-prey infections, R_{xx}. (a) and (b) show the parameter values leading to predator survival (red dots) and disease persistence (black squares) for $R_{yy} = 1.26$ in the one-prey and two-prey models respectively. The coordinates of the markers indicate the parameter values. (c) and (d) show the outcomes for the same two models, but for $R_{yy} = 5.01$. Given the magnitude of change of R_{yy}, the changes in predator survival are relatively minor.

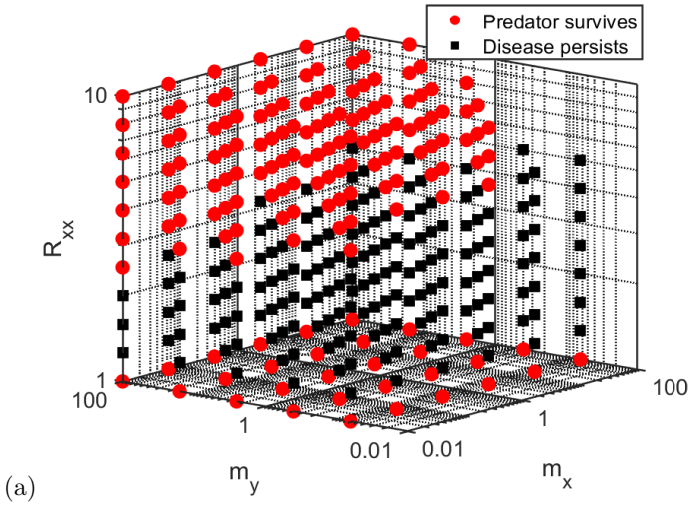
Parameter sweep with alternative functional form

As an argument for the generality of our conclusions, we repeated the parameter sweep of the main paper, but using the classical Lotka-Volterra equations to model predator-prey interactions:

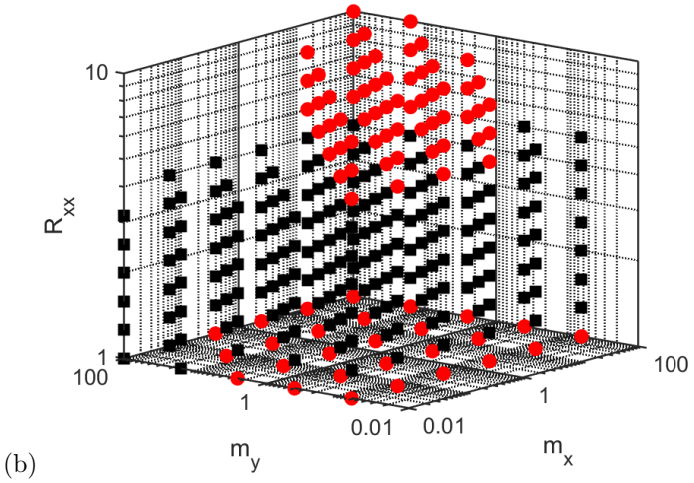
$$\frac{dx}{dt} = \alpha x - \phi xy \tag{1}$$

$$\frac{dy}{dt} = \nu yx - \delta y. \tag{2}$$

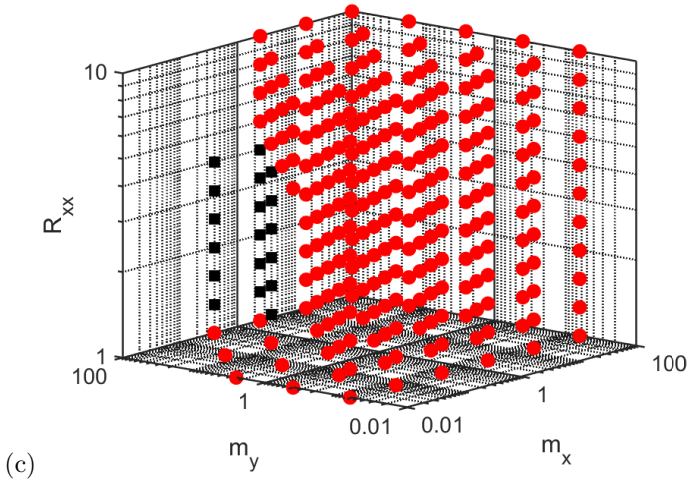
We here choose the parameterisation $\alpha = \frac{1}{50} m_x^{-1/4}$, $\delta = \frac{1}{50} m_y^{-1/4}$, $\phi = 10^{-3} m_y^{3/4}/m_x$, $\nu = \phi \frac{m_x}{10 m_y}$. The reasoning behind this parameterisation can be found in our previous paper [1]. We have slightly modified the parameterisation of α and δ in order to get parameter values of the same order of magnitude as those used in the main model.



(a)



(b)



(c)

Figure 4: This figure shows parameter sweeps analogous to the ones shown in the main paper, but using the classical Lotka-Volterra equations with no prey carrying capacity and linear functional and numerical responses. The appearance of parameter space is highly similar to that of the main paper. (a) shows the case with immune predators, (b) the case with disease spillover from prey to predator, and (c) the case with two prey species, one susceptible and one immune.

References

[1] A. Eilersen and K. Sneppen, "Applying allometric scaling to predator-prey systems," *Phys. Rev. E*, vol. 99, no. 2, 2019.

Chaos in disease outbreaks among prey

Authors: Andreas Eilersen*, Mogens Høgh Jensen*, Kim Sneppen*

*The Niels Bohr Institute, University of Copenhagen, Copenhagen, Denmark

My contributions: I contributed to the development of the model, to the associated computational and analytical work, to figure production, and to the writing of the article.

Publication status: Published in *Scientific Reports* (2020).

doi: 10.1038/s41598-020-60945-z

Notes on relation to master's project: The predator-prey-pathogen equations of the model presented here were derived from the same SIR/Lotka-Volterra framework as those presented in the master's thesis. However, the model is different from any of those presented in the master's thesis in that it includes two prey and one predator species. None of the work which this article is based on was done as part of my master's project.

SCIENTIFIC REPORTS
nature research

OPEN

Chaos in disease outbreaks among prey

Andreas Eilersen*, Mogens H. Jensen & Kim Sneppen

Epidemics are highly unpredictable, and so are real-world population dynamics. In this paper, we examine a dynamical model of an ecosystem with one predator and two prey species of which one carries a disease. We find that the system behaves chaotically for a wide range of parameters. Using the allometric mass scaling of animal and disease lifetimes, we predict chaos if (a) the disease is infectious enough to persist, and (b) it affects the larger prey species. This provides another example of chaos in a Lotka-Volterra system and a possible explanation for the apparent randomness of epizootic outbreaks.

Real ecosystems are full of noise and unpredictable dynamics. In systems as unpredictable as the ecosphere, it seems reasonable to look for chaos. Nonetheless true chaos has long been regarded as unlikely in nature. When a model of an ecological system gives rise to chaotic behaviour, it has been taken as an argument against the existence of such a system[1], although this view has been gradually changing for the past few decades[2].

A component of ecosystems that is known to frequently be unpredictable is disease[3]. In this paper, we will therefore examine whether the interplay between a generalist predator (here modelled as a predator with two prey species) and a pathogen that affects one of the preys can cause chaotic dynamics. We will prove that under our assumptions it can, and derive the conditions for this to happen, as well as examine on what timescale chaotic effects become noticeable. Our main argument is that chaos may be behind the unpredictability of epizootics.

A truly chaotic system depends so sensitively on initial conditions that it appears to be unpredictable despite being deterministic. Even a tiny change in initial conditions (the proverbial "flap of the butterfly's wings") can drastically change the outcome. Dynamics like these are particularly interesting in the context of epidemiology.

One of the most well-known examples of chaos is the logistic map, originally a discrete map model of animal reproduction[4]. Despite some prominent early examples of chaos originating in ecology, the focus in the study of chaos was elsewhere. Fewer instances of chaos were found in ecological models in continuous time such as the one presented here, and the topic therefore received less attention. Nonetheless, chaotic behaviour has been predicted from continuous time mathematical models in some cases[2]. A few important examples include models with two competing prey species[5–8], N competing species[9], and an omnivore-prey-resource system[10]. See also the review by Hastings *et al.*[2] for an overview of the earlier work on this topic. A more recent discovery is the fact that chaos occurs even in a simple discrete time Lotka-Volterra system[11].

More recently, the view of chaos as a solely destabilising factor that leads to ecosystem collapse has therefore mellowed a bit. Earn *et al.*[12] even suggest that chaos might have a stabilising effect by desynchronising separate ecosystems and thus enabling species re-immigration, the so-called rescue effect[13]. Chaos is thus increasingly thought to be an inherent dynamic in ecosystems[2]. The study presented here lends further support to this view.

Below, we will see that in a system governed by the classical Lotka-Volterra equations, chaos should often be expected when a predator-prey system is exposed to a serious disease. More precisely, we will show that in a system where a generalist predator subsists on two prey species, a disease becoming endemic in one of them can lead to chaos. This system is visualized in Fig. 1, including a typical trajectory for the four variables.

The assumption that the dynamics can be described well by the Lotka-Volterra equations with linear functional responses is fairly restrictive. In nature, it will correspond to an ecosystem where prey equilibrium populations are far from the carrying capacity and it thus makes the applicability of this study more narrow. Nonetheless, some of the basic features of the model presented below are present in nature (e.g. predator-prey oscillations[14] and periodic epidemics[15]), and we therefore believe it to capture some essential features of real ecosystems, although it may be very approximate.

Most epidemics are notoriously unpredictable[16], particularly in the cases of zoonoses and vector-borne diseases[17], and it has previously been considered whether seasonality or noise might drive chaos in the dynamics of certain childhood diseases[16,18,19]. Furthermore, a few papers have examined the potential chaotic behaviour of enzootics in predator-prey ecosystems, usually with infection in the predator[20,21], but also in the prey[22]. However,

University of Copenhagen, Niels Bohr Institute, Blegdamsvej 17, 2100, København Ø, Denmark. *email: andreaseilersen@nbi.ku.dk

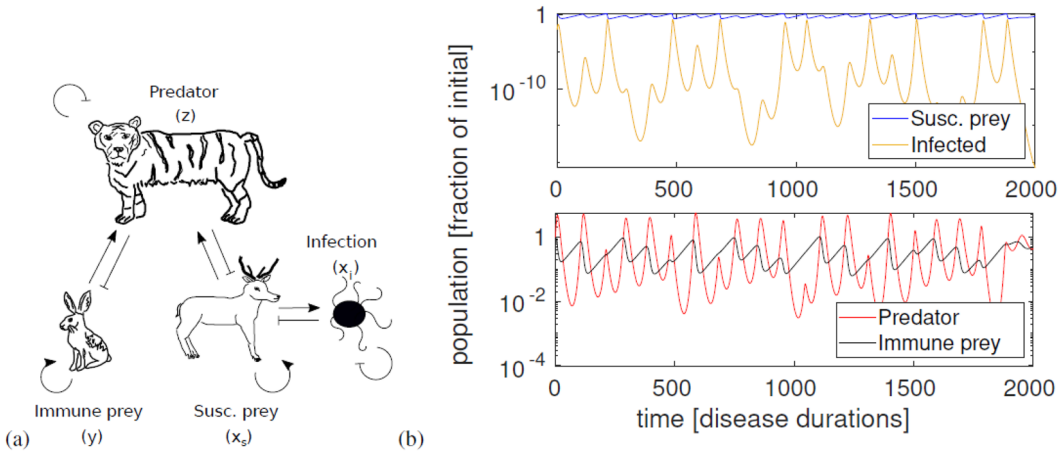


Figure 1. Diagram of the system and time series of the numerical solution. (**a**) A visual representation of the system. Pointy arrows indicate a positive effect, while blunt arrows indicate a negative one. Self arrows indicate reproduction/death that is independent of the other species. Not shown is the interaction between predators and infected individuals, which the predator can eat, although it cannot itself be infected. We expect the number of infected prey to be so small at any given time that it usually does not contribute much to predator growth. The system is still chaotic if we remove the predator reproduction term from eating infected prey. The panel (**b**) shows a simulation of the system. Note especially the large variations in number of infected. Parameters used are $a = 7/400$, $b = 0.0208$, $c = 2$, $d = 0.3098$, $R = 1.5$.

to our knowledge, chaos and disease have not been studied in the context of generalist predators, which must be assumed to be more common in nature than entirely specialist predators. Neither has chaos been discovered before in the simplest possible predator-prey-disease models with linear functional and numerical responses and mass-action disease dynamics. It is this discovery that we will present and analyse in this paper. Hopefully, it will further elucidate the role of chaos in ecosystems and the interplay between disease and predation.

The model

We base our model upon the classical Lotka-Volterra equations[23], to which we add a second prey species and combine it with the SIR model[24]. To simplify the analysis, we assume that the disease is 100 percent fatal, as recovery with immunity would give rise to a number of new steady states and complex scenarios. This model is in part based on an earlier model[25], though this had non-linear functional and numerical responses. The older model also exhibits chaotic behaviour, and we will examine it in the supplement.

In order to achieve the most basic system of equations, we assume that all functional and numerical responses are linear, that there is no competition between the prey species, and that there is no carrying capacity on the prey nor the predator. It should however be noted that chaos is still possible in systems with nonlinear functional responses and a prey carrying capacity. We further assume that healthy and infected animals are equally difficult to catch and equally nutritious for the predator. Thus, we arrive at the following equations:

$$\begin{aligned}
\frac{dx_s}{dt} &= \alpha_x x_s - \beta x_s x_i - \epsilon_x x_s z \\
\frac{dx_i}{dt} &= \beta x_s x_i - \epsilon_x x_i z - \gamma x_i \\
\frac{dy}{dt} &= \alpha_y y - \epsilon_y y z \\
\frac{dz}{dt} &= \eta_x (x_s + x_i) z + \eta_y y z - \delta z,
\end{aligned} \tag{1}$$

where x_s and x_i are healthy and infected populations of the susceptible prey species, y is the population of the immune prey species, and z is the population of predators. $\alpha_{x,y}$ are the reproduction rates of the prey species, β is the infection coefficient of the disease, and γ is the death rate of infected individuals. Finally, $\varepsilon_{x,y}$ are the coupling constants signifying the rate of individual prey of species x and y being eaten per predator, while $\eta_{x,y}$ are the predator reproduction rates from eating prey of species x and y, and δ is the predator starvation rate in the absence of prey. This is far too many parameters to allow us to get any meaningful information out of the system. Therefore, we rescale it to a timescale and characteristic populations that reflect the dynamics of the system.

First, let $\tilde{t} \equiv \gamma t$, meaning that we will measure time in units of the time it takes an infected prey to die. Thus, the lifetime of the disease is our timescale. We also choose the unit of prey population sizes to be the Lotka-Volterra equilibrium populations in a prey-predator system with only that prey species and the predator. The unit of predator population is similarly chosen as the equilibrium predator population in a Lotka-Volterra system with only the susceptible prey and the predator. Finally, the unit population of infected prey is the same as for susceptible prey. This means that $\tilde{x}_s \equiv \frac{\eta_x}{\delta} x_s$, $\tilde{x}_i \equiv \frac{\eta_x}{\delta} x_i$, $\tilde{y} \equiv \frac{\eta_y}{\delta} y$ and $\tilde{z} \equiv \frac{\epsilon_x}{\alpha_x} z$. We then have

$$\begin{aligned}
\frac{d\tilde{x}_s}{dt} &= a\tilde{x}_s - R\tilde{x}_s\tilde{x}_i - a\,\tilde{x}_s\tilde{z} \\
\frac{d\tilde{x}_i}{dt} &= R\tilde{x}_s\tilde{x}_i - a\tilde{x}_i\tilde{z} - \tilde{x}_i \\
\frac{d\tilde{y}}{dt} &= b\tilde{y} - ac\tilde{y}\tilde{z} \\
\frac{d\tilde{z}}{dt} &= d(\tilde{x}_s + \tilde{x}_i + \tilde{y})\,\tilde{z} - d\tilde{z},
\end{aligned} \tag{2}$$

where $a \equiv \alpha_x/\gamma$, $b \equiv \alpha_y/\gamma$, $c \equiv \varepsilon_y/\varepsilon_x$, $d \equiv \delta/\gamma$ and $R \equiv \beta\, x(0)/\gamma = \beta\delta/\gamma\eta_{1x}$. R (often called R_0 for a given epidemic) is the basic reproduction number of the disease at the initial susceptible prey density $x_s = 1$. The basic reproduction number is defined as the number of new infections caused by one infected individual dropped into a susceptible population. In real epidemics R varies from around 1, the minimum required for the epidemic to start, up to around 18 in extremely infectious diseases such as measles[26]. This rescaling allows us to reduce the number of unknown parameters from nine to five, and also rids us of the coupling constants $\eta_{x,y}$ whose size and relationship with $\varepsilon_{x,y}$ are hard to determine. From now on, we will drop the tildes and simply use x,y,z etc. to refer to dimensionless variables. With so relatively few parameters, extracting information from the equations should be easy. This is what we will do in the following section.

Stability analysis. If for some parameter values the system has only one stable fixed point, we will intuitively expect it to not be chaotic. Therefore, we should be able to learn more about the potential for chaos in the system by looking at the stability of the fixed points. The physically possible (i.e. where no populations are strictly negative), nontrivial fixed points of the system 2 are

$$\begin{pmatrix} x_s \\ x_i \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ \frac{b}{ac} \\ 1 \end{pmatrix}, \begin{pmatrix} \frac{1}{R} \\ \frac{a}{R} \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \frac{b+c}{Rc} \\ \frac{ac-b}{Rc} \\ \frac{b}{ac} \\ \frac{R-1-a}{R} \end{pmatrix} \tag{3}$$

For the first fixed point, where the susceptible prey and the predator reach an equilibrium, we have two real and two purely imaginary eigenvalues of opposite sign. The real eigenvalues are $(b - ac, R - 1 - a)$. We expect this fixed point to be (marginally) stable if the real eigenvalues are both negative, i.e. if $b < ac$ and $R < 1 + a$. This translates into a situation where the disease is infectious enough to spread and the immune prey reproduces slowly relative to the susceptible prey.

The second fixed point, where the immune prey and the predator coexist in a Lotka-Volterra equilibrium, similarly has two conjugate imaginary eigenvalues and two real ones, $(-1 - b/c, a - b/c)$. As b and c are always positive, we expect the first eigenvalue to be strictly negative. The second will be negative and give rise to stability if $b > ac$, opposite of the other fixed point.

The third fixed point, an analogue to the Lotka-Volterra steady state but with the pathogen replacing the predator, has two imaginary and two real eigenvalues like the others. One of the real ones is c, however, which we know is always positive, so this fixed point is unstable as long as there are two prey species.

Finally, the fixed point with coexistence of all populations has eigenvalues that are the fourth roots of some function of the parameters. As we know that the fourth roots of any number will be four complex numbers at orthogonal angles in the complex plane, their real parts will always have mixed signs, and this equilibrium will be unstable. Our stability analysis shows that there are two stable fixed points in which the system might end up: If the disease is less infectious than some threshhold $R < 1 + a$, and it holds that the immune prey reproduces slower than the threshhold $b < ac$, it will end up in the first fixed point, where only the susceptible prey and the predator persist. If the opposite is true, $b > ac$, the system will end up in the second fixed point regardless of R. Two diagrams of the conditions for stability can be seen in Fig. 2(a,b). On the other hand, if both

$R > 1 + a$ and $b < ac$, none of the fixed points will be stable. We expect that if any chaos occurs, it will happen in this region. However, in order to prove that the system becomes chaotic, we will first have to numerically solve the equations and measure the Lyapunov exponent λ.

Lyapunov exponent and chaotic transitions. In order to determine whether the system is truly chaotic, we use a variant of the Benettin algorithm[27] to estimate λ. Instead of repeatedly measuring and renormalising a unit perturbation vector using the Jacobian, we simultaneously integrate two adjacent trajectories, measure and renormalise their separation vector, integrate again starting with the renormalised separation vector and so on. This method should be equivalent to the Benettin algorithm, as long as we do not change the direction of the separation vector upon normalisation. We thereby determine whether the (small) perturbation vector grows or contracts. In a chaotic system, we should expect two infinitesimally close trajectories to drift apart at an exponential rate, giving a positive Lyapunov exponent. To reduce the number of false positive tests for chaos, we estimate λ at two different points in the time series for each set of parameters and choose the lowest of the two values. As can be seen in Fig. 2(c,d) the measurements of the Lyapunov exponent fit with what we expect from linear stability

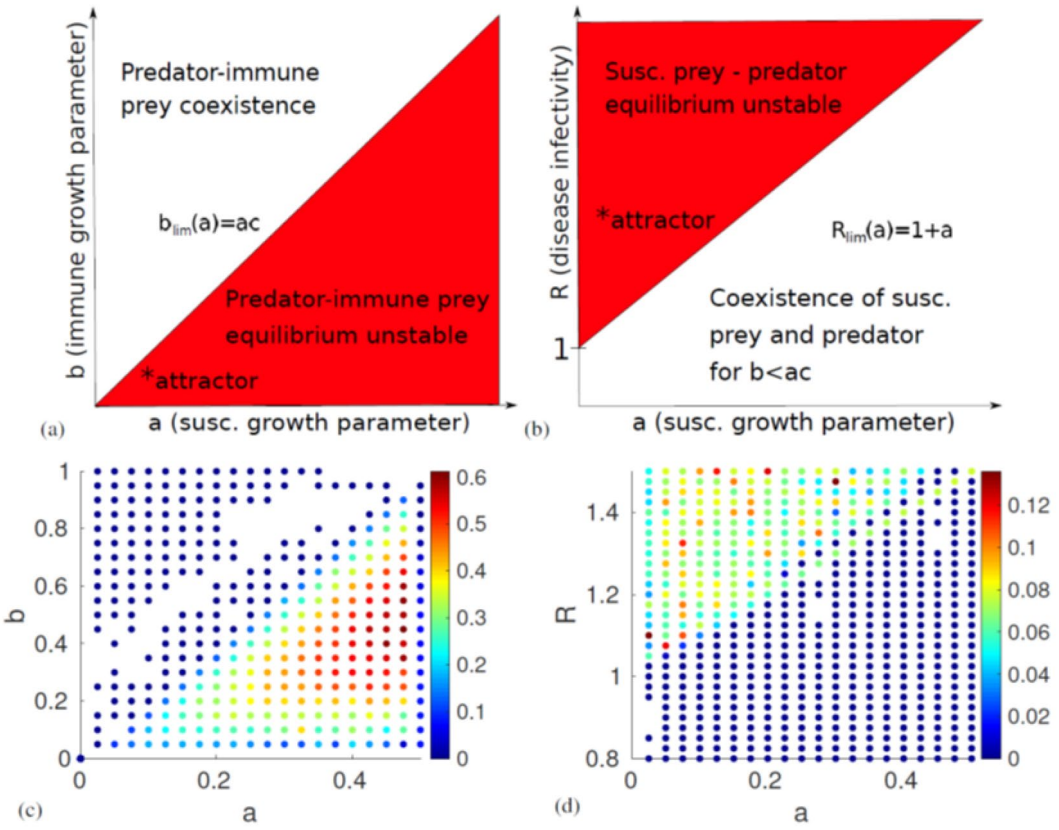


Figure 2. Diagrams of the analytically derived linear stability of the system and the numerical measurements of the Lyapunov exponent. (**a**) Stability of the predator-immune prey equilibrium as a function of the susceptible and immune prey reproduction parameters a and b. We see that the system will end up in this fixed point if $b > ac$. (**b**) Stability of the predator-susceptible prey equilibrium as a function of a and the disease basic reproduction number R. The system will end up at this fixed point if $b < ac$ and $R < 1 + a$. Thus, we can conclude that chaos is possible only if $b < ac$ and $1 + a < R$. The approximate location of the chaotic attractor plotted in Fig. 3 is marked with an asterisk. (**c**,**d**) show all parameter values resulting in a Lyapunov exponent $\lambda > 0$. Note that when looking at the timeseries, we observe no chaos for $\lambda < 0.01$. In (**c**), the system is expected to become periodic for $a > 0.5$, as $1 + a > R$. This explains the low λ-estimate near this value. The colour of the markers indicate the magnitude of λ. The diagrams look almost exactly as expected based on linear stability analysis. Parameter values used: (**c**) $c = 2$, $d = 0.3098$, $R = 1.5$, (**d**) $b = 0.0208$, $c = 2$, $d = 0.3098$.

analysis, with a far higher λ in regions where chaos is possible. Inspection of the time series of the numerical solutions reveal that the system is in fact not chaotic when the estimated $\lambda < 0.01$.

In Fig. 3 we see (a) the trajectory of the system for R close to the transition to chaos where it seems to be quasiperiodic, and (b) a chaotic attractor in $x_s - x_i - z$-space. The attractor has a fairly regular shape with a streaked surface as expected for a structure with a fractal dimension. Here, the variable y (immune prey) has been left out for ease of plotting. It was left out because it represents the predator interactions with prey that are less specific than the interactions with susceptible prey. The transition to chaos occurs rapidly and is almost discontinuous. We can therefore rule out the period-doubling route to chaos, as is confirmed by a plot of the peak values of the time series of x_s (Supplemental Fig. 2). Based on the appearance of the trajectory in phase space near the transition (Fig. 3a), we instead believe the transition to happen through a region with quasiperiodic behaviour.

Discussion

A central question to ask is what parameter values that are realistic. To address this, we will look at the four parameters of the model a, b, c and d in terms of the mass of the involved animals. It is a well-known fact that the reproduction rate and metabolism scale with animal mass[28,29]. Disease duration has also been shown to exhibit such scaling[30]. Yodzis & Innes, Dobson, Weitz & Levin, and Eilersen & Sneppen have all attempted to use this to derive information about population dynamics and epidemiology[31–34]. Using the relations found in Peters[28] and Cable *et al.*[30], we have that the reproduction rate is $\alpha \approx \frac{1}{400} m_x^{-1/4}$ $[days^{-1}]$ and the disease infection period, calculated as the difference between disease death time and time until first symptoms, is $t_i = \gamma^{-1} \approx k m_x^{1/4}$ $[days]$. Here, m_x, m_y, m_z are the masses of the susceptible prey, immune prey, and predator (in kilograms), and k depends on the disease. For example, k is approximately 7 $[days/kg^{1/4}]$ for rabies[30]. We will use this value in our calculations below, as we need an order of magnitude estimate of the coefficient. Note, however, that the disease described here is a hypothetical disease and not in fact rabies.

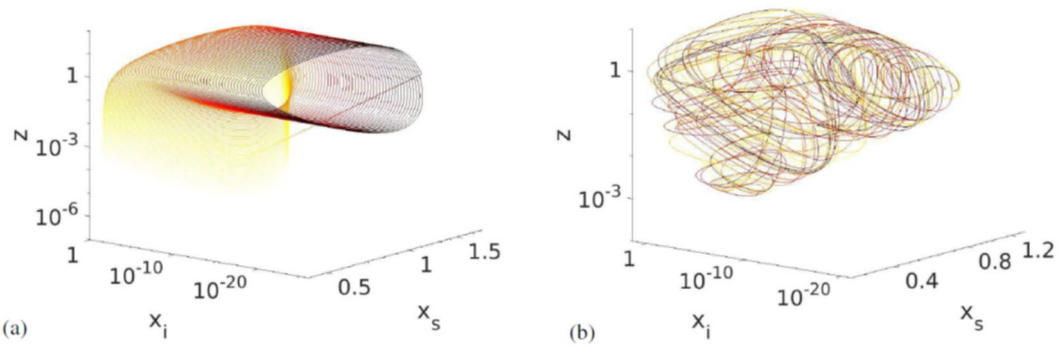


Figure 3. The evolution of the trajectories in $x_s - x_i - z$-space for different values of R. (**a**) at $R = 1.025$ where the system becomes quasiperiodic near the chaotic trasition. Other parameters were set to $a = 7/400$, $b = 0.0208$, $c = 2$, $d = 0.3098$. The fourth dimension, y, has been projected out. (**b**) Chaotic attractor for $R = 1.5$. The attractor has a characteristically streaked surface, and we estimate its dimension as $D_0 = 3.8 \pm 0.1$. The colour of the line indicates time, with darker colours signifying later times. Here, we have removed the initial transient.

Similarly, the predator starvation time is approximately its body mass divided by its metabolism, giving $t_s = \delta^{-1} \approx 20 \frac{m_z}{m_z^{3/4}} = 20 m_z^{1/4}$ [*days*]. Now, we can find approximate relations for all the dimensionless parameters:

$$a = \frac{\alpha_x}{\gamma} \approx \frac{7}{400};\ b = \frac{\alpha_y}{\gamma} \approx \frac{7}{400}\left(\frac{m_x}{m_y}\right)^{\frac{1}{4}};\ c = \frac{\epsilon_y}{\epsilon_x} \approx \frac{m_x}{m_y};\ d = \frac{\delta}{\gamma} \approx \frac{7}{20}\left(\frac{m_x}{m_z}\right)^{\frac{1}{4}}.$$

To estimate c, we have assumed that the number of individuals of a given prey species eaten by the predator will be roughly inversely proportional to the body mass of that species.

We see from these estimates that for a given disease there are in fact only two parameters that matter: The size ratio of the susceptible and immune prey species, and the size ratio of the susceptible prey and the predator. Interestingly, we also see that the condition for chaos $b < ac$ translates to $\frac{7}{400}\left(\frac{m_x}{m_y}\right)^{1/4} < \frac{7}{400}\left(\frac{m_x}{m_y}\right)$, which implies $\frac{m_x}{m_y} > 1$. Thus, our model predicts that an epizootic may cause chaos if it affects the larger of a predator's prey species. As chaos should occur for a wide range of realistic parameter values, it should be fairly common in the wake of epizootics. This of course assumes that our model is a good approximation of nature, which we will discuss below. Furthermore, two-prey one-predator systems should be closer to most real ecosystems than the classical Lotka-Volterra system, since the classical model assumes an entirely specialist predator, which is rare in nature. Even mostly specialist predators are known to have alternative prey in times of scarcity[35].

What does chaos mean for the ecosystem? The immediate consequence is unpredictable population dynamics. As mentioned in the introduction, the onset of epidemics and epizootics is already known to be unpredictable, but this is often ascribed to stochastic randomness. Our discovery that chaotic dynamics are plausible in ecosystems with three or more species where one is host to a pathogen adds an alternative explanation for this unpredictability.

Based on this model, in the wild one would expect to observe the disease lying dormant for long periods between outbreaks while the infected population is low, and then suddenly and unpredictably breaking out. The model presented here predicts very low minimum infected populations for some parameter values, so in some cases it would probably correspond to real-world pathogens that either die out locally but persist globally, or to pathogens that can survive in some dormant state outside a host. In any case, our assumption that a predator has multiple prey species of different sizes that are susceptible to different diseases should hold almost universally, making our model quite general.

Despite this, some important limitations of the model must be noted. The Lotka-Volterra equations upon which our work is based present a highly simplified and slightly pathological image of ecosystems. One issue is the fact that the classical Lotka-Volterra model with linear functional responses and growth terms gives rise to center equilibria whose orbits depend on initial populations. To avoid this, one could assume that prey species grow logistically. We briefly tried this, and it seemingly makes chaotic dynamics transient by damping oscillations. The conclusion that chaos is widespread is thus dependent on the validity of our assumptions, especially that the Lotka-Volterra model can be used as a rough approximation of real ecosystem dynamics. This is more likely to be true when prey populations are far from the carrying capacity.

In this paper, we have shown that following a disease outbreak, chaos will occur in a two-prey one-predator ecosystem for a wide range of parameters - specifically when the disease affects the larger prey species and is infectious enough to become enzootic. As opposed to previous, similar studies by e.g. Hastings & Powell[5] or Tanabe & Kumi[10], chaos occurs in the system studied here even without prey-prey competition or nonlinear functional responses. Our model therefore represents an example of a relatively minimal system that still shows complex dynamics, where the eco-epidemiological models discussed in the introduction require nonlinear responses to show chaotic behaviour. Furthermore, some light has been shed on the interaction between generalist predation

and disease. The discussion of how alternative prey species may cause chaos in the context of epizootics has likewise not been explored in the eco-epidemiological studies referenced above. The results presented here are based on the classical Lotka-Volterra model and our conclusions thus depend on its validity, which may be questionable. Nonetheless, the fact that we do not have to impose very tight constraints on parameter values is an argument for the existence and importance of chaos in ecosystems. Most importantly, our results provide one plausible explanation for the apparent unpredictability of epizootics.

Received: 30 November 2019; Accepted: 17 February 2020;
Published online: 03 March 2020

References

1. Ginzburg, L. R., Burger, O. & Damuth, J. The may threshold and life-history allometry. *Biol. Lett.* **6**, 850–853 (2010).
2. Hastings, A., Hom, C. L., Ellner, S., Turchin, P. & Godfray, H. C. J. Chaos in ecology: is mother nature a strange attractor? *Annu. review ecology systematics* **24**, 1–33 (1993).
3. Musso, D., Rodriguez-Morales, A. J., Levi, J. E., Cao-Lormeau, V.-M. & Gubler, D. J. Unexpected outbreaks of arbovirus infections: lessons learned from the pacific and tropical america. *The Lancet Infect. diseases* **18**, e355–e361 (2018).
4. May, R. M. Simple mathematical models with very complicated dynamics. *Nature* **261**, 459–467 (1976).
5. Hastings, A. & Powell, T. Chaos in a three-species food chain. *Ecology* **72** (1991).
6. Gilpin, M. E. Spiral chaos in a predator-prey model. *The Am. Nat.* **113**, 306–308 (1979).
7. Takeuchi, Y. & Adachi, N. Existence and bifurcation of stable equilibrium in two-prey, one-predator communities. *Bull. mathematical Biol.* **45**, 877–900 (1983).
8. Klebanoff, A. & Hastings, A. Chaos in one-predator, two-prey models: cgeneral results from bifurcation theory. *Math. biosciences* **122**, 221–233 (1994).
9. Gardini, L., Lupini, R. & Messia, M. Hopf bifurcation and transition to chaos in lotka-volterra equation. *J. Math. Biol.* **27**, 259–272 (1989).
10. Tanabe, K. & Namba, T. Omnivory creates chaos in simple food web models. *Ecology* **86**, 3411–3414 (2005).
11. Elsadany, A.-E. A., El-Metwally, H., Elabbasy, E. & Agiza, H. Chaos and bifurcation of a nonlinear discrete prey-predator system. *Comput. Ecol. Softw.* **2**, 169 (2012).
12. Earn, D. J., Rohani, P. & Grenfell, B. T. Persistence, chaos and synchrony in ecology and epidemiology. *Proc. Royal Soc. London. Ser. B: Biol. Sci.* **265**, 7–10 (1998).
13. Brown, J. H. & Kodric-Brown, A. Turnover rates in insular biogeography: effect of immigration on extinction. *Ecology* **58**, 445–449 (1977).
14. May, R. M. Limit cycles in predator-prey communities. *Science* **177**, 900–902 (1972).
15. Duncan, S. R., Scott, S. & Duncan, C. J. Smallpox epidemics in cities in britain. *The J. Interdiscip. Hist.* **25**, 255–271 (1994).
16. Rand, D. A. & Wilson, H. B. Chaotic stochasticity: A ubiquitous source of unpredictability in epidemics. *Proc. Royal Soc. B: Biol. Sci.* **246**, 179–184 (1991).
17. Higgs, S. The 2005–2006 chikungunya epidemic in the indian ocean. *Vector-Borne Zoonotic Dis.* **6**, 115–116 (2006).
18. Olsen, L. F., Truty, G. L. & Schaffer, W. M. Oscillations and chaos in epidemics: a nonlinear dynamic study of six childhood diseases in copenhagen, denmark. *Theor. population biology* **33** (1988).
19. Billings, L. & Schwartz, I. Exciting chaos with noise: unexpected dynamics in epidemic outbreaks. *J. mathematical biology* **44**, 31–48 (2002).
20. Stiefs, D. *et al*. Evidence of chaos in eco-epidemic models. *Math. Biosci. Eng* **6**, 855–871 (2009).
21. Das, K., Samanta, S., Biswas, B. & Chattopadhyay, J. Occurrence of chaos and its possible control in a predator-prey model with disease in the predator population. *The J. Ecol.* **108**, 306–319 (2014).
22. Chatterjee, S., Kundu, K. & Chattopadhyay, J. Role of horizontal incidence in the occurrence and control of chaos in an eco-epidemiological system. *Math. Medicine Biol. A J. IMA* **24**, 301–326 (2007).
23. Lotka, A. J. Analytical note on certain rhythmic relations in organic systems. *Proc. Natl. Acad. Sci. United States Am.* **6**, 410–415 (1920).
24. Kermack, W. O., McKendrick, A. G. & Walker, G. T. Contributions to the mathematical theory of epidemics. iii. further studies of the problem of endemicity. *Proc. Royal Soc. Lond. A* **141**, 94–122, https://doi.org/10.1098/rspa.1933.0106 (1933).
25. Eilersen, A. & Sneppen, K. The uneasy coexistence of predators and pathogens. *BioRxiv*, https://doi.org/10.1101/721506 (2020, In preparation).
26. Moss, W. J. & Griffin, D. E. Global measles elimination. *Nat. Rev. Microbiol.* **4** (2006).
27. Benettin, G., Galgani, L., Giorgilli, A. & Strelcyn, J.-M. Lyapunov characteristic exponents for smooth dynamical systems and for hamiltonian systems; a method for computing all of them. part 1: Theory. *Meccanica* **15**, 9–20 (1980).
28. Peters, R. H. *The ecological implications of body size* (Cambridge University Press, Cambridge, 1983).
29. West, G. B., Brown, J. H. & Enquist, B. J. A general model for the origin of allometric scaling laws in biology. *Science* **276**, 122–126 (1997).
30. Cable, J. M., Enquist, B. J. & Moses, M. E. The allometry of host-pathogen interactions (allometry and disease). *PLoS ONE* **2** (2007).
31. Yodzis, P. & Innes, S. Body size and consumer-resource dynamics. *The Am. Nat.* **139**, 1151–1175 (1992).
32. Dobson, A. Population dynamics of pathogens with multiple host species. *The Am. Nat.* **164**, 64–78 (2004).
33. Weitz, J. & Levin, S. Letter: Size and scaling of predator-prey dynamics. *Ecol. Lett.* **9**, 548–557 (2006).
34. Eilersen, A. & Sneppen, K. Applying allometric scaling to predator-prey systems. *Phys. Rev. E* **99** (2019).
35. Sundell, J. & Ylönen, H. Specialist predator in a multi-species prey community: boreal voles and weasels. *Integr. Zool.* **3**, 51–63 (2008).

Acknowledgements

Our research has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme under Grant Agreement No. [740704].

Author contributions

Models were designed by K.S. and parameterisation was done by K.S. and A.E. Code was written and figures prepared by A.E. Method for analysing chaos was devised by M.H.J. A.E., M.H.J. and K.S. edited the paper.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41598-020-60945-z.

Correspondence and requests for materials should be addressed to A.E.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Supplementary material to the article "Chaos in disease outbreaks among prey"

Andreas Eilersen[1,*], **Mogens H. Jensen**[1], **and Kim Sneppen**[1]

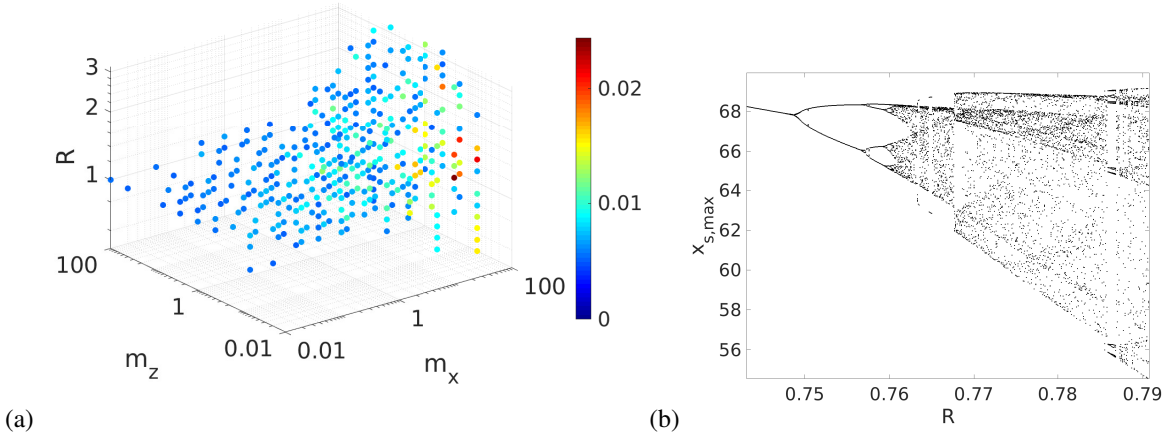
[1]University of Copenhagen, Niels Bohr Institute, Blegdamsvej 17, 2100 København Ø, Denmark
[*]andreaseilersen@nbi.ku.dk

As mentioned in the main paper, Eilersen & Sneppen[1] have previously explored an alternative predator-prey-disease model with non-linear functional and numerical response functions. The proposed system of equations is

$$\begin{aligned}
\frac{dx_s}{dt} &= \alpha x_s(1-(x_s+x_i)/K) - \beta_{xx}x_s x_i - \varepsilon\frac{z}{x_s+x_i+y+k_s}x_s \\
\frac{dx_i}{dt} &= \beta_{xx}x_s x_i - \varepsilon\frac{z}{x_s+x_i+y+k_s}x_i - \gamma x_i \\
\frac{dy}{dt} &= \alpha y(1-y/K) - \varepsilon\frac{z}{x_s+x_i+y+k_s}y \\
\frac{dz}{dt} &= \eta\frac{x_s+x_i+y}{x_s+x_i+y+k_s}z - \delta\frac{k_s}{x_s+x_i+y+k_s}z.
\end{aligned} \tag{1}$$

Here, we have introduced the half-saturation constant k_s and the prey carrying capacity K. It is also worth noting that the constants α, δ, ε, and η now represent maximum values of their respective terms and thus have to be estimated anew. In the cited paper, these quantities are estimated to be

$$\alpha \approx 1/t_g \approx \frac{1}{50}m_x^{-1/4} \quad , \quad \eta \approx \frac{1}{50}m_y^{-1/4} \approx \delta \quad [1/days] \quad , \quad K \approx 200m_x^{-3/4} \quad [prey/km^2],$$



(a) (b)

Figure 1. (a) A plot showing all parameter combinations that result in chaos, here defined as $\tilde{\lambda} > 5\cdot 10^{-3}$. The colour of each data point represents the value of $\tilde{\lambda}$. Chaos mainly occurs around $R = 1$ and the range of R that leads to chaos expands at higher prey mass and lower predator mass. In (b), we see a diagram of the peak values of x_s as a function of R, equivalent to a bifurcation diagram. The diagram shows that the chaotic transition happens through period doubling bifurcations. The masses used are $m_x = m_y = 0.4$ kg.

using the allometric scaling method described in the discussion. The model being correct thus depends on the mass parameterisation being correct, but the parameterisation on the other hand allows us to model the system with only three free parameters. We have assumed that both prey species have similar mass for simplicity. Nonetheless, the linear stability analysis of this system is fairly complicated and will not necessarily give us any useful information. We shall therefore restrict ourselves to measuring the Lyapunov exponent of this system as a function of prey mass, predator mass, and disease reproduction number, as well as examining the nature of the transition to chaos.

Even without stability analysis, we expect a transition to take place near $R = 1$, as this is where the epidemic becomes viable at Lotka-Volterra equilibrium populations. We measure the Lyapunov exponent using the same method as above. As the equations are no longer scaled to the natural time scale of the system, we rescale the measured exponents as $\tilde{\lambda} = \lambda/\gamma$. Thereby, we get a new exponent, $\tilde{\lambda}$, which gives the speed of divergence of neighbouring trajectories in units of disease durations. As can be seen in figure 1 (a), chaos in this system exists primarily at low R. The range of R-values where chaos can occur generally grows with prey size and shrinks slightly with predator size.

Figure 1 (b) shows a diagram of the peaks of the susceptible prey (x_s) time series. From this diagram, we see clearly how the transition to chaos takes place through a series of period-doubling bifurcations.

This documents that when using nonlinear functional and numerical responses, chaos can occur in a predator-prey-disease system. It even happens when the prey masses are equal, which we otherwise would not predict using the simpler model with linear response functions. As saturating response functions are probably more realistic in many ecological systems, this result strengthens our argument for the ubiquity of chaos in the wake of disease in predator-prey systems.

Chaotic transition of the system studied in the main paper

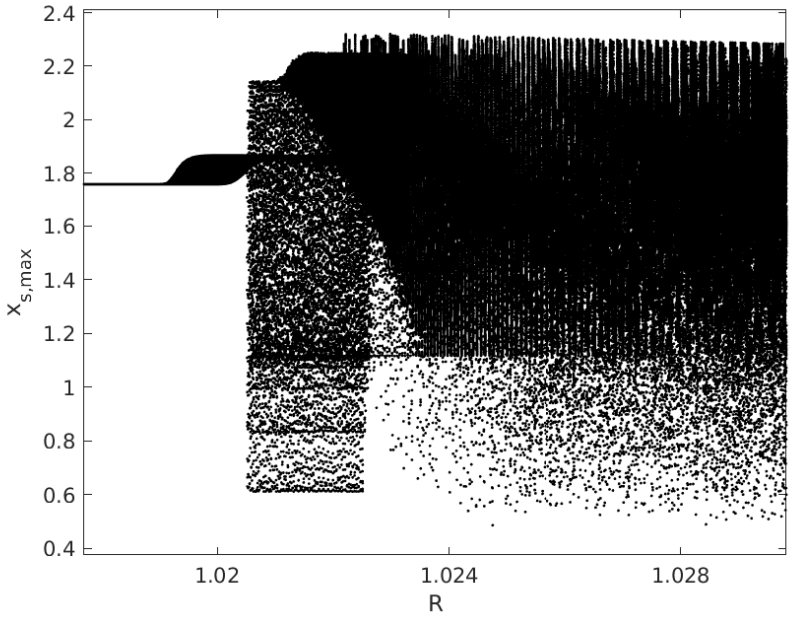


Figure 2. A plot showing the maximal values of x_s for the model system studied in the main paper as a function of the disease basic reproduction number R. The plot is analogous to fig. 1 (b). We see no signs of period doubling bifurcations, nor any other bifurcations, and we can therefore rule out that the transition to chaos takes place through this route. Instead, we believe that the transition is quasiperiodic. The trajectories near the transition, shown in fig. 3 (a) in the main paper, further support this conclusion. Parameter values used: $a = 7/400, b = 0.0208, c = 2, d = 0.3098$.

References

1. Eilersen, A. & Sneppen, K. The uneasy coexistence of predators and pathogens. *Under review* (2020).

Articles for chapter 2

The following articles form the basis of the chapter on pandemic modelling. All work for these articles was done during this PhD project.

1. Eilersen, A. and Sneppen, K. (2020). Cost–benefit of limited isolation and testing in COVID-19 mitigation, *Scientific reports* **10** (1): 1–7.
2. Eilersen, A. and Sneppen, K. (2021). SARS-CoV-2 superspreading in cities vs the countryside, *Apmis*.
3. Nielsen, B. F., Eilersen, A., Simonsen, L. and Sneppen, K. (2021). Lockdowns exert selection pressure on overdispersion of SARS-CoV-2 variants, *medRxiv*.
4. Eilersen, A., Li, R., Stenseth, N. C. (2021). Evolutionarily stable strategies (ESS) in a socially age-structured epidemic model, *In preparation*.
5. Eilersen, A. and Sneppen, K. (2021). Tradeoff between speed and infectivity in pathogen evolution, *In preparation*.

Cost-benefit of limited isolation and testing in COVID-19 mitigation

Authors: Andreas Eilersen* and Kim Sneppen*

*The Niels Bohr Institute, University of Copenhagen, Copenhagen, Denmark

My contributions: I contributed to the development of the model, to the computational work, to the production of figures and data processing, and to the writing of the article.

Publication status: Published in *Scientific Reports* (2020).

doi: 10.1038/s41598-020-75640-2

scientific reports

Check for updates

OPEN

Cost–benefit of limited isolation and testing in COVID-19 mitigation

Andreas Eilersen[1,2]✉ & Kim Sneppen[1,2]

The international community has been put in an unprecedented situation by the COVID-19 pandemic. Creating models to describe and quantify alternative mitigation strategies becomes increasingly urgent. In this study, we propose an agent-based model of disease transmission in a society divided into closely connected families, workplaces, and social groups. This allows us to discuss mitigation strategies, including targeted quarantine measures. We find that workplace and more diffuse social contacts are roughly equally important to disease spread, and that an effective lockdown must target both. We examine the cost–benefit of replacing a lockdown with tracing and quarantining contacts of the infected. Quarantine can contribute substantially to mitigation, even if it has short duration and is done within households. When reopening society, testing and quarantining is a strategy that is much cheaper in terms of lost workdays than a long lockdown. A targeted quarantine strategy is quite efficient with only 5 days of quarantine, and its effect increases when testing is more widespread.

The 2020 coronavirus (COVID-19) pandemic has raised the need for mitigation efforts that could reduce the peak of the epidemic[1,2]. To fulfill this need, theoretical modelling can play a crucial role. Traditional epidemiological models that assume universal or constant infection parameters are not sufficient to address case specific strategies like contact tracing. Therefore, we have developed an agent-based epidemiological model which takes into account that disease transmission happens in distinct arenas of social life that each play a different role under lockdown: The family, the workplace, our social circles, and the public sphere. This subdivision becomes especially important when discussing such efforts as contact tracing. Using an estimated weight of social contacts within each of these four spheres[3] we discuss the effect of various mitigation strategies.

At the time of writing, both classical mean field models[4,5] and agent-based models[2,6,7] of the COVID-19 epidemic have already been made. The models often assume contact rates and disease transmission to be stratified by age[3,8,9]. In our model, we focus on social and work networks. This will directly allow us to test the effectiveness of localized quarantine measures. In addition we allow a fraction of the contacts to be non-specific, representing random meetings.

Within families, several age groups may live together. At the same time, disease transmission within the family is probably the variable that is the most difficult to change through social distancing. Furthermore, there is doubt as to what extent children carry and transmit the disease[10]. By ignoring age as a factor our agent-based model implicitly weights children on equal footing with anyone else, and our model is not designed to address scenarios where one specifically targets older people.

Analysing what role each area of social life plays also allows us to separately treat leisure activities, and since these play a smaller economic role than work, they may be restricted with a smaller toll on society. Furthermore, if widespread testing and contact tracing is implemented, a compartmentalisation like the one we are assuming here will help in assessing which people should be quarantined and how many will be affected at any one time.

In the following, we will investigate two closely related questions. First, how a lockdown is most effectively implemented, and second, how society is subsequently reopened safely, and yet as fast as possible. To answer the first, we must examine the relative effects of reducing the amount of contacts in the workplace, in public spaces, and in closely connected groups of friends. For the second question, we will look for viable strategies for mitigation that do not require a total lockdown. Here, we will focus on the testing efficiency and contact tracing[11]. Our results will hopefully be helpful in informing future containment and mitigation efforts.

Methods

Our proposed model divides social life into family life which accounts for 40% of all social interactions, work life accounting for 30%, social life in fixed friend groups accounting for 15%, and public life which accounts for another 15%[3]. Fig. 1 shows a schematic representation of the model. Interactions in public are taken to be completely random and not dependent on factors such as geography, density or graph theoretical quantities.

[1]Niels Bohr Institute, University of Copenhagen, 2100 Copenhagen, Denmark. [2]These authors contributed equally: Andreas Eilersen and Kim Sneppen. ✉email: andreaseilersen@nbi.ku.dk

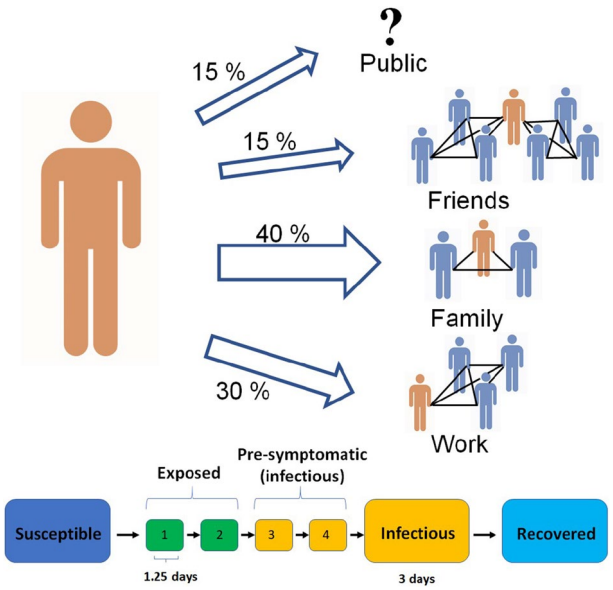


Figure 1. A diagram of the model structure. Each agent has a network consisting of a family, a workplace and two groups of friends. The family accounts for 40% of interactions. Work accounts for 30% and socialisation with friends accounts for a further 15%. The members of each of these 3 groups are fixed throughout the simulation. Finally, 15% of interactions happen "in public", which we implement as an interaction with a randomly chosen other agent. Everyone in the work and friend sub-graphs are assumed to be connected to each other. Below the graph, the underlying mechanisms of the disease are shown. We divide the exposed state into four in order to get a more naturalistic gamma distribution of incubation periods. The two last exposed states are infectious, but asymptomatic, meaning that individuals will not get tested. This is to include presymptomatic infection. In our simulation we set the family groups to an average of 2 people, and the work network to 10 completely interconnected people. The friend network consists of two groups with five in each.

Within families, workplaces, and friend groups, everyone is assumed to know everyone. Each agent is assigned one family and workplace, as well as two groups of friends. Workplaces on average contain ten people, whereas each friend group on average contains five. In the simulation runs presented here, we use a population of $N = 5000$ agents. Increasing the number of agents changes the outcome very little, except for minimising stochastic noise. We also do not allow migration in or out of the system.

We use a discrete-time stochastic algorithm. At each time-step (0.5 days), each person has one interaction with some other person. A "die roll" decides whether the person will interact with family, friends, work, or the public. The respective odds are the above-mentioned percentages 40:30:15:15. If the public is chosen, an entirely random person is selected, otherwise a person is drawn from a predefined group (family etc.). For each interaction, an infectious person has a fixed probability of passing on the disease to the person they interact with.

The family size distribution of is based on the distribution of Danish households[12]. The average number of people per household is approximately 2, and large households of more than 4 people have been ignored, as they account for less than 10% of the population. We believe that in a country where family sizes are larger and there are fewer singles, the family would be more important to the spread of disease. We test the effect of larger families in the supplement (Figs. S1–S3) and find that it does not change our overall conclusions.

We simulate the progression of disease using an SEIR model with four exposed states, $E = E_1 + E_2 + E_3 + E_4$, each lasting on average 1.25 days, corresponding to a mean incubation period of 5 days. The exposed states are presymptomatic, meaning that people will not get tested in the incubation period. We let stages E3;4 be as infectious as the I-stage, as data suggest that a substantial fraction of COVID-19 transmission happens before the onset of symptoms[13]. Multiple exposed states are included in order to get a naturalistic distribution of incubation periods. Li et al.[10] report that the mean incubation period is approximately five days and the reported distribution is fitted well by the gamma distribution we obtain from our four E-stages.

A further problem is the duration of the infectious period (I). Viral shedding has been observed to last up to eight days in moderate illness[14]. On the other hand, according to Linton et al.[15], the median time from onset to hospitalisation is three days. A bedridden patient (even if not hospitalised) is likely to transmit the disease less. To fit the observed mean serial intervals of 4.6 days of Nishiura et al.[13] we model the infectious period as a single state with an average duration of three days. In addition, the infectious presymptomatic period lasts on average 2.5 days. In comparison Ref.[16] uses a serial interval distribution with mean of 6.5 days. Other authors have suggested a longer serial interval[10] with presymptomatic infections.

Finally, the transmission rate of the disease is estimated from an observed rate of increase of 23% per day in fatalities in the USA. This also fits the observation of a growth rate of ICU admissions of about 22.5% per day in Italy[17]. With our parameters this is reproduced by a basic reproduction number $R_0 \sim 3$ (as we allow transmission

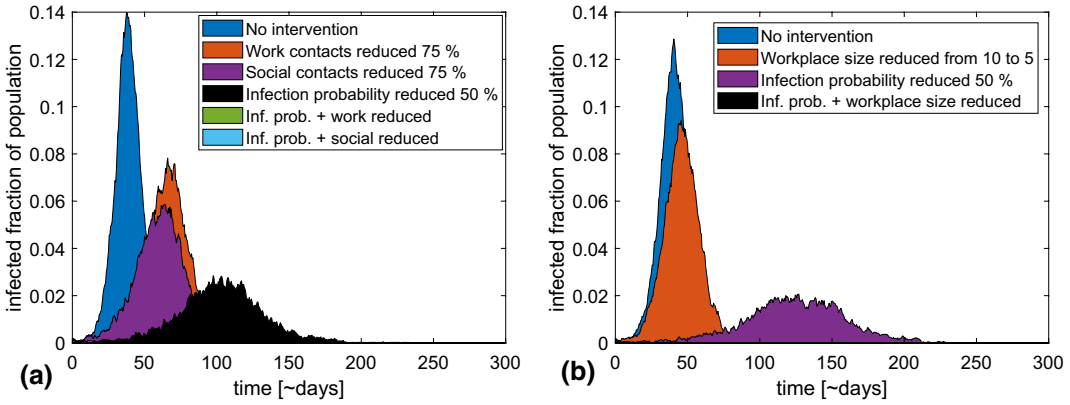


Figure 2. Comparison of various strategies with and without a reduction in transmission probability per encounter. (**a**) Reducing contacts in different social contexts by 75% through a lockdown. It can be seen that work and social contacts play roughly the same part in disease transmission. A reduction of infection probability makes the strategies relatively more effective. Reducing work contacts, for example, reduces the peak height by roughly 40% (relative to no intervention) if infection probability is high. If the infection probability is lowered, the strategy completely eliminates the epidemic. The apparent missing graphs in the figure are due to the epidemic dying out completely when a lockdown is combined with hygienic measures. (**b**) A similar comparison of the effects of reducing workplace sizes by half. This strategy is also relatively more effective if infection probability is reduced. The strategy also eliminates the epidemic at a lowered infection probability versus only a 30% reduction if infection probability remains high.

in both directions when selecting two people, this is simulated by a rate for transmission of 0.2 per encounter). Li et al.[10] estimate R_0 at 2.2 based on a growth rate of 10% per day in confirmed COVID-19 cases in Wuhan prior to Jan. 4.

Having calibrated the model in this way, we want to explore mitigation strategies for the corona epidemic. Specifically, we will investigate the relative importance of the areas of social life, and the extent that reducing workplace size reduces disease spread. Moreover, we will examine the possible gain and cost by simple contact tracing and light quarantine practices.

Results: mitigation strategies

To illustrate the relative importance of the workplace and public life, we consider the scenarios in Fig. 2a. In the first scenario, nothing is done. In the second, contacts within the workplace are reduced by 75%, while in the third, contacts with friends and the public are reduced. Finally, we compare these with similar scenarios, but where good hygiene or keeping a distance reduces the probability of infection from all types of encounters by half.

In the figure, we see that the effects of reducing workplace and social contacts are roughly of the same magnitude. This reflects the assignment of 30% weight to each of these contact types. The slightly larger effect of social contacts reflects our assumption that these connections are less clustered than the workplace network. The two latter graphs show the scenarios where we both reduce infection probability within one group by 75% and overall infection probability by 50%. They show that an effective lockdown requires both restrictions of the time spent in the workplace and in the public sphere, and measures that reduce infection probability by increased hygiene and physical distancing.

The above results provide one useful piece of information. If the effect of workplace and social contacts are of the same order, it is of little importance which one is restricted. Ideally, both will be restricted for a period. However, when restrictions need to be lifted, authorities will primarily be able to control the workplace, whereas the social sphere relies on local social behavior. Obviously, it is economically more sustainable to lift the one with the largest social consequences first, by allowing people to return to work while encouraging keeping social gatherings at a minimum.

If restrictions are lifted before a substantial level of immunity is achieved, the epidemic will re-ignite. Therefore, we now examine what can be done to minimise spread in the reopened workplaces.

One possible strategy is to reduce the number of people allowed at any one time in each workplace. In Fig. 2b, we compare an epidemic scenario where the average number of employees per workplace is 10 with an epidemic where this number is reduced to 5. We further assume that the number of contacts per coworker remains the same, meaning that the number of contacts per person drops when workplace size is reduced.

It can be seen that fragmentation of physical spaces at workplaces could have a significant effect on the peak number of infected. In a situation with a risk of straining the healthcare system, this could be part of a mitigation strategy. Once again, the strategy becomes relatively more effective if the infection probability per encounter is also reduced. Compared to the cases with no workplace size reduction, making workplaces smaller leads to a greater relative reduction in peak size if infection probability is lower, completely eliminating the epidemic at an infection probability reduction of 50%.

A more local strategy that can be employed when reopening society is widespread testing and contact tracing. As mentioned above, Hellewell et al.[11] have suggested that this can be effective in containing COVID-19

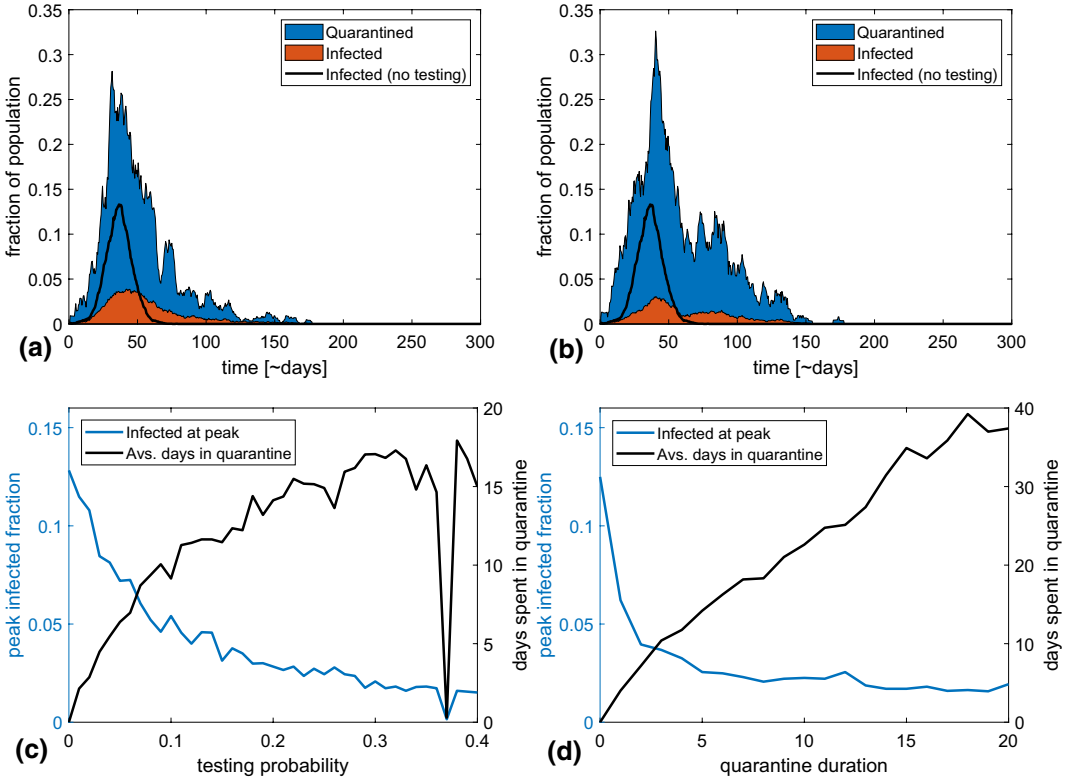


Figure 3. The effect of quarantine duration and testing probability. (**a**,**b**) show examples of epidemic trajectories for a quarantine length of 5 days and a daily testing probability of 10 and 20% respectively. The blue section of the curve shows the fraction of people who are in quarantine but healthy or presymptomatic, while the orange section shows the fraction who are ill. (**c**,**d**) show the peak fraction of population infected (left y-axis) and time spent in quarantine (right y-axis) as a function of testing chance and quarantine length. The number of days in quarantine was calculated using our standard group sizes which connect each person to approximately 20 others. The average quarantine time scales proportionally with this assumed connectivity. (**c**) With a quarantine length of 5 days, it is possible to reduce the peak number of infected by eight percentage points, corresponding to a 60% drop, if the probability of infected people being tested is only 10% per day of illness. However, the price of this is that each person is on average quarantined once during the epidemic. If testing is more widespread, the epidemic peak can be further reduced, until it finally becomes unstable at a testing probability of around 40% per day. (**d**) Epidemic peak and time spent in quarantine as a function of quarantine length for a testing probability of 20% per day. The average time spent in quarantine increases linearly with the length of quarantine. On the contrary, the effect of quarantine on the peak height appears to stagnate at approximately 5 days.

outbreaks provided high efficiency in detecting infected individuals. Contact tracing has previously been modeled in relation to other epidemics[18], and used successfully against smallpox[19] and SARS[20].

One obstacle to the widespread implementation of this strategy is the difficulty of tracing contacts. Therefore, we will here implement a crude form of contact tracing where we (1) close the workplaces of people who are tested positive for the disease, (2) isolate their regular social contacts for a limited period, and (3) keep symptomatic individuals in quarantine until they recover. We will see that such a 1 step tracing and quarantine strategy (1STQ) can give a sizeable reduction in disease spread while costing fewer lost workdays than overall lockdown. Our simulations include the limitations imposed by not being able to trace the estimated 15% of infections from random public transmissions. Thus, the strategy does not require sophisticated contact tracing but could be implemented based on infected people being able to recollect their recent face-to-face encounters with friends.

It should be noted that we here quarantine persons in their own households, thereby making our contact tracing strategy easier to implement in practice. In particular, family members of a quarantined person are still free to interact outside their home if they are not themselves tested positive. The drawback of such light quarantine practices is that infected persons in quarantine may still transmit the infection to their families.

Figure 3 examines how increased testing efficiency systematically improves our ability to reduce the peak disease burden. This would then be a more cost efficient way to mitigate the pandemic than a complete lockdown where each person would lose several man-months. Even detecting as little as 5% of COVID-19 infected per day (which with an average symptomatic disease duration of 3 days corresponds to finding approximately 15% of the infected) can potentially reduce the peak number of cases by 50%. If 10% efficiency is possible, corresponding to detecting about a third of infectious cases, then peak height could be reduced by a factor of almost three with

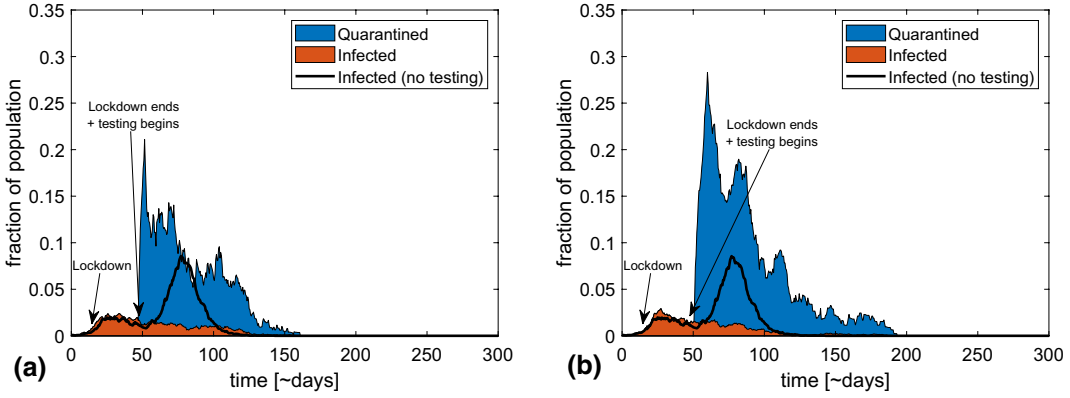


Figure 4. Various trajectories of the epidemic when combining a lockdown and a late-onset 1STQ strategy. (**a**) shows a possible course of an epidemic where restrictions in public and work life (by 75%) are implemented when 1% are infected and lifted after 30 days, being replaced by a testing and tracing regime with a testing probability of 20% per day and quarantine duration of 5 days. The black line shows the fraction of infected if no testing is implemented. We see that this level of testing and quarantine is sufficient to prevent a resurgence of the epidemic. (**b**) is similar, but here quarantine lasts 10 days. This is about as effective as (**a**) but costs a lot more in terms of number of people in quarantine.

less than two weeks in quarantine per person during the entire epidemic. This is illustrated in Fig. 3a where peak height is reduced from 0.13 to 0.04 at 10% testing efficiency.

The main cost of the quarantine option is the quarantine time. Figure 3d examines the efficiency versus cost of as a function of quarantine length. It can be seen that there is little gain in extending the quarantine period beyond the 5-day duration of the incubation period. For this reason we opted for 5 days in quarantine in panel (a, b). As a consequence, an average person will stay around 12 days in quarantine during the course of the epidemic with a testing probability of 10% per day. This time can be reduced if people can be convinced of smaller work environments and fewer face-to-face contacts per week. Fragmentation of our networks into smaller groups will reduce both quarantine overhead and the direct transmission of the disease (Fig. 2b, orange curve).

A prolonged lockdown will hugely disrupt society, and it is questionable whether a complete eradication of the virus is possible anyway. Therefore, most governments have aimed at softening the epidemic curve, with varying degrees of success. The one step contact tracing with testing and quarantine is a means to this end and would work most effectively in combination with other efforts to reduce R_0.

Finally, we investigate whether an aggressive testing and contact tracing strategy could work if implemented at a late stage in an epidemic. This could be relevant if for example the strategy is part of an effort to reopen society after a period of lockdown.

In Fig. 4, we show two possible scenarios where testing and contact tracing is implemented after a 30-day lockdown with a 75% reduction of the work and social spheres. The lockdown is initiated when 1% of the population is infected. In (a) we subsequently test and quarantine the infected and their contacts for 5 days, while in (b) the required quarantine is set to 10 days. We assume a testing efficiency of 20% chance of detection for each day a person is symptomatic. The progression of the epidemic without testing is marked by a black graph for comparison.

From the figure one sees that the strategy of even relatively short quarantines also works with a late onset. At a realistic detection probability, it prevents a resurgence of the epidemic. Nonetheless, it is quite costly initially, with a very high peak in number of quarantined people. Importantly, the effect does not increase with a longer quarantine period, but the cost is substantially larger.

Discussion

Pandemics such as the one caused by COVID-19 can pose an existential threat to our social and economic life. The disease itself is serious and leaves specific epidemic signatures and characteristics that make traditional contact tracing difficult. In particular it is highly infectious, can sometimes be transmitted already two days after exposure, and a large fraction of transmission happens before the onset of symptoms. As such it is difficult to contain without a system-wide lockdown of society. Nonetheless, a successful containment in South Korea used contact tracing. This motivated us to explore a one-step contact tracing/quarantine strategy (1STQ).

Using reasonable COVID-19 infection parameters we find that the 1STQ strategy can contribute to epidemic mitigation, in the sense that it can reduce the peak number of infected individuals by about a factor of two even with a realistic testing rate of 10% per day of illness. This was illustrated systematically in Fig. 3. The main cost was people in self-quarantine and not contributing to the workforce. In comparison one has to consider that a society-wide lockdown with similar reduction in peak height would have to last for about 100 days (see Fig. 2). Thus, the lockdown would require of order 100 days of quarantine (or at least extensive social distancing) per person, whereas testing and isolation only requires on average around 15 days per person with a 5-day quarantine

even at high testing probabilities. Importantly these numbers can be reduced if people are able to lower their number of contacts.

A noticeable objection to the 1STQ strategy is the fraction of cases with so weak symptoms that people do not contact health authorities. The effect of such limitations is in our model parameterized through the detection probability. From Fig. 3c one sees that when the detection probability goes below 3% (a rate of 1% per day) the peak reduction of the 1STQ strategy becomes only of the order 1 percentage point. It should also be noted that, since we rely on symptoms to determine who stays in quarantine, and people in the infectious/symptomatic stage are assumed to always stay in quarantine, we implicitly assume that all infected persons develop at least some symptoms at some point. This may be a break from reality.

The increasing availability of tests may also change the perspectives of the 1STQ strategy. With widely available rapid tests, it will be possible to test everyone regularly, and to test all quarantined persons before they leave quarantine. Supplementary Figure S4 deals with the results of such a testing strategy and finds that it makes it possible to totally control the epidemic, or to mitigate it without quarantining any healthy individuals. To put this into perspective, the drawbacks of widespread, but slow testing is examined in the supplementary Fig. S5. Here, we find that the 1STQ strategy is most efficient with no test delay, and that delayed contact tracing is comparable to a primitive lockdown.

One interesting point which we have not examined here, is that real-world social networks are heterogeneous, with a large variance in number of contacts. It may be expected, for example, that workers in customer-facing positions in shops will have a high risk of catching the disease and passing it on. The effects of this heterogeneity is examined more closely in Ref.[21] Here, it is concluded that heterogeneity in the number of contacts enhances the effect of contact tracing, since persons with many contacts are both more likely to pass on the disease and more likely to be quarantined.

In Ref.[11], the authors suggest a 1STQ strategy similar to the one we here model. The main points of the present analysis is the focus on mitigating instead of eradicating the epidemic, our suggestion of a shorter quarantine length, and the implementation of quarantine together with other members of the household instead of total isolation. Our stochastic, agent-based approach also allows for local failures due to the limited duration of quarantine (people may not yet be symptomatic when exiting quarantine) and the non-traceable public contacts (set to 15%).

Finally, one noticeable finding is that contact tracing and reduction of contacts per person is still feasible even at a later stage of the epidemic. As can be seen in Fig. 4, a lockdown and subsequent reopening with testing and contact tracing is highly effective in controlling the epidemic. Our study that lockdowns have an important role to play in epidemic mitigation, but that they can be replaced by a 1STQ strategy once the epidemic is under control.

The COVID-19 pandemic has set both governments, health professionals, and epidemiologists in a situation that is more stressful and more rapidly evolving than anything in recent years. Due to the uncertainties caused by a situation in flux, it is difficult to predict anything definite about what works and what does not. The empirical observation that lockdowns worked in both China, and in a milder form in Denmark shows that our assumption of a 75% reduction in specific infection rates under lockdown is realistic. Our main result is that some of these restrictions can be replaced by testing, one-step contact tracing and short periods of quarantine. This is far cheaper than total lockdowns. Perhaps most importantly, these measures work best in combination. As is highly relevant to the current epidemic stage of COVID-19, we pinpoint that 1STQ can be successfully implemented also at a late stage of the epidemic where testing may become massively available.

Data availability

Plots of alternative variants of our model (including alternative testing strategies and larger family sizes) can be found in the supplementary material. The code used to produce the plots shown in this article is available on Figshare under the URL https://doi.org/10.6084/m9.figshare.12206735.v4.

Received: 24 April 2020; Accepted: 16 October 2020
Published online: 29 October 2020

References

1. Anderson, R. M., Heesterbeek, H., Klinkenberg, D. & Hollingsworth, T. D. How will country-based mitigation measures influence the course of the covid-19 epidemic?. *Lancet* **395**, 931–934 (2020).
2. Ferguson, N. *et al.* Report 9: impact of non-pharmaceutical interventions (NPIS) to reduce Covid-19 mortality and healthcare demand. *Imp. Coll. Lond.* **10**, 77482 (2020).
3. Mossong, J. *et al.* Social contacts and mixing patterns relevant to the spread of infectious diseases. *PLoS Med.* **5**, e74 (2008).
4. Peng, L., Yang, W., Zhang, D., Zhuge, C. & Hong, L. Epidemic analysis of covid-19 in china by dynamical modeling. *arXiv preprint* arXiv:2002.06563 *(2020).*
5. Prem, K. *et al.* The effect of control strategies to reduce social mixing on outcomes of the COVID-19 epidemic in Wuhan, China: a modelling study. *Lancet Public Health* **5**, e261–e270 (2020).
6. Chang, S. L., Harding, N., Zachreson, C., Cliff, O. M. & Prokopenko, M. Modelling transmission and control of the covid-19 pandemic in australia. *arXiv preprint* arXiv:2003.10218 *(2020).*
7. Li, R. *et al.* Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (sars-cov2). *Science* **368**, 489–493 (2020).
8. Klepac, P., Kissler, S. & Gog, J. Contagion! the BBC four pandemic: the model behind the documentary. *Epidemics* **24**, 49–59 (2018).
9. Klepac, P. *et al.* Contacts in context: large-scale setting-specific social mixing matrices from the BBC pandemic project. *medRxiv* https://www.medrxiv.org/content/medrxiv/early/2020/03/05/2020.02.16.20023754.full.pdf (2020).
10. Li, Q. *et al.* Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. *New Engl. J. Med.* **382**, 1199–1207 (2020).
11. Hellewell, J. *et al.* Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts. *Lancet Glob. Health* https://doi.org/10.1016/S2214-109X(20)30074-7 (2020).

12. Fam44n: Families 1. January by municipality, type of family, size of family and number of children. *Stat. Den.* https://statistikbanken.dk/statbank5a/SelectVarVal/Define.asp?Maintable=FAM44N&PLanguage=1 *(Accessed October 7 2020).*
13. Nishiura, H., Linton, N. M. & Akhmetzhanov, A. R. Serial interval of novel coronavirus (COVID-19) infections. *Int. J. Infect. Dis.* **93**, 284–286 (2020).
14. Coronavirus disease 2019 (covid-19) pandemic: increased transmission in the EU/EEA and the UK—seventh update. *Eur. Centre Dis. Control. Prev.* https://www.ecdc.europa.eu/en/publications-data/rapid-risk-assessment-coronavirus-disease-2019-covid-19-pandemic (2020).
15. Akhmetzhanov, A. R., Mizumoto, K., Jung, S. M., Linton, N. M., Omori, R., & Nishiura, H. Epidemiological characteristics of novel coronavirus infection: a statistical analysis of publicly available case data. *medRxiv* https://doi.org/10.1101/2020.04.24.20077800 (2020).
16. Flaxman, S. *et al.* Report 13: estimating the number of infections and the impact of non-pharmaceutical interventions on COVID-19 in 11 European countries. *Imp. Coll. Lond.* https://doi.org/10.25561/77731 (2020).
17. Remuzzi, A. & Remuzzi, G. COVID-19 and Italy: what next?. *Lancet* **395**, 1225–1228 (2020).
18. Klinkenberg, D., Fraser, C. & Heesterbeek, H. The effectiveness of contact tracing in emerging epidemics. *PLoS ONE* **1**, e12 (2006).
19. Fenner, F. *et al. Smallpox and Its Eradication, No. 6* (World Health Organization, Geneva, 1988).
20. Donnelly, C. A. *et al.* Epidemiological determinants of spread of causal agent of severe acute respiratory syndrome in Hong Kong. *Lancet* **361**, 1761–1766 (2003).
21. Nielsen, B. F., Sneppen, K., Simonsen, L. & Mathiesen, J. Heterogeneity is essential for contact tracing. *medRxiv* https://doi.org/10.1101/2020.06.05.20123141 (2020).

Acknowledgements

We thank Gorm Gruner, Bjarke Frost Nielsen, Andreas Roepstorff, and Lone Simonsen for enlightening discussions. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program under Grant Agreement No. 740704.

Author contributions

A.E. and K.S. both participated in devising the model. Code was written and plots produced by A.E.. The functionality of the code was checked by comparison with an alternative algorithm written by K.S. A.E. and K.S. wrote and edited the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41598-020-75640-2.

Correspondence and requests for materials should be addressed to A.E.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Supplementary figures for the article "Cost-benefit of limited isolation and testing in COVID-19 mitigation"

Andreas Eilersen[1,*,+] **and Kim Sneppen**[1,+]

[1]University of Copenhagen, Niels Bohr Institute, 2100 København Ø, Denmark
[*]andreaseilersen@nbi.ku.dk
[+]these authors contributed equally to this work

The effects of family size

To examine the effect of the family size distribution on our conclusions, we have repeated the figures of the main article with twice as large families. That is to say, single households now contain two people, previously two-person households contain four etc. We see that while this somewhat affects the shape of the epidemic curve, it does not significantly influence any of our conclusions.

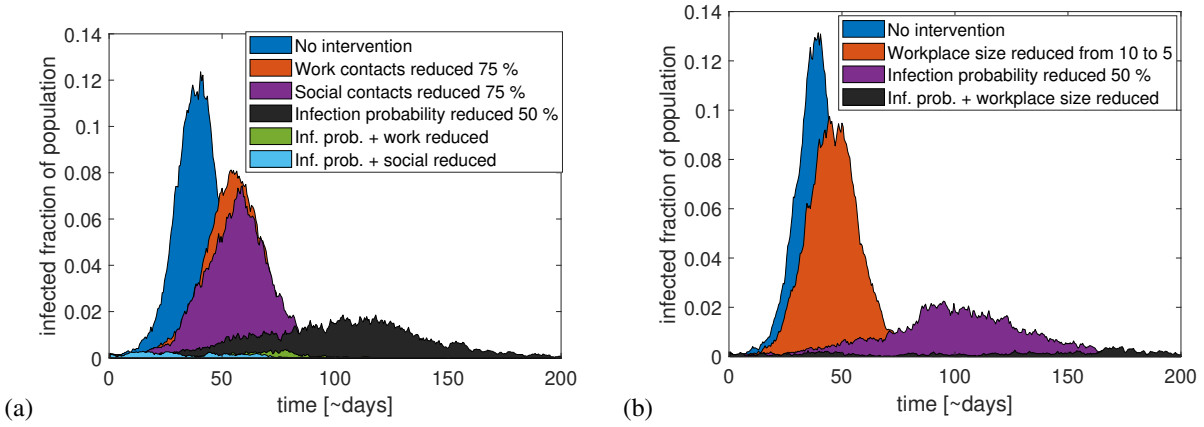


Figure S1. (a) A comparison of various containment strategies that include a lockdown reducing social or work contacts by 75 %. We see that if we combine these measures with improved hygiene or social distancing, reducing the transmission risk per encounter by half, this is sufficient to stop the epidemic completely. This is the same conclusion as in the main article. (b) The effect of reducing workplace sizes by half, from 10 to 5 people per workplace on average. As in the main article, this has a significant effect on the epidemic, and in this version of the model, it is also enough to fully mitigate it if combined with hygienic measures that halve the infection probability.

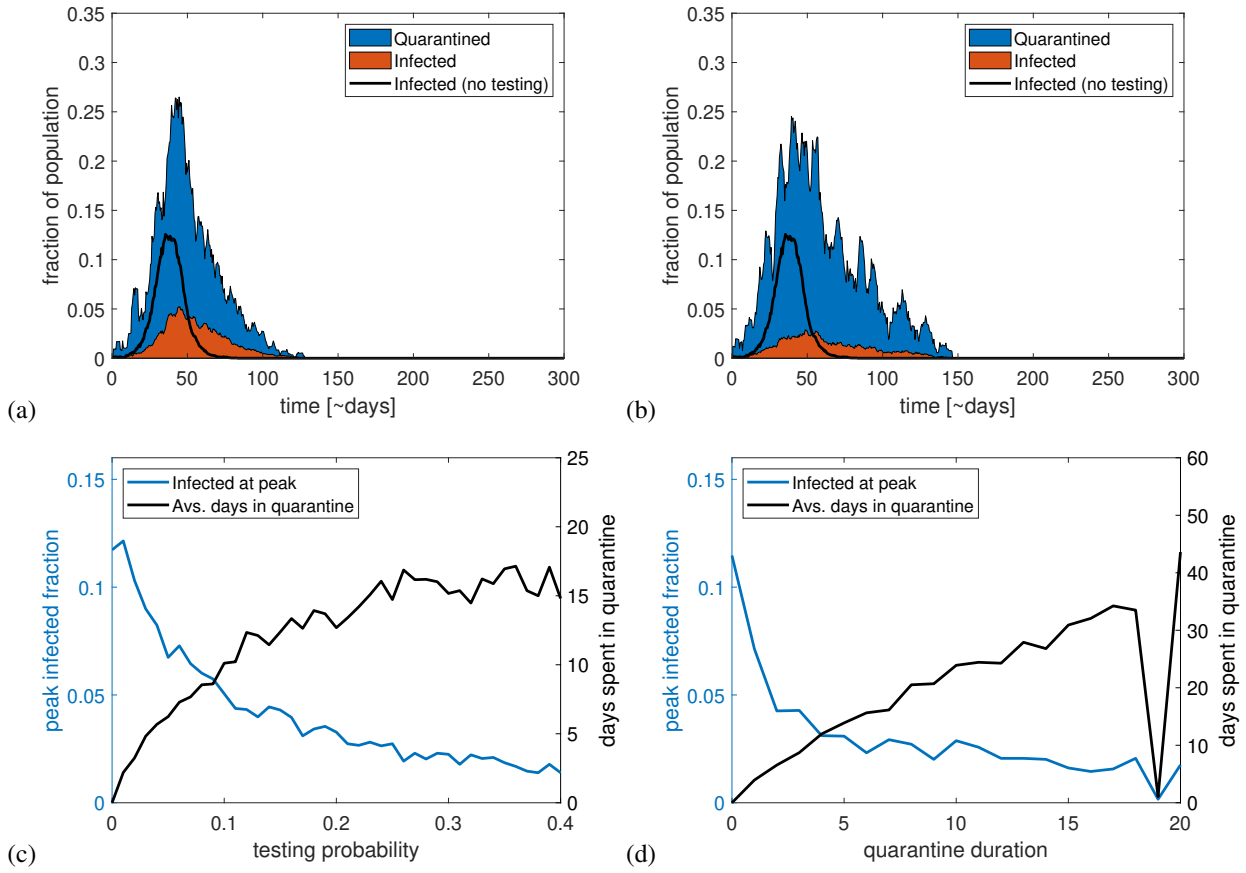


Figure S2. (a) and (b) show epidemic trajectories and fraction of people in quarantine for testing probabilities of 10 and 20 % per day respectively. The black line shows the epidemic trajectory in the absence of testing and contact tracing. (c) shows the infected fraction of the population at the peak of the epidemic (left axis) and the average number of days each person spends in quarantine during the epidemic as a function of daily testing probability while symptomatic. (d) shows the same variable but as a function of quarantine duration. It can be seen here as well that little is gained from quarantine beyond five days. When quarantine lasts longer than about 18 days, the epidemic becomes unstable, sometimes dying out.

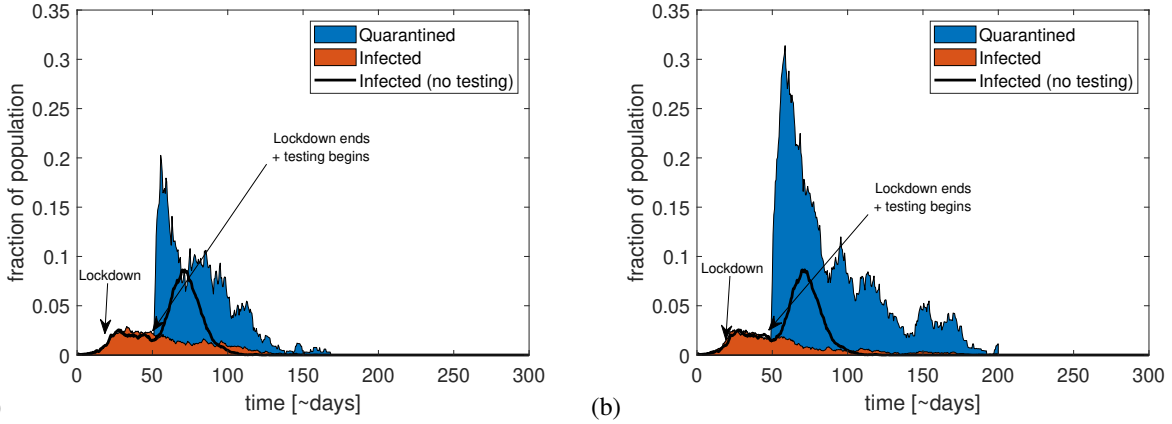


Figure S3. Two examples of epidemic trajectories when combining a lockdown with a 1STQ strategy. When 1 % of the population is infected, a lockdown is implemented, reducing social and work contacts by 75 %. It is lifted after 30 days and replaced by a testing and contact tracing strategy with a daily testing probability of 20 % for symptomatic individuals. The duration of the quarantine in panel (a) is five days and in (b) it is ten days. We see that the longer quarantine does not change the effect on the epidemic much, but it does increase the number of people in quarantine.

Weekly tests

In this section, we let all agents get tested regularly at a one-week interval in addition to a 20 % daily testing probability for symptomatic individuals. This strategy will become increasingly feasible with increasing availability of rapid tests. It can be seen that weekly tests are sufficient to contain the epidemic if contacts of the infected are quarantined (fig. S4(a)). If only the infected themselves are quarantined, it is still enough to mitigate the epidemic as seen in panel (b), giving us a situation where no healthy persons are unnecessarily quarantined. With as widespread testing as this, it is assumed that everyone is tested before leaving quarantine, meaning that no presymptomatic individuals are let out of quarantine.

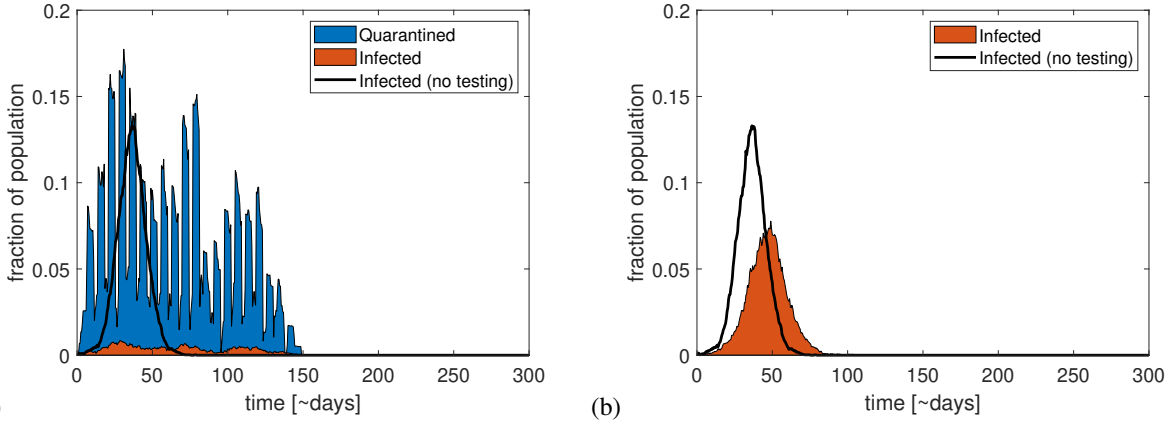


Figure S4. (a) Infected and quarantined fraction of the population if everyone, in addition to tests of the symptomatic, take one weekly test, thus also discovering any presymptomatic cases. This is enough to keep the epidemic in check. In (b) the same strategy is followed, but only the infected themselves are quarantined. This still significantly reduces the epidemic peak. The unmitigated epidemic trajectory is shown by the black curves.

Delayed test results

In the following, we examine the effect of a delay in obtaining test results. We assume that people who are tested have to wait for a number of days before getting the result, thus delaying contact tracing and quarantine. We find that the effect of contact tracing only subsides slowly with increasing test delay, as seen in fig. S5. This we believe is due to the effect of the large number of people in quarantine at the peak of the epidemic. Even though the contact tracing efforts are less effective when test results are delayed, having a large fraction of the population in quarantine works as a primitive lockdown, lowering the epidemic peak by "brute force".

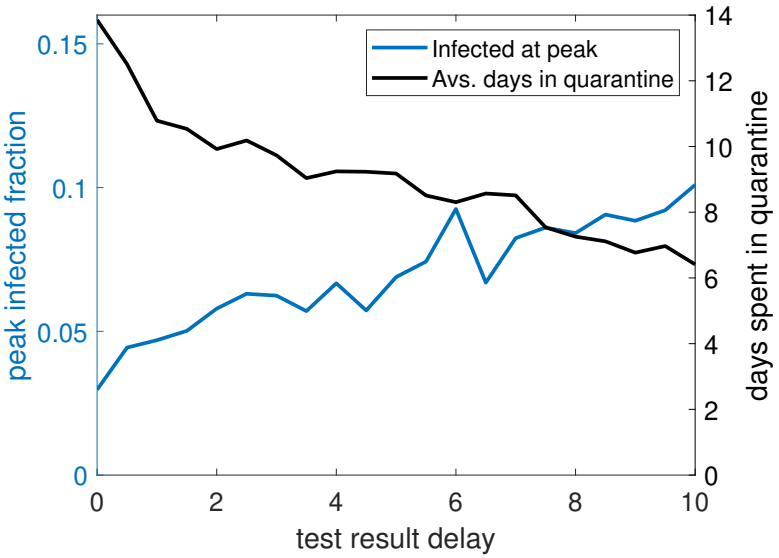


Figure S5. This figure shows the effect of delayed test results on the peak fraction of infected and the average number of days spent in quarantine during the epidemic. We see that the peak fraction of infected grows linearly with the delay, until the effect is nearly gone after ten days. At the same time, the amount of time spent in quarantine decreases. Part of the mitigation at long delays is expected to stem from the fact that a significant portion of the population is quarantined, and therefore less infectious, at all times. The testing probability is here set to 20 % per day of symptomatic illness and the quarantine length is set to five days. The testing delay is assumed to not affect people's ability to leave quarantine.

SARS-CoV-2 superspreading in cities vs the countryside

Authors: Andreas Eilersen* and Kim Sneppen*

*The Niels Bohr Institute, University of Copenhagen, Copenhagen, Denmark

My contributions: I contributed to developing and programming the model, to the production of figures and data processing, and to the writing of the article.

Publication status: Published in *APMIS* (2021).

doi: 10.1111/apm.13120

APMIS JOURNAL OF PATHOLOGY, MICROBIOLOGY AND IMMUNOLOGY

APMIS 129: 401–407

DOI 10.1111/apm.13120

SARS-CoV-2 superspreading in cities vs the countryside

ANDREAS EILERSEN and KIM SNEPPEN

Niels Bohr Institute, University of Copenhagen, København Ø, Denmark

Eilersen A, Sneppen K. SARS-CoV-2 superspreading in cities vs the countryside. APMIS. 2021; 129: 401–407.

The first wave of the COVID-19 pandemic was characterized by an initial rapid rise in new cases followed by a peak and a more erratic behaviour that varies between regions. This is not easy to reproduce with traditional SIR models, which predict a more symmetric epidemic. Here, we argue that superspreaders and population heterogeneity would predict such behaviour even in the absence of restrictions on social life. We present an agent-based lattice model of a disease spreading in a heterogeneous population. We predict that an epidemic driven by superspreaders will spread rapidly in cities, but not in the countryside where the sparse population limits the maximal number of secondary infections. This suggests that mitigation strategies should include restrictions on venues where people meet a large number of strangers. Furthermore, mitigating the epidemic in cities and in the countryside may require different levels of restrictions.

Key words: COVID-19; model; superspreading; population density; epidemiology.

Kim Sneppen, Niels Bohr Institute, University of Copenhagen, Blegdamsvej 17, 2100 København Ø, Denmark. e-mail: sneppen@nbi.ku.dk

At its onset the coronavirus disease 2019 (COVID-19) pandemic shocked the world, with the number of new cases and deaths growing more than 20% per day in the main hotspots [1]. With a growth rate this high, the disease was expected to spread through the population in less than six months without mitigation, and to reach a peak after three months, at which point 30% of the population would have had the disease [2]. This, however, was not how the initial wave of the epidemic played out [3].

While most of the epidemic undoubtedly was halted due to mitigation efforts, it is striking that even in countries that have implemented a very light lockdown, such as Sweden, the epidemic peaked long before herd immunity was achieved. Furthermore, even societies that slowly reopened businesses and public life did not experience an immediate explosive resurgence of the epidemic expected given the low levels of immunity and the speed with which the disease spread initially [3].

Here, we will propose an agent-based lattice model of an infectious disease that spreads in a geographically heterogeneous population. The model is a simplified depiction of the dynamics of viral infection with severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), with its characteristic parameters. We will examine the effect of the heterogeneous infection pattern that is so characteristic of this disease, using a gamma distributed infectiousness with dispersion factor $k = 0.1$ from [4]. Infection heterogeneity is a feature of several epidemic diseases [5] and plays a particularly important role for COVID-19 [4,6–8]. In individual events, a single person has caused dozens of infections [9]. At the same time, the attack rate within households has been reported to be very low, at less than 20% [10], despite prolonged close contact. This suggests that the majority of COVID-19 patients infect very little. In [11] an agent-based model was used to demonstrate that the Achilles heel of an epidemic driven by superspreaders was public social contacts, while the repeated contacts to smaller family and work groups were less dangerous.

Looking at COVID-19 data from the first wave in the United States in the analysis by [12], it is clear that a rapidly spreading epidemic occurs primarily in densely populated areas. In less densely populated areas, the epidemic onset is delayed, and in rural areas, the epidemic never really starts, with most cases appearing to be spillover from the cities. The daily per capita mortality at the first peak in [12] varies by a factor of ten between the most and least densely populated areas in the United States.

Received 27 November 2020. Accepted 20 January 2021

The density dependence of COVID epidemics could in principle be explained by the higher chance of meeting infected people in dense areas. However, only 1% of COVID-19 spread was outdoors in China [13], suggesting that only indoor meetings count. Furthermore, only few infections happen within households [10], suggesting that it is social visits and meetings in confined areas that facilitate infection [14]. We will base our model on these observations.

METHODS

Our model plays out on a lattice of side L with periodic boundary conditions. Each lattice site may either be empty or contain one agent. The agents can be in one of three states, susceptible, infectious or recovered. Individuals interact with their neighbours with a frequency f_{meet} which is constant and fixed along with the mean infection probability per meeting to give the desired number of secondary infections at the onset of the epidemic, R_0. The distance that agents travel we draw from the distribution $p(r) = \frac{1}{r_0} e^{-r/r_0}$ with a mean of $r_0 = 10$ sites unless stated otherwise. The real probability distribution of travel distances for cars has in Italy been determined to be an exponential function below 20 km, and then a power law with a steeper cut-off around 500 km [15]. The interaction radius r_0 is not by itself a meaningful parameter. Rather, it is the number of neighbours within this radius, given by $\pi\rho r_0^2$ (where ρ is the population density) which determines the behaviour of the system, as will be demonstrated below. A diagram of the model can be seen in Fig. 1.

COVID-19 is known to be transmitted before the onset of symptoms [16]. For simplicity, rather than using the susceptible-exposed-infectious-recovered (SEIR) modelling framework, we combined the exposed, presymptomatic state and the overlapping infectious period into one single infectious period of $1/\gamma = 10$ days. Given the geometry of the system, this leads to a measured serial interval of approximately 6 days, slightly higher than most sources estimate [17,18]. However, it is a small difference, and given the simplifications of our model we judged that it would not be meaningful to fit this parameter precisely to the observed value. Infected agents randomly leave the infectious state with a rate of γ per day, meaning that the duration of the disease is exponentially distributed. The implications of using a different distribution of disease durations are explored in the supplement.

When an infectious agent i interacts with a susceptible agent, the susceptible agent will become infected with a probability p_i that is specific to the infecting agent. We draw these probabilities from a gamma distribution with dispersion factor $k = 0.1$ [4,5], within the range observed for COVID-19 [4,6,8]. The distribution is normalized to give an average reproduction number R_0 of 3 at a population density of 1 at the onset of the epidemic.

The geographical heterogeneity of the population is modelled by placing a square 'city' of side $L/5$ on a lattice with periodic boundary conditions and lattice size L. The city has the population density 1, that is all sites are occupied, and the city population is thus $L^2/25$. The city is surrounded by 'countryside' with a population density and total population of 24 $L^2/25$.

Importantly, we assume that the rate of interactions per agent is kept the same in both city and countryside, meaning that we assume that people are equally social. If an agent in the countryside attempts to interact with an empty site, the attempt is counted as failed, and new attempts are made until the number of contacts is the same as in the city, where all sites are occupied.

Thus, people in the countryside interact with a smaller set of people while still spending the same amount of time on social activities. In a wider perspective, this proposes that density dependence of disease spreading is more due to difference in diversity of contacts than due to differences in time spent around other people. Thereby our model assumes an infection rate that depends on density, but not in a simple linear fashion as sometimes assumed [19].

We seed the disease within the city. Even if we were to seed it randomly, the city would usually be hit early on provided that the epidemic catches on.

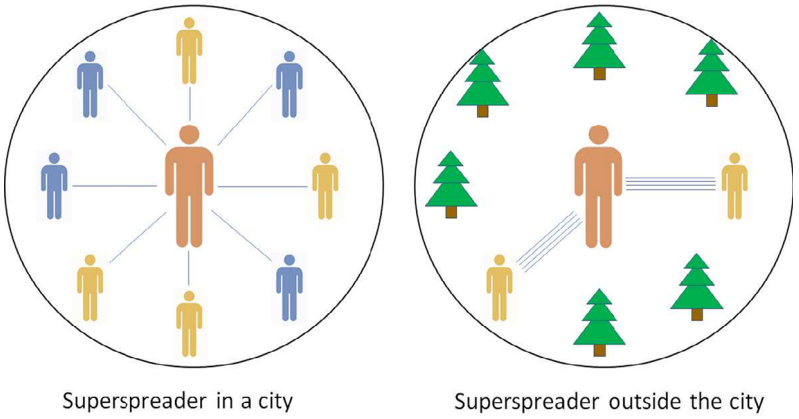


Fig. 1. Model: A superspreader in a city interacts a little with a lot of people and will infect some fraction of them. On the other hand, a superspreader outside the city will interact a lot with each of a smaller set of people. The superspreader then infects practically all of them, but there is a lower cap on the number of secondary infections.

RESULTS

In Fig. 2, we consider a disease with heterogeneous infection rates and study how simulated epidemic trajectories depend on population density. One sees that the dynamics resembles that of a SIR model at values much lower and much higher than the critical density. In the low regime, the epidemic only spreads in the city which is relatively well-mixed, whereas in the high regime, the countryside begins to resemble the city more and more. In both extremes, the fraction of infected individuals rises and falls symmetrically. The figure illustrates how the epidemic is 'stretched out' in the intermediate density range. The epidemic has the longest lifetime when is just above the percolation threshold $_{\text{crit}}$, as the disease still spreads in the countryside, but is nonetheless slowed down by the lower population density.

In Fig. 3A,B, we measure the attack rate of the epidemic in a homogeneously distributed population in order to find the percolation threshold, below which the epidemic will stop propagating. Panel (A) identifies $_{crit} \approx 0.01$ for a homogeneously spreading disease. In contrast, a disease with an overdispersion of $k = 0.1$ has a much higher critical density $_{crit} \approx 0.04$, as seen in (B). The same figure demonstrates that it is not the density alone that determines the ability of the disease to percolate, but rather the number of neighbours, proportional to r_0. The overdispersed simulation and the simulation with homogeneous infection were done with same average disease transmission rate and the factor ~ 4 difference in critical density comes about because a disease with $k = 0.1$ has 10% of the infected being responsible for 80% of the infections. Thus, most people do not transmit the disease, and it is therefore the density of the few people who do spread the disease that sets the critical threshold.

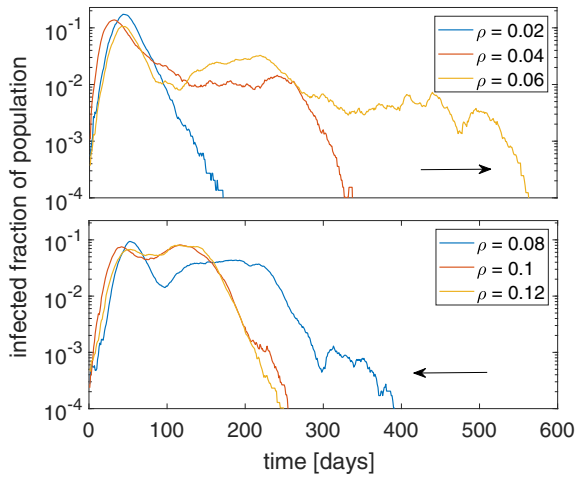


Fig. 2. Epidemic trajectories: Infected fraction of the population over time changes with the countryside population density ρ. In the low- regime, increasing the population density stretches the curve, as the epidemic spreads further from the city. When ρ is above ~ 0.06, the epidemic again approaches the behaviour of a SIR model, as the epidemic now spreads unhindered across the whole system. Around ρ_{crit}, there is a large variation in the duration of the epidemic. The parameters used are $\gamma = 0.1$, $r_0 = 10$, $f_{meet} = 10$ and dispersion parameter $k = 0.1$.

Since our model analysis centres on superspreaders as a main driver of the epidemic, Fig. 3C compares epidemics with and without superspreaders. It can be seen that with no superspreaders, the epidemic will spread unhindered in the countryside, albeit more slowly since the countryside is geographically larger than in the city.

This leads to a graph similar to two superimposed SIR-like models. If superspreaders are present, however, the epidemic may spread both slowly and erratically in the countryside and continue long after herd immunity is achieved in the city.

The dependence of the percolation threshold on the dispersion parameter k is further investigated in Fig. 4, which shows the attack rate as a function of both the number of neighbours and k. The figure shows that the percolation threshold increases drastically at low k, meaning that a superspreader-driven epidemic requires more social contacts per person in order to spread. Once the epidemic becomes sufficiently overdispersed ($k < 0.05$), it is no longer viable.

In Fig. 5, we try to replicate the data compiled by [12] and see that local disease incidence in a model with heterogeneous infectivity is indeed much more population density dependent than a model assuming homogeneous infectivity. This fits well with the cited data, which suggest a strong dependence of COVID-19 incidence on population density. It has already been known for years that the spread of epidemics is population density dependent [19]. Here, we show that this dependence is enhanced by heterogeneous infectivity. Importantly, as opposed to traditional disease models, we assume that everyone is equally social, but that the set of available contacts is smaller in sparsely populated regions. The significance of this will be discussed further below.

The delayed onset and erratic behaviour of the countryside epidemic obviously depend on the density and other characteristics of the countryside and the assumed travelling pattern of individuals. Also, real countryside contains a diverse pattern of smaller and large settlements. Therefore, we considered

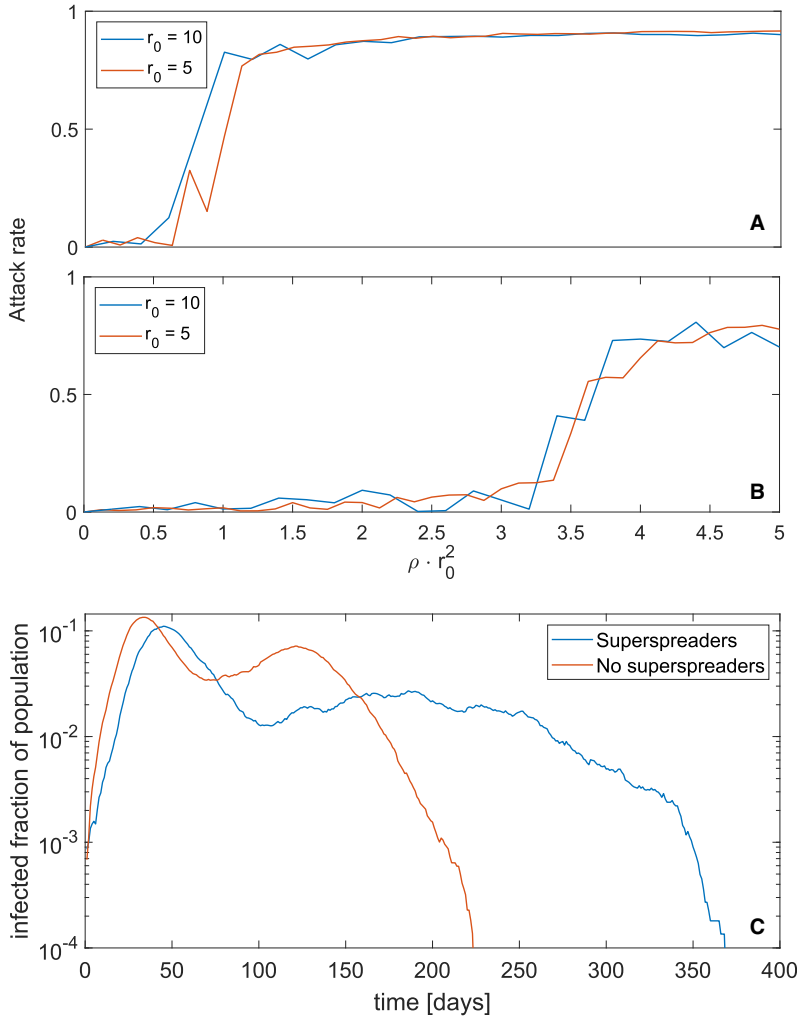


Fig. 3. Comparison of models without and with superspreaders. Panel (A) shows the attack rate as a function of the number of neighbours within the radius of interaction ($\sim \rho r_0$) in a population where everyone infects with the same rate. (B) shows attack rate with heterogeneous infection rates, using a gamma distribution with dispersion factor $k = 0.1$. The two overlaid curves demonstrate that the parameter r_0 does not affect the physics of the system, and what really determines the ability of the disease to percolate is the number of neighbours, proportional to ρr_0. (C) Epidemic trajectory when superspreaders dominate (blue) and when infectiousness is evenly distributed (red) for equal countryside population density and radius of interaction ($\rho r_0 = 6$). When superspreaders are the main drivers of the epidemic, it is strongly impeded once the city has reached herd immunity. When everyone infects equally, the epidemic simply spreads radially out from the city, leading to a 'second wave' in the countryside. Parameters are as in Fig. 2.

a system with several cities distributed on a lattice with side length $L = 1000$ (Fig. 6). The system mimics the observed city size distribution, which is fairly close to the Zipf law [20,21]. The random distribution of smaller cities is not entirely naturalistic, since evidence suggests that real cities are organized in a fractal pattern [22,23]. A figure using a city distribution closer to a fractal can be found in the supplement. For illustrative purposes, we chose a density in the countryside that is close to but below the percolation threshold for the disease ($\rho = 0.03$). With this below-critical spreading, we observe cities that are spared and cities that are nearly completely infected. In reality assuming near critical spreading in the countryside would not be necessary for global spread, since people occasionally travel long distances [15], facilitating rare direct transmissions between distant cities.

DISCUSSION

Based on the above, we suggest that the lopsided appearance of COVID-19 epidemic curves can be

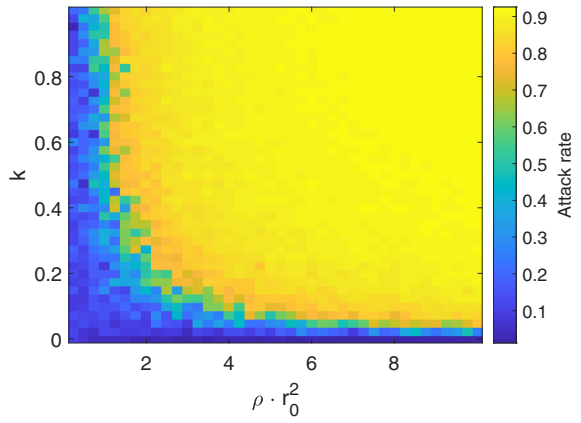


Fig. 4. Dependence of attack rate on density and k. Since the variable determining percolation is not the absolute density, but the number of neighbours, we plot ρr_0 on the x-axis rather than ρ. It is seen that the disease percolates much more easily at a higher k, which implies a more homogeneous infectivity. The more overdispersed the disease (corresponding to lower k), the more neighbours are required for percolation.

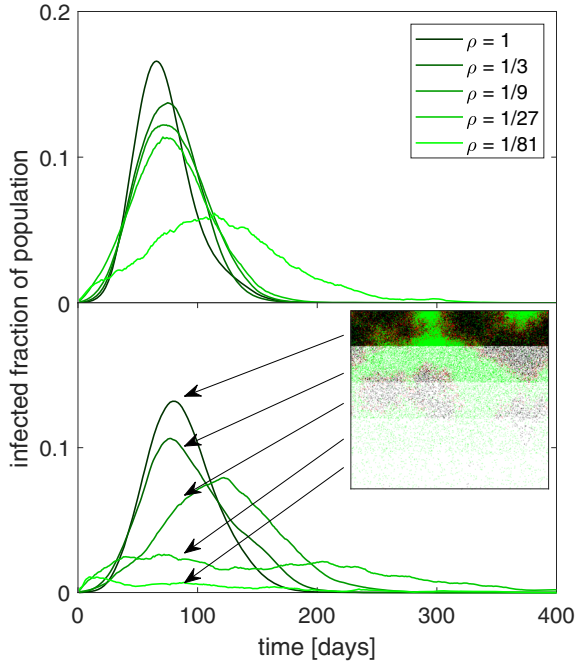


Fig. 5. Epidemic trajectory in regions of varying density. When infectivity is homogeneous (top), the epidemic is a lot less sensitive to a lower population density than when the epidemic is driven by superspreaders (bottom). Here, the epidemic is nearly absent in the low-density regions and appears to be driven by spillover. Inset illustrates the layout of the lattice.

explained by heterogeneous infection ability combined with a geographically heterogeneous population. However, our model makes a number of

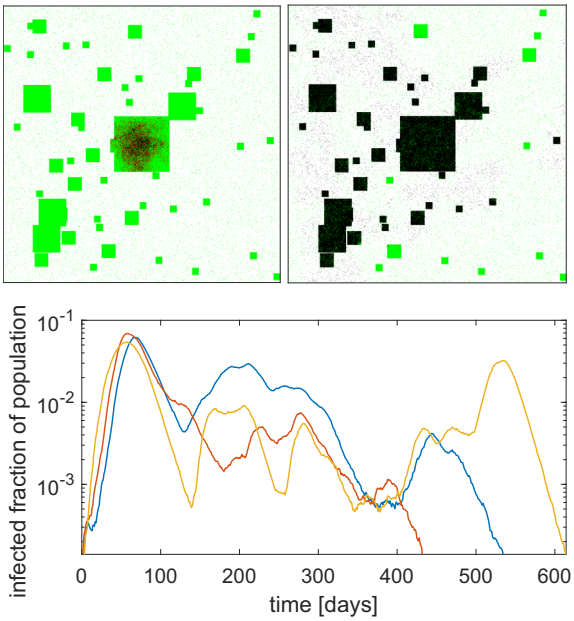


Fig. 6. Simulation with multiple cities: Infected individuals are shown in red, susceptibles are green, empty sites are white and recovered are black. City size distribution mimics Zipf's law [23], such that there is one city of with 40,000 inhabitants, 4 with 10,000, and so forth. The graphs show the fraction of the population currently infected as a function of time for three example runs of the simulation. Here, $\rho = 0.03$ and the other parameters are the same as the above figures.

assumptions and breaks from reality, whose importance must be discussed.

First and foremost, the distribution of full and empty lattice sites does not represent the geographical distribution of people as such. Rather, it represents the density of contacts. The density ratio between people in the countryside compared to an inner city is much lower than what is used in our model. On the other hand, persons living outside the cities may be more mobile than city-dwellers. The density contrast of agents on our lattice represents the combined effect of these factors.

A further complication is the distribution of disease duration and incubation periods. We here assume a simple exponential distribution of infectious period duration, whereas the real mechanics of COVID-19 includes presymptomatic transmission and a broad gamma distribution of incubation periods [17]. A different infectious period distribution might complicate our findings. Therefore, we examine the effect of a gamma distributed infectious period duration in the supplement. We find that, while changing this distribution has an effect on the percolation threshold, it does not change our fundamental conclusions.

It has long been discussed how disease transmission rate depends on population density [19,24,25]. We here present a new way of looking at this problem. A classical SIR model with pseudo-mass action transmission would assume a simple linear dependence of R_0 on population density, implicitly assuming that people become more social by living in a densely populated area.

Our model makes a different assumption: People will be equally social regardless of population density, but when population density is lower, the groups that they spend time with will be less diverse. The number of possible unique contacts for each individual declines linearly with population density, but people in sparsely populated areas are likely to have multiple encounters with the same persons, as their communities are smaller.

If, however, the epidemic is dominated by superspreaders who only need one or a few encounters to transmit the disease, the duration of each encounter becomes less important as even a rather brief contact to a superspreader is likely to lead to infection. Instead, what limits the action of superspreaders is their number of unique contacts. Superspreaders that interact with only a small, tight-knit group will inherently be highly limited in how many secondary infections they can generate.

If superspreading makes a disease vulnerable to variations in population density, we should conversely expect to see that diseases with a homogeneous infectivity, that is a dispersion factor close to or above 1, exhibit little variation with population density. One example of a disease with a homogeneous infectivity is influenza, with an estimated dispersion factor of $k = 0.94$ for the 1918 pandemic flu [26]. When examining the incidence of seasonal influenza, which we assume to have a similar dispersion factor, [27] find no consistent variation with population density. This is a point in favour of the link between superspreading and population density dependence.

Finally, the effect of lockdowns and changes of social behaviour is important. A previous paper [11] suggests that even moderate mitigation may limit the action of superspreaders by reducing the maximal number of people any person can be in contact with. If this hypothesis is true, bans on gatherings and a reduction in public social life would lead to early peaks in the number of new cases. Our study compounds this finding and suggests that a change in behaviour is not strictly necessary to cause an epidemic peak well before herd immunity has been achieved. Mitigation strategies that primarily target cities may well be sufficiently effective in bringing down the epidemic. However, large events like funerals, weddings or festivals are not included in our model and will of course facilitate spreading in any location. Finally, if there is a large difference in attitude towards the disease in the city and the countryside, this can significantly change the behaviour of the epidemic in the countryside relative to the city. All in all, how governments and populations have responded to the pandemic has likely had a crucial effect on its trajectory. Our model investigates how it would have played out in the absence of these complicating factors.

CONCLUSION

Despite some caveats, our model reproduces the main aspects of geographical heterogeneity and suggests a new view on density dependence of epidemic dynamics. An epidemic with a large heterogeneity in infection rates is predicted to be most intense in large cities while it slowly tapers off in the countryside. This is consistent with what we see in data from the first half year of the COVID-19 pandemic [12,28]. Our results thus favour the hypothesis that the COVID-19 pandemic is driven by superspreaders, and that the observed quick exponential growth phase, early peak, and slow, erratic recovery phase are in part consequences of combining heterogeneity of infectivity with a heterogeneous population density.

Of course, restrictions limiting social contact were crucial to the overall COVID-19 pandemic, and behavioural societal changes remain a central part of a pandemic that still after a year is far from reaching herd immunity. Our model predicts that anything that limits the diversity of contacts, be it population density or limits on gatherings, will have an outsized effect on disease spread. In comparison, simply limiting the number or duration of contacts will not be as effective unless the diversity is also decreased. A superspreader still spreads quite effectively given ten encounters with ten different individuals, but is much less effective given 20 encounters with only five different people. The present work suggests that future epidemiological models should take contact diversity into account if they are to properly capture the dynamics of COVID-19 and other diseases with large transmission heterogeneity.

We thank Lone Simonsen, Viggo Andreasen and Bjarke Frost Nielsen for enlightening discussions.

FUNDING INFORMATION

Our research has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme under grant agreement No. (740704).

ETHICS STATEMENT

This study involved no experimental work, and ethical approval was therefore not relevant.

CONFLICT OF INTEREST

The authors declare no conflicting interests.

AUTHOR CONTRIBUTIONS

KS proposed the model design. AE wrote the code used for the simulations presented here, produced the figures and wrote the initial draft of the paper. KS and AE edited the paper.

DATA AVAILABILITY STATEMENT

The code used to produce the figures shown in this article is available on Figshare [29].

REFERENCES

1. Remuzzi A, Remuzzi G Covid-19 and Italy: what next? Lancet. 2020;395:1225–8.
2. Anderson RM, Heesterbeek H, Klinkenberg D, Hollingsworth TD How will country-based mitigation measures influence the course of the covid-19 epidemic? Lancet. 2020;395:931–4.
3. Dong E, Du H, Gardner L An interactive webbased dashboard to track covid-19 in real time. Lancet Infect Dis. 2020;20:533–4.
4. Endo A, Abbott S, Kucharski AJ, Funk S Estimating the overdispersion in covid-19 transmission using outbreak sizes outside china. Wellcome Open Res. 2020;5:67.
5. Lloyd-Smith JO, Schreiber SJ, Kopp PE, Getz WM Superspreading and the effect of individual variation on disease emergence. Nature. 2005;438:355–9.
6. Bi Q, Wu Y, Mei S, Ye C, Zou X, Zhang Z, et al. Epidemiology and transmission of covid-19 in 391 cases and 1286 of their close contacts in shenzhen, china: a retrospective cohort study. Lancet Infect Dis. 2020;20(8):911–9.
7. Adam D, Wu P, Wong J, Lau E, Tsang T, Cauchemez S, et al. Clustering and superspreading potential of severe acute respiratory syndrome coronavirus 2 (sars-cov-2) infections in hong kong, Research Square - preprint 2020.
8. Miller D, Martin MA, Harel N, Kustin T, Tirosh O, Meir M, et al. Full genome viral sequences inform patterns of sars-cov-2 spread into and within Israel. Nat Commun. 2020;11:5518. https://doi.org/10.1038/s41467-020-19248-0
9. Shen Y, Li C, Dong H, Wang Z, Martinez L, Sun Z, et al. Community outbreak investigation of sars-cov-2 transmissionamong bus riders in eastern China. JAMA Intern Med. 2020;180(2):1665–1671.
10. Park SY, Kim Y-M, Yi S, Lee S, Na B-J, Kim CB, et al. Early release-coronavirus disease outbreak in call center, South Korea. Emerg Infect Dis. 2020;26:1666–70.
11. Sneppen K, Taylor RJ, Simonsen L Impact of superspreaders on dissemination and mitigation of covid19. medRxiv. 2020.
12. Medhi N Covid-19 cases in the U.S. by population density. https://www.endcoronavirus.org/projects-1/covid-19-cases-in-the-us-by-population-density. Accessed 20 Aug 2020.
13. Carlson CJ, Chipperfield JD, Benito BM, Telford RJ, O'Hara RB Species distribution models are inappropriate for covid-19. Nature Ecol Evol. 2020;4:770–1.
14. Nishiura H, Oshitani H, Kobayashi T, Saito T, Sunagawa T, Matsui T, et al. Closed environments facilitate secondary transmission of coronavirus disease 2019 (covid-19). medRxiv. 2020.
15. Pappalardo L, Rinzivillo S, Qu Z, Pedreschi D, Giannotti F Understanding the patterns of car travel. Eur Phys J Spl Topics. 2013;215:61–73.
16. Tindale LC, Stockdale JE, Coombe M, Garlock ES, Lau WYV, Saraswat M, et al. Evidence for transmission of covid-19 prior to symptom onset. Elife. 2020;9:e57149.
17. Nishiura H, Linton NM, Akhmetzhanov AR Serial interval of novel coronavirus (COVID-19) infections. Int J Infect Dis. 2020;93:284–6.
18. Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, et al. Early transmission dynamics in wuhan, china, of novel coronavirus–infected pneumonia. N Engl J Med. 2020;382:1199–207.
19. Anderson RM, Anderson B, May RM Infectious diseases of humans: dynamics and control. Oxford: Oxford University Press; 1992.
20. Gabaix X Zipf's law for cities: an explanation. Q J Econ. 1999;114:739–67.
21. Soo KT Zipf's law for cities: a cross-country investigation. Reg Sci Urban Econ. 2005;35:239–63.
22. Yook S-H, Jeong H, Barabási A-L Modeling the internet's large-scale topology. Proc Natl Acad Sci. 2002;99:13382–6.
23. Mori T, Smith TE, Hsu W-T Common power laws for cities and spatial fractal structures. Proc Natl Acad Sci. 2020;117:6469–75.
24. Finkenstädt B, Keeling M, Grenfell B Patterns of density dependence in measles dynamics. Proc R Soc Lond B Biol Sci. 1998;265:753–62.
25. Finkenstädt B, Grenfell B Empirical determinants of measles metapopulation dynamics in England and Wales. Proc R Soc Lond B Biol Sci. 1998;265:211–20.
26. Fraser C, Cummings DA, Klinkenberg D, Burke DS, Ferguson NM Influenza transmission in households during the 1918 pandemic. Am J Epidemiol. 2011;174:505–14.
27. Bonabeau E, Toubiana L, Flahault A The geographical spread of influenza. Proc R Soc Lond B Biol Sci. 1998;265:2421–5.
28. Smith M, Yourish K, Almukhtar S, Collins K, Ivory D, Harmon A Covid-19 data repository. https://github.com/nytimes/covid-19-data. Accessed 03 Sept 2020.
29. Code available under the URL https://doi.org/10.6084/m9.figshare.12919709.v1

COVID-19 superspreading in cities versus the countryside - supplement

Andreas Eilersen and Kim Sneppen

Different distribution of infectious periods

In order to take into account that the duration of illness caused by COVID-19 does not actually follow a neat exponential distribution, we have here considered the case where the duration of infection is gamma distributed as well, though with a dispersion factor greater than one ($k \approx 4$). We repeat the central figures 3 and 5 from the main article in order to see if a different distribution of infectious period durations changes our results. The gamma distribution should give fewer people with a very short disease duration, though both distributions have a relatively long tail. The mean duration of infectiousness is the same as in the main article (10 days). An important complication, that individuals with a long period of illness are likely to be hospitalised or otherwise isolated for part of it and thus less infectious, has not been accounted for here.

We find that, while the different distribution of infectious periods does lower the percolation threshhold slightly and also diminishes the population density dependence of the epidemic, it does not fundamentally change our conclusions.

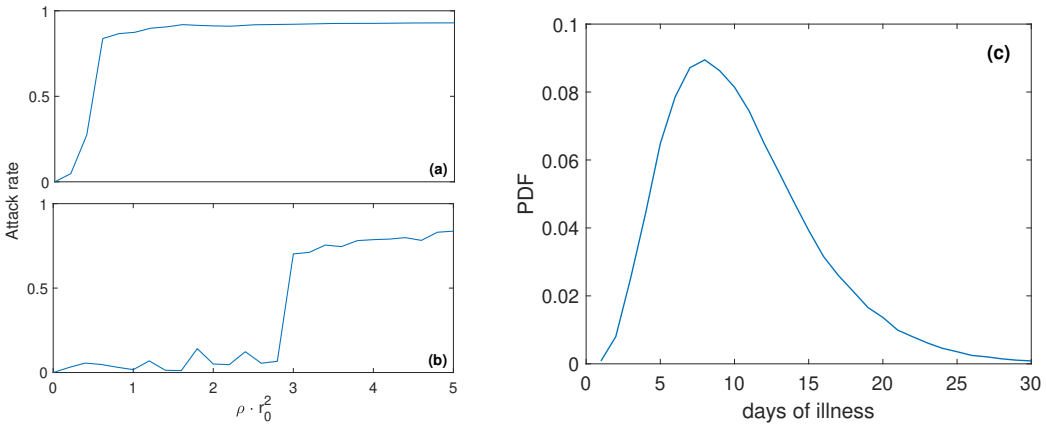


Figure 1: (a) and (b) show attack rates as a function of ρr_0^2, a measure of the number of neighbours within a radius of r_0. (a) shows the case with no infectivity overdispersion ($k = 1$) and (b) shows the case where infectivity is overdispersed with $k = 0.1$. We see that the system behaves similar to the main model, but with a slightly lower percolation threshhold in both cases. (c) shows the probability distribution function for disease duration used here.

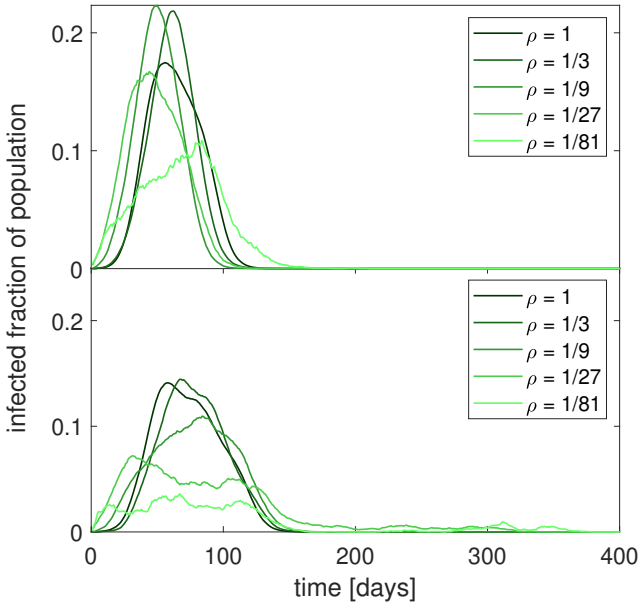


Figure 2: A comparison of the effect of population density on an epidemic that is not overdispersed (top) and an overdispersed epidemic (bottom). As above, the duration of illness is gamma distributed. The effect of density is somewhat reduced in the overdispersed case, but still noticeably greater than the non-overdispersed case. The smaller number of people with very short disease durations using this distribution seemingly also helps the epidemic spread more effectively.

Fractal city distribution

The random distribution of cities used in fig. 6 of the main article is not entirely realistic. Rather, there is evidence suggesting that real geographical distribution of cities follows a fractal pattern [1,2]. We therefore repeat the figure here with a (partially) fractal distribution of cities. The results of this simulation are fairly similar to the random city distribution.

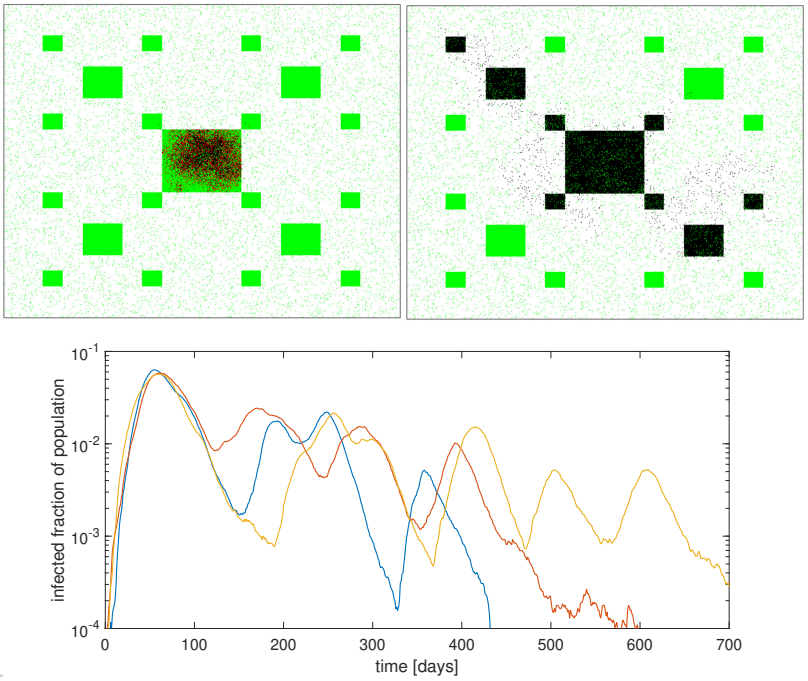


Figure 3: A simulation where, instead of a random distribution, the cities are laid out in a fractal pattern. We see that this makes little difference from the random distribution shown in the main article. The lower panel shows epidemic trajectories from three test runs.

1. Modeling the Internet's large-scale topology, Soon-Hyung Yook, Hawoong Jeong, and Albert-László Barabási, PNAS 99 (21) pp. 13382-13386 (2002)

2. Common power laws for cities and spatial fractal structures, Tomoya Mori, Tony E. Smith, and Wen-Tai Hsu, PNAS 117 (12) pp. 6469-6475 (2020)

Lockdowns exert selection pressure on overdispersion of SARS-CoV-2 variants

Authors: Bjarke Frost Nielsen*, Andreas Eilersen**, Kim Sneppen**, Lone Simonsen*

*Department of Science and Environment, Roskilde University, Roskilde, Denmark

**The Niels Bohr Institute, University of Copenhagen, Copenhagen, Denmark

My contributions: I participated in the development of the stochastic extinction model, wrote a program confirming the results on stochastic extinction, and read and commented on the manuscript.

Publication status: In review. A preprint is available on *medrxiv* (2021).

doi: 10.1101/2021.06.30.21259771

Lockdowns exert selection pressure on overdispersion of SARS-CoV-2 variants

Bjarke Frost Nielsen[a], **Andreas Eilersen**[a], **Lone Simonsen**[b], **and Kim Sneppen**[a,2]

[a]Niels Bohr Institute, University of Copenhagen, Blegdamsvej 17, 2100 Copenhagen, Denmark.; [b]Department of Science and Environment, Roskilde University, Universitetsvej 1, 4000 Roskilde, Denmark.

Preprint dated June 30, 2021

The SARS-CoV-2 ancestral strain has caused pronounced superspreading events, reflecting a disease characterized by overdispersion, where about 10% of infected people causes 80% of infections. New variants of the disease have different person-to-person variations in viral load, suggesting for example that the Alpha (B.1.1.7) variant is more infectious but relatively less prone to superspreading. Meanwhile, mitigation of the pandemic has focused on limiting social contacts (lockdowns, regulations on gatherings) and decreasing transmission risk through mask wearing and social distancing. Using a mathematical model, we show that the competitive advantage of disease variants may heavily depend on the restrictions imposed. In particular, we find that lockdowns exert an evolutionary pressure which favours variants with lower levels of overdispersion. We find that overdispersion is an evolutionarily unstable trait, with a tendency for more homogeneously spreading variants to eventually dominate.

Overdispersion | Evolution | Superspreading | Non-pharmaceutical interventions

One of the major features of the coronavirus pandemic has been overdispersion in transmission, manifesting itself as superspreading. There is evidence that around 10% of infected individuals are responsible for 80% of new cases (1–4). This means that some individuals have a high personal reproductive number, while the majority hardly infect at all. A recent study has shown this is reflected in the distribution of viral loads which is extremely wide, with just 2% of of SARS-CoV-2 positive individuals carrying 90% of the virus particles circulating in communities (5). Overdispersion is in fact a key characteristic of certain diseases (6–8). However, this is by no means a universal signature of infectious respiratory diseases. Pandemic influenza, for example, is characterized by a much more homogeneous transmission pattern (9–11).

As an emerging virus evolves, its transmission patterns may change and it may become more or less prone to superspreading. The Alpha (B.1.1.7) variant of SARS-CoV-2 has been reported to be $\sim 50\%$ more transmissible than the ancestral SARS-CoV-2 virus under varying degrees of lockdown (12–14). Meanwhile, others have shown that the Alpha variant possesses a higher *average* viral load and a reduced variability between infected persons, compared to the ancestral strain (15, 16). It remains to be seen how this reduced variability affects the transmission patterns of the virus.

The altered viral load distributions seen in persons infected with the Alpha variant have also been investigated at the level of individual mutations. The spike protein of the Alpha variant prominently features the N501Y substitution (asparagine replaced by tyrosine at the 501 position) as well as the ΔH69/V70 deletion (histidine and valine deleted at the 69 and 70 positions). Investigators found that the viral load is, on average, three times as great for the Alpha variant compared with the ancestral strain (16). Furthermore, viral load distributions in samples taken from persons infected with a variant with the ΔH69/V70 show a lower variance, whether or not they also have tyrosine at the 501 position. However, the difference in variance was most pronounced for those samples which had the deletion as well as the 501Y mutation. Similarly, an analysis of samples with the N501Y mutation show that they have a higher median viral load as well as a substantially diminished variance compared to those without it. Using data from Ref. (15), we calculate that the viral loads in samples of the Alpha variant are associated with a lower coefficient of variation of approximately 2, compared to 4 for the ancestral strain. Importantly, the exact relation between viral load and infectiousness is not well understood; however, a higher viral load is logically expected to increase the risk of disease transmission. By this logic, the decreased variability in the viral load for the Alpha variant may translate into a reduced overdispersion in transmission.

In this paper, we use a mathematical model to study the competition between idealized variants which differ in their level of overdispersion (k) and their mean infectiousness. Our focus is on exploring whether overdispersion confers any evolutionary (dis)advantages, and whether non-pharmaceutical interventions which restrict social network size and transmissibility change the fitness landscape for variants with varying degrees of overdispersion. While it is evident that a higher mean infectiousness confers an evolutionary advantage to an emerging pathogen, it is not *a priori* obvious if a competitive

Significance

One of the most important and complex properties of viral pathogens is their ability to mutate. The SARS-CoV-2 pandemic has been characterized by overdispersion – a propensity for superspreading, which means that around 10% of those who become infected cause 80% of infections. However, evidence is mounting that this is not a stable property of the virus and that the Alpha variant spreads more homogeneously. We use a mathematical model to show that lockdowns exert a selection pressure, driving the pathogen towards more homogeneous transmission. In general, we highlight the importance of understanding how non-pharmaceutical interventions exert evolutionary pressure on pathogens. Our results imply that overdispersion should be taken into account when assessing the transmissibility of emerging variants.

The authors declare no competing interests.

[2]To whom correspondence should be addressed. E-mail: sneppen@nbi.dk

advantage can be gained by specifically altering the *variability* in infectiousness (while keeping transmissibility unchanged). Our recent studies have shown that the presence of overdispersion makes a pandemic far more controllable than influenza pandemics when mitigating by limiting non-repetitive contacts (17) and personal contact network size (18). We therefore speculate that restrictions which alter social contact structure may, conversely, provide a fitness advantage to variants with more homogeneous transmission, and may thus play a role in viral evolution.

Across several diseases, individual variations in infectiousness have been approximated by a Gamma distribution (6) characterized by a certain mean value and a dispersion parameter known as k, which is related to the coefficient of variation (CV) through $CV = 1/\sqrt{k}$. In the simplest of cases (a well-mixed population), infection attempts are modeled as a constant-rate (Poisson) process, which leads to a personal reproductive number which follows a negative binomial distribution. The dispersion parameter k characterizes the degree of transmission heterogeneity; a *lower* k corresponds to greater heterogeneity. For small values of k, it approximately corresponds to the fraction of infected individuals responsible for 80% of new infections The value for the SARS-CoV-2 ancestral virus is around 10%, corresponding to a k-value of approximately 0.1. Other coronaviruses are also prone to superspreading, with the k-values of SARS-CoV-1 and MERS estimated at 0.16 (6) and 0.26 (19), respectively. To explore questions of how such overdispersion affects fitness and pathogen evolution, we use an agent-based model of COVID-19 spreading in a social network, as originally developed in Ref. (18).

Overdispersion in personal reproductive number – i.e. superspreading – is a phenomenon that requires *means* (biological infectiousness) as well as *opportunity* (social context). Superspreading can have diverse origins, ranging from purely behavioural to biological (8, 20). However, a recent meta-review (21) compared the transmission heterogeneity of influenza A (H1N1), SARS-CoV-1 and SARS-CoV-2 and found that higher variability in respiratory viral load was closely associated with increased transmission heterogeneity. This suggests that biological aspects of individual diseases are decisive in determining the level of overdispersion, and thus the risk of superspreading.

Initial survival of variants

The words *fitness* and *competitive advantage* may take on several meanings in an evolutionary context. For our purposes, it is especially important to distinguish between the ability of a pathogen to *avoid stochastic extinction* and to *reproduce effectively* in a population.

To quantify the ability to avoid stochastic extinction we use a branching process to simulate an outbreak of a variant with a given level of overdispersion in a naive population. We then record whether it survives beyond the first 10 generations of infections, as a measure of the ability of that variant to take hold. Repeating these simulations multiple times allows us to compute the survival chance of each variant as a function of its infectiousness and overdispersion, in the absence and presence of mitigation (Fig. 1). Since we are dealing with a few related quantities, some definitions must be made. By the *basic reproductive number* (R_0) we mean the average number of new infections which each infected person gives rise to *when all contacts are susceptible*. This is in contrast to the effective reproductive number (known variously as R, R_t and R_e), which is affected by population immunity. Note that R_0 as well as R_e are context dependent, since behaviour (and mitigation strategies) will affect e.g. the number of contacts that a person has and thus the reproductive number. Another parameter entirely is the *(biological) mean infectiousness*, by which we mean the rate at which transmission occurs *when* an infected person is in contact with a susceptible person. This is a property of the disease and not of the social environment. In Fig. 1, the independent variables are thus the mean infectiousness and the dispersion parameter, both of which are assumed to be properties of the disease. The details of the calculation can be found in the Materials and Methods section.

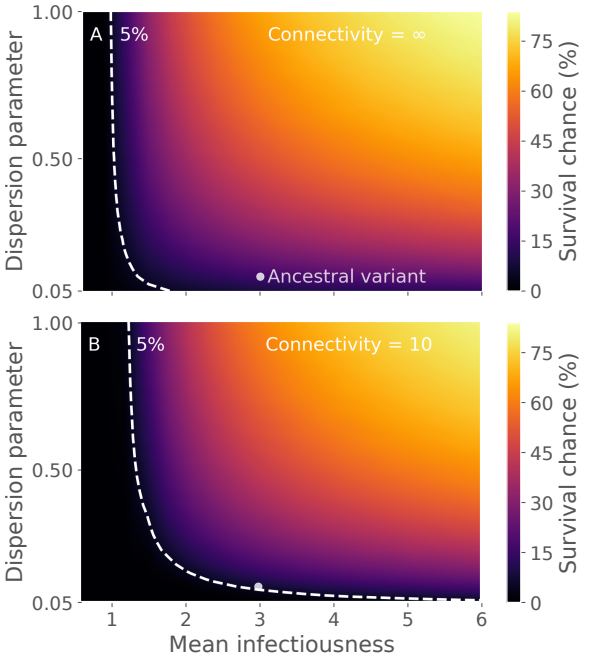


Fig. 1. Initial survival chance depends strongly on overdispersion and moderately on lockdown status. A) The epidemic spreads in an unrestricted setting (homogeneous mixing contact structure) **B)** The epidemic spreads in a situation with limited social connectivity (modeled as an Erdos-Renyi network of average connectivity 10). The survival chance is computed by simulating several outbreaks, each starting from a single infected individual in a susceptible population. This initial individual is infected with a variant of a given overdispersion. For each outbreak, the variant is recorded as having *survived* if it does not go extinct within 10 generations. The dashed white line indicated parameters for which the variant has a 5% chance of surviving. The biological mean infectiousness (horizontal axis) has been scaled such that it equals the basic reproductive number (R_0) in the homogeneous mixing scenario of panel A. For details on these calculations, see the Materials and Methods section.

In the unmitigated scenario (Fig. 1A), the procedure is relatively straightforward. A single infected individual is initially introduced, with a personal reproductive number z drawn from a negative binomial distribution $P_{\mathrm{NB}}[Z; R_0, k]$ with mean value R_0 and dispersion parameter k. Thus, this individual gives rise to z new cases, and the algorithm is reiterated for each of these subsequent infections.

In the case of a lockdown scenario, in terms of restrictions of the number of social contacts (Fig. 1B), the algorithm is slightly more involved. In this case, a *degree* c (the number of contacts) is first drawn from a degree distribution (in this case

a Poisson distribution, to mimic an Erdös-Renyi network). A biological reproductive number ξ (the *infectiousness*) is then drawn from a Gamma distribution with mean value R_0 and dispersion parameter k. The actual personal reproductive number z is then drawn from the distribution

$$P(z;\xi,c) = \binom{c}{z}\left(1-e^{-\xi/c}\right)^{z}\left(e^{-\xi/c}\right)^{(c-z)}. \qquad [1]$$

This reflects that the personal reproductive number z is, naturally enough, limited by the number of distinct social contacts c. This algorithms is then reiterated for each of the z new cases.

Similar results can be obtained analytically by considering the probability that an infection chain dies out in infinite time. Let that probability be d and let p_i, $i \in \{0, 1, \dots\}$ be the distribution of personal reproductive number (i.e. p_i is the probability that a single infected individual will infect i others). Then the extinction risk d is the sum:

$$d = p_0 + p_1 d + p_2 d^2 + \dots \qquad [2]$$

where the first term on the right hand side is the extinction risk due to the index case producing no new infections, the second term is the case where the index case gives rise to one branch of infections which then dies out (this being the reason for the single factor of d in the second term) and so on. Since each new branch exists independently of the other, the extinction events are independent and the probabilities may be combined by simple multiplication as in Eq. Eq. (2).

We find that the survival chance depends very strongly on overdispersion (Fig. 1), with more homogeneous variants ($k \sim 1$) having a good chance of survival while highly overdispersed variants ($k \leq 0.1$) are very unlikely to survive beyond 10 generations. This finding fits well with the general pattern of overdispersed spreading, namely that many individuals hardly become infectious while a few pass the disease onto many others. The uneven distribution of infectiousness makes heterogeneous diseases more fragile in the early stages of an epidemic, and thus more prone to stochastic extinction.
For the case of homogeneous mixing (Fig 1A) and the number of generations tending to infinity, Lloyd-Smith et al (6) performed a similar calculation using the generating function method described in Eq. 2. For a disease with $R_0 = 3$ and a k value of 0.16 (similar to what they estimated for SARS-CoV-1), the survival chance was found to be 24%. Our model yields the same figure in the unmitigated connectivity$\to\infty$ limit.

To assess the effect of lockdown-like non-pharmaceutical interventions on the initial survival chances of a pathogen, we performed an analogous computation in a socially restricted setting (Fig. 1B). Compared with the unmitigated scenario of Fig. 1A, it can be seen that the mitigation has an effect on the survival chance, affecting highly overdispersed variants (small k) much more than their more homogeneous counterparts (with the same mean infectiousness). This result is parallel to the effect of lockdown-like interventions on the *competitive advantage* of a variant, which we explore in the next section.

In Ref. (20), the authors study stochastic extinction of a superspreading disease under a targeted intervention they call *cutting the tail*. They introduce a cutoff value N_{cutoff} for the personal reproductive number, and if a person has a personal reproductive number $z \geq N_{\text{cutoff}}$, a new z is drawn until one below the threshold is obtained. Since the disease is highly heterogeneous, this process is analogous to "removing" a potential superspreading event and replacing it with a much lower personal reproductive number (typically $z = 0$). This is exactly why the intervention is rightly called *targeted*. Their approach is thus based on viewing superspreading entirely as an event-based phenomenon, where one can directly remove superspreading events above some threshold size, and instead let the individuals take part in other less risky events. Our approach, on the other hand, assumes superspreading to be due to a combination of high individual biological infectiousness and opportunity, e.g. a large number of social contacts. These two viewpoints are complementary in obtaining a comprehensive description of superspreading phenomena, rather than mutually exclusive (17).

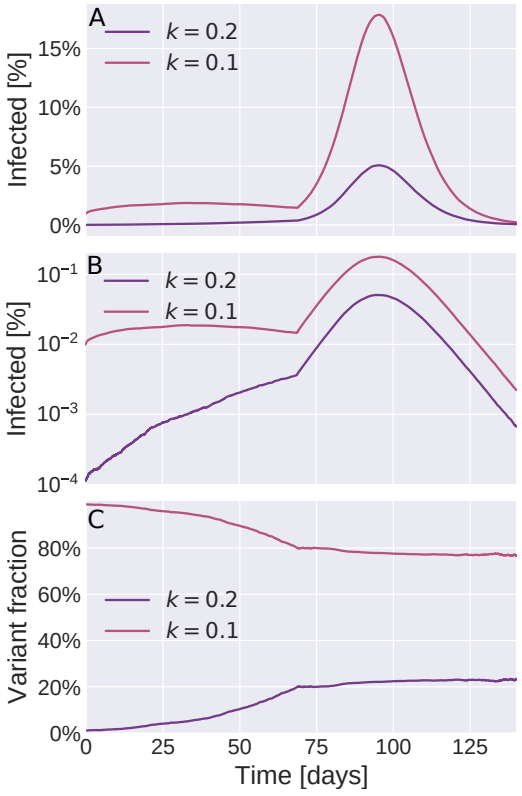


Fig. 2. Simulations of the emergence of a new variant. An initially dominant ("ancestral") strain with dispersion parameter $k = 0.1$ (red) has initially infected 1% of the population. The figure follows the emergence of a new variant (purple), which has the same biological mean infectiousness, but is more homogeneous ($k = 0.2$). Initially, 0.01% of the population is infected with the emerging variant. The two variants exhibit perfect cross-immunity. The initial scenario is a partially locked-down society (modeled as an Erdös-Renyi network with 10 contacts/person). When the new variant reaches 20% of all current infections (around day 65), the lockdown is completely lifted (modeled by a homogeneous mixing contact structure with the same *total* social time available per person). **A)** Incidence of each strain as a function of time since the new variant was introduced. Notice that the new variant spreads approximately exponentially until day 65 (see also panel B), whereas the ancestral strain stays at about 1% incidence. When restrictions are lifted, both surge. **B)** Same data as panel A, but plotted on a logarithmic scale. In this plot, exponential growth shows up as a straight line, and it is thus clear that the new variant spreads approximately exponentially during the lockdown phase. **C)** The relative proportions of the old and new variants. In the locked-down society, the new variant has a distinct fitness advantage, as revealed by its increasing share of infections. Once restrictions are lifted around $t = 65$ days, the fitness advantage is lost and the two variants spread equally well.

Competitive advantage is determined by context

We now turn to the competition between two variants which have already managed to gain a foothold, and so have moved past the initial risk of stochastic extinction. This is a separate aspect of "fitness", distinct from the initial survival ability described in the last section. Fig. 2 explores the competition between two strains which differ only in their level of overdispersion. The ancestral variant has a broad infectiousness distribution ($k = 0.1$) while the other – the *new variant* – is more narrowly distributed ($k = 0.2$). In the initial partial lockdown scenario, each person is only allowed contact with 10 others, At first, the fraction of infections due to the new variant is observed to grow rapidly. When it reaches a 20% share of active infections, around day 65, the lockdown is lifted (simulated by a shift to a homogeneous mixing contact structure). Naturally, this more permissive contact structure causes a surge in both variants (Fig. 2c). However, the fraction of infections owing to *each* variant suddenly stabilizes, indicating that the more homogeneous new variant has lost its competitive advantage in the unmitigated scenario.

This sudden loss of competitive advantage demonstrates conceptually that the fitness of variants with different patterns of overdispersion depends on context, in the form of non-pharmaceutical interventions or the absence thereof. To quantify this dependence, we separately simulate the spread of several pathogen variants, each with its own specified mean infectiousness and dispersion parameter k, and measure the resulting basic reproductive numbers. In each case we let the pathogen spread in an Erdös-Renyi network with a mean connectivity of either 10 or 50, to simulate scenarios with either a restricted or fairly open society. The results are shown in Fig. 3, where the competitive (dis)advantage of each variant is plotted as a function of its a given biological mean infectiousness and dispersion. The infectiousness is given relative to the SARS-CoV-2 ancestral strain which is set to average infectiousness $= 1$ and has dispersion $k = 0.1$. This average infectiousness of 1 corresponds to a basic reproduction number of $R_0 = 3$ in a well-mixed scenario, representative of COVID-19 (22). In the socially restricted case with only 10 contacts, the competitive advantage depends strongly on the dispersion parameter, as evidenced by the contour lines in Fig. 3A. The dashed white contour in the figure indicates variants which spread *as well* as the ancestral strain. Concretely, a variant with just half the biological infectiousness of the ancestral strain has no substantial competitive disadvantage, provided it is sufficiently homogeneous ($k \gtrsim 1.0$). In the more socially connected scenario (Fig. 3B), the competitiveness of a strain is observed to depend less strongly on dispersion, and is primarily determined by biological mean infectiousness. Viewed more broadly, these results imply that an observed increase in R_0 for an emerging variant may be due to a *combination* of changes in transmission patterns (k) and biological mean infectiousness

So far, our focus has been on mitigation strategies which rely on reductions in contact network. However, even when societies reopen by allowing contact with an increased number of individuals, non-pharmaceutical interventions which decrease transmission risk per encounter may be in force. These may include face masks and regular testing. In the Supporting Information, we show that interventions which decrease the transmission risk per encounter (i.e. per unit of contact time) in fact decrease the competitive advantage of more homogeneous variants. These types of interventions thus have essentially the opposite effect, relative to strategies which reduce social connectivity.

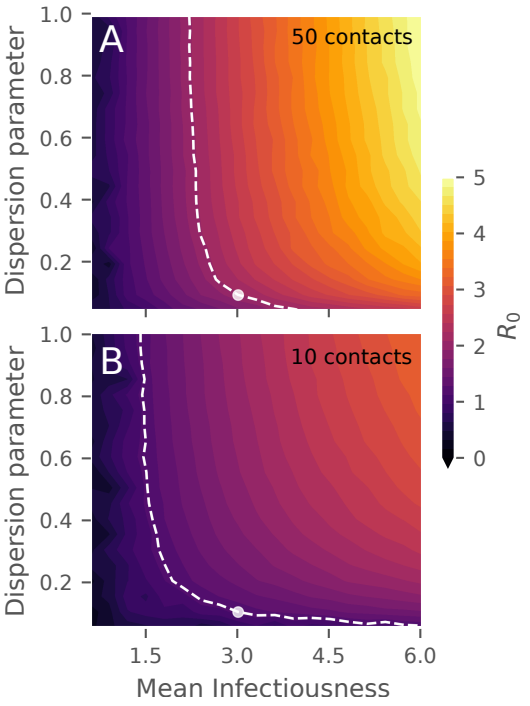


Fig. 3. Relative fitness of variants. The color indicates the basic reproductive number that each variant exhibits under the given circumstances. The dashed white line indicates variants which have the same fitness as the ancestral strain, which is estimated to have $k = 0.1$. The biological mean infectiousness (horizontal axis) has been scaled such that it equals the basic reproductive number (R_0) in a homogeneous mixing scenario. **A)** Spread of the disease in a connectivity 10 Erdös-Renyi network, corresponding to a partial lockdown. **B)** Spread of the disease in a connectivity 50 Erdös-Renyi network, corresponding to a mostly open society.

Interventions exert selection pressure

As the observed differences in the viral load distributions of the Alpha (B.1.1.7.) variant and the ancestral strain suggest, overdispersion is not a fixed property, but rather one that may evolve over time. Furthermore, the SARS-CoV-2 pathogen has been estimated to mutate at a rate of approximately 2 substitutions per genome per month (23), translating to about one mutation per three transmissions. In Fig. 4, we explore the consequences of overdispersion as an evolving feature of the pathogen. In these simulations, the virus has a mutation probability of 1/3 at each transmission. When it mutates, the overdispersion factor is either increased (by a factor of 3/2) or decreased (by a factor of 2/3). Thus, we assume no drift on the microscopic scale, but one may arise macroscopically due to selection pressure from the environment. It should of course be noted that while the assumed mutation rate is realistic for SARS-CoV-2, many mutations will be neutral and only very few mutations will affect transmission dynamics. As such, the present model will likely overestimate the *magnitude* of the drift in overdispersion. It is however conceptually robust – decreasing the mutation rate merely slows down the drift, but the tendency remains.

In our simulations, we find that there is always a tendency

for overdispersion to decrease (i.e. for the k value to *increase*), leading to more homogeneous disease transmission. This makes sense, since we have already established that heterogeneous disease variants are more likely to undergo stochastic extinction (Fig. 1) and that they have a competitive disadvantage as soon as contact structures are anything but well-mixed (Fig. 3). In the absence of any interventions, the tendency to evolve towards homogeneity is quite weak (Fig. 4A), but when a partial lockdown is instituted, the picture changes dramatically and the k value increases exponentially. The conclusion is thus that lockdowns exert a selection pressure on the virus when it comes to overdispersion, towards developing a less superspreading-prone phenotype.

One may of course object that the scenarios of Fig. 4A (unrestricted spread) and 4B (partial lockdown) are not directly comparable, since the epidemic in 4A unfolds much more rapidly. For this reason, we have included the scenario shown in 4C, where the transmission rate per encounter has been lowered, but social structure is unrestricted. The transmission rate is lowered such that the *initial* daily growth rates in Fig. 4B and 4C are identical (11%/day averaged over the first 14 days). This slightly increases the growth of k over the course of the epidemic, but to a much lower level than in the lockdown scenario, demonstrating that it is indeed the restriction of social network that provides the selection pressure driving k upwards.

Discussion

With this paper we have demonstrated that the relative success and survival of mutants of a superspreading disease depends on the type of mitigation strategies employed within a population. The choice of a certain mitigation strategy may well amount to selecting the next dominant variant. If, for example, a simple lockdown is enacted while still allowing people to meet within restricted social groups, the evolution of more homogeneously spreading disease variants may become favoured.

The spreading of an emerging virus in a human society is a complex phenomenon, where the actual reproductive number depends on sociocultural factors, mitigation policies and self-imposed changes in the behaviour of citizens as awareness grows in the population. The spread of a disease such as COVID-19 cannot simply be characterized by a single fitness quantity like the basic reproductive number R_0, but will also depend on the heterogeneities of transmission patterns within the population. If schools are open, mutants which spread more easily among children may be selected for, whereas rapid self-isolation of infected individuals may tend to favor variants which temporally separate disease transmission from the development of symptoms. We have focused on modeling the evolutionary effects of biological superspreading in the context of mitigations such as lockdowns which have been implemented globally during the COVID-19 pandemic. We found that such lockdowns will favour the emergence of homogeneously spreading variants over time.

Our findings also have implications for the assessment of new variants. They highlight the importance of taking overdispersion into account when evaluating the transmissibility of an emerging variant. We have shown that the disease can spread more effectively not only by increasing its biological mean infectiousness, but also by changing its pattern of transmission to become more homogeneous. Practically, this means that transmission data obtained under even partial lockdown can lead to an overestimation of the transmissibility of an emerging variant. We thus call for an increased focus on measuring the overdispersion of variants, as this may be critical for estimating the reproductive number of new variants. These estimates in turn determine the required vaccination levels to reach herd immunity.

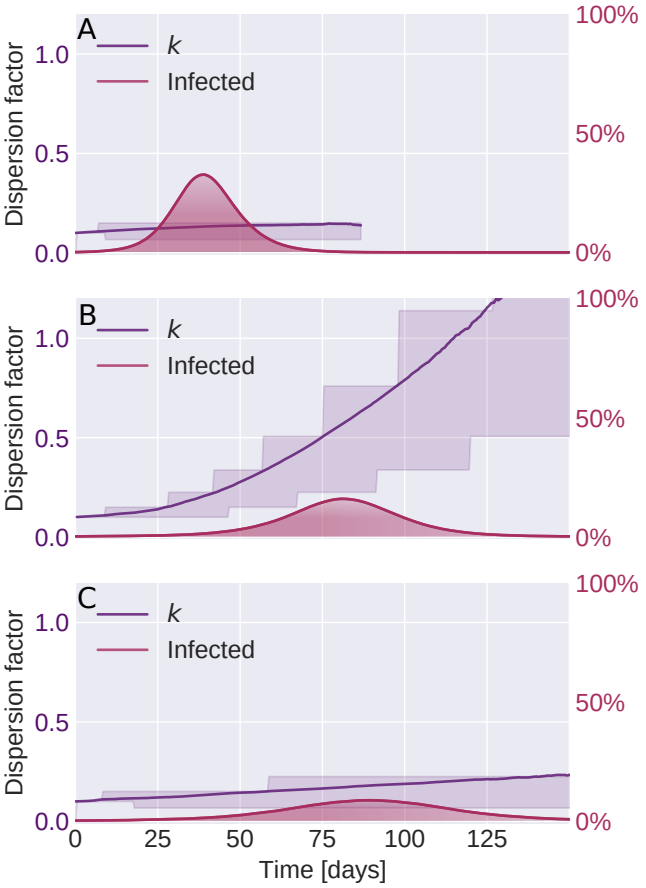


Fig. 4. Evolution of overdispersion is driven by imposed restrictions. In these simulations, random mutations occur which alter the level of transmission overdispersion in a non-directed fashion. However, external evolutionary pressures are seen to drive the disease towards developing more homogeneous spreading patterns. The filled red curve shows the combined incidence of all strains. The purple curve shows the average dispersion factor k in the infected population (with higher k corresponding to a more homogeneous infectiousness). The shaded purple area shows the 25% and 75% percentiles of the distribution of dispersion factors in the infected population. **A)** The pathogen evolves in an open society with no restrictions imposed (homogeneous mixing contact structure). **B)** Partial lockdown, with an average social network connectivity restricted to 15 persons. **C)** No restrictions on social network, but infectiousness lowered by other means (e.g. face masks).

Materials and Methods

We use an individual-based (or agent-based) network model of disease transmission as originally developed in Ref. (18). In this section, we present only a brief overview of the basic model, and refer to Ref. (18) for a more detailed description. We then go on to describe in detail the simulations and calculations which are particular to this manuscript.

The disease progression model consists of four overall states, **S**usceptible, **E**xposed, **I**nfected and **R**ecovered. The exposed state has an average duration of 2.4 days and is subdivided into two consecutive states with exponentially distributed waiting times (i.e.

having constant probability rate for leaving the state) of 1.2 days each, thus constituting a gamma distributed state when viewed as a whole. The infectious state is divided into two states as well, of 1.2 and 5 days in duration, respectively.

Each individual in the model is associated with a fixed social network. Only a subset of edges are activated in each timestep, to simulate a contact event. In the simulations of this work, we always use either an Erdös-Renyi network with finite mean connectivity, or a homogeneous-mixing contact structure, which is also obtainable as the infinite connectivity limit of an Erdös-Renyi network.

When an edge connecting a susceptible and an infectious individual is active, there is a certain probability per unit of time for disease transmission to occur. This rate is determined by the individual infectiousness r_i of the infectious agent, which is drawn from a gamma distribution with dispersion parameter k before the individual has become infectious. As such, the infectiousness for any given individual is assumed constant throughout the infectious stage of the disease. The infectiousness distribution determines an upper bound on size Δt of the the timesteps in the model, since the inequality $r_i \cdot \Delta t < 1$ must hold for all agents. A timestep of size $\Delta t = 30$min was used throughout, since this was sufficient to ensure that the inequality was satisfied.

Below we go into more detail as to how the simulations involving multiple strains were performed.

Stochastic extinction. The stochastic extinction (or, conversely, survival) plots of Figure 1 in the main text rely entirely on a branching process algorithm with sampling of probability distributions with an analytic description. In practice, we have performed the computation by numerical sampling.

In each generation of the epidemic, the computation is reiterated. Without loss of generality, we therefore here describe a single generation which initially has I infected individuals. Note that for the initial generation, $I = 1$ infected individuals.

- For $i \in \{1, \ldots, I\}$:
  - Draw individual infectiousness ξ_i from Gamma distribution $P_\xi(\xi; k, \mu)$
  - Draw number of contacts c from a Poisson distribution with a given mean connectivity.
  - Given number of contacts c, draw personal reproductive number z_i from the distribution Eq. (3)

$$P_z(z; \xi, c) = \binom{c}{z} \left(1 - e^{-\xi/c}\right)^z \left(e^{-\xi/c}\right)^{(c-z)}. \quad [3]$$

- Let the number of newly infected be $I = \sum_i z_i$ and repeat the algorithm with this new value of I.

If the number of infected I ever drops to zero, the outbreak is said to have undergone stochastic extinction in that generation. By performing multiple such branching process simulations for each value of the parameters μ (mean infectiousness) and k (dispersion factor) we build up a statistic of the survival chance of each specific variant. To generate Figure 1, this is repeated for two different values of the mean connectivity c.

Two-strain competition simulations. In Fig. 2, two strains spread simultaneously in the population of $N = 10^6$ individuals. Initially, 0.99% of the population are infected with the heterogeneous "old" variant ($k = 0.1$), while 0.01% are infected with the more homogeneous "new" variant ($k = 0.2$). Once a person with a given variant infects a susceptible individual, the characteristics of the variant are passed on to the newly infected individual, such that the infectiousness of this person is drawn from a Gamma distribution with dispersion parameter k set by the variant. In other words, these simulations assume that no further mutations affecting overdispersion occur, allowing us to track solely the competition of two differently-dispersed variants within a population.

Evolutionary model. In Fig. 4, we allow the pathogen to stochastically mutate upon transmission, with the mutations affecting the degree of overdispersion. In the simulations, the pathogen mutates on average once for each new host it is transmitted to (i.e. with mutation probability $p = 1/3$) and the mutations are assumed to always affect overdispersion, by either increasing the k value by a factor of 3/2 (i.e. $k \to 3k/2$) or decreasing it by a factor of 2/3 (i.e. $k \to 2k/3$). On a microscopic level, the dispersion level thus performs an unbiased (multiplicative) random walk. The value of this step-size parameter is arbitrarily chosen, and as such the simulations can only be regarded as qualitative and conceptual. However, although no intrinsic bias is built into the mutation mechanism, external selection pressures may drive the level of overdispersion in the population up or down, as is explored in Fig. 4.

In Fig. 4C, the average infectiousness of the strain is lowered so as to produce an initial growth rate that is identical to that of 4A, namely 11% per day in the first 14 days of the epidemic.

ACKNOWLEDGMENTS. We thank Robert J. Taylor, and Julius B. Kirkegaard for enlightening discussions. Our research has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme, grant agreement No. 740704, as well as from the Carlsberg Foundation under its Semper Ardens programme (grant # CF20-0046).

1. D Miller, et al., Full genome viral sequences inform patterns of SARS-CoV-2 spread into and within Israel. *Nat. communications* **11**, 1–10 (2020).
2. A Endo, S Abbott, AJ Kucharski, S Funk, , et al., Estimating the overdispersion in covid-19 transmission using outbreak sizes outside china. *Wellcome Open Res.* **5**, 67 (2020).
3. C Pozderac, B Skinner, Superspreading of sars-cov-2 in the usa. *Plos one* **16**, e0248808 (2021).
4. JB Kirkegaard, K Sneppen, Variability of individual infectiousness derived from aggregate statistics of covid-19. *medRxiv* **0** (2021).
5. Q Yang, et al., Just 2% of sars-cov-2-positive individuals carry 90% of the virus circulating in communities. *Proc. Natl. Acad. Sci.* **118** (2021).
6. JO Lloyd-Smith, SJ Schreiber, PE Kopp, WM Getz, Superspreading and the effect of individual variation on disease emergence. *Nature* **438**, 355–359 (2005).
7. AP Galvani, RM May, Dimensions of superspreading. *Nature* **438**, 293–295 (2005).
8. ME Woolhouse, et al., Heterogeneities in the transmission of infectious agents: implications for the design of control programs. *Proc. Natl. Acad. Sci.* **94**, 338–342 (1997).
9. C Fraser, DA Cummings, D Klinkenberg, DS Burke, NM Ferguson, Influenza transmission in households during the 1918 pandemic. *Am. journal epidemiology* **174**, 505–514 (2011).
10. J Brugger, CL Althaus, Transmission of and susceptibility to seasonal influenza in switzerland from 2003 to 2015. *Epidemics* **30**, 100373 (2020).
11. MG Roberts, H Nishiura, Early estimation of the reproduction number in the presence of imported cases: pandemic influenza h1n1-2009 in new zealand. *PloS one* **6**, e17835 (2011).
12. MS Graham, et al., Changes in symptomatology, reinfection, and transmissibility associated with the sars-cov-2 variant b. 1.1. 7: an ecological study. *The Lancet Public Heal.* **6**, e335–e345 (2021).
13. E Volz, et al., Assessing transmissibility of sars-cov-2 lineage b. 1.1. 7 in england. *Nature* **593**, 266–269 (2021).
14. NG Davies, et al., Estimated transmissibility and impact of sars-cov-2 lineage b. 1.1. 7 in england. *Science* **372** (2021).
15. M Kidd, et al., S-variant SARS-CoV-2 lineage B1.1.7 is associated with significantly higher viral loads in samples tested by ThermoFisher TaqPath RT-qPCR. *The J. infectious diseases* **223** (2021).
16. T Golubchik, et al., Early analysis of a potential link between viral load and the N501Y mutation in the SARS-COV-2 spike protein. *medRxiv* **0** (2021).
17. K Sneppen, BF Nielsen, RJ Taylor, L Simonsen, Overdispersion in COVID-19 increases the effectiveness of limiting nonrepetitive contacts for transmission control. *Proc. Natl. Acad. Sci.* **118** (2021).
18. BF Nielsen, L Simonsen, K Sneppen, COVID-19 superspreading suggests mitigation by social network modulation. *Phys. Rev. Lett.* **126**, 118301 (2021).
19. A Kucharski, CL Althaus, The role of superspreading in middle east respiratory syndrome coronavirus (mers-cov) transmission. *Eurosurveillance* **20**, 21167 (2015).
20. BM Althouse, et al., Superspreading events in the transmission dynamics of sars-cov-2: Opportunities for interventions and control. *PLoS biology* **18**, e3000897 (2020).
21. PZ Chen, et al., Heterogeneity in transmissibility and shedding sars-cov-2 via droplets and aerosols. *Elife* **10**, e65774 (2021).
22. A Billah, M Miah, N Khan, Reproductive number of coronavirus: A systematic review and meta-analysis based on global level evidence. *PLOS ONE* **15**, 1–17 (2020).
23. M Worobey, et al., The emergence of sars-cov-2 in europe and north america. *Science* **370**, 564–570 (2020).

Supporting Information: Lockdowns exert selection pressure through overdispersion of SARS-CoV-2 variants

Bjarke Frost Nielsen,[1,*] Andreas Eilersen,[1,†] Lone Simonsen,[2,‡] and Kim Sneppen[1,§]

[1] *Niels Bohr Institute, University of Copenhagen, Blegdamsvej 17, 2100 Copenhagen, Denmark.*

[2] *Department of Science and Environment, Roskilde University, Universitetsvej 1, 4000 Roskilde, Denmark.*

(Dated: June 30, 2021)

INTERVENTIONS TARGETING TRANSMISSION RISK

In the main text we have primarily considered the evolutionary pressures exerted by mitigation strategies which rely on reductions in social connectivity, such as lockdowns. However, other mitigation strategies may rely not on limiting the number of distinct individuals that a person interacts with, but rather on decreasing the transmission rate when contacts *do* occur. Examples of such strategies include the use of face masks, which work primarily by decreasing the number of emitted virions, and physical distancing which exploits the decrease in viral concentration as distance to the source is increased. In Fig. 1, we consider two scenarios in which a more homogeneous variant with dispersion factor $k = 0.2$ emerges in the background of the $k = 0.1$ ancestral strain. Note that the curves in Fig. 1 show the relative *share* of infections owing to the emerging variant and not the absolute incidence. In both of the cases studied, social connectivity is initially quite restrictive, with each person allowed only 10 contacts. At time $t = 25$days, restrictions are relaxed, simulated by increasing average connectivity to 50. This is where the two scenarios diverge: in one of them, the infection risk *per encounter* is halved, while in the other it stays the same. It is of course expected that halving the infection risk will cause a lower overall epidemic, but what is not obvious is how it will affect the competition between the two variants. As expected from Fig. 3 of the main text, the competitive advantage of the new variant is reduced by going from 10 to 50 social connections, as is reflected in both of the curves in Fig. 1 bending off at $t = 25$days. However, more notable is the fact that the competitive advantage of the new variant is substantially reduced when restrictions are put in place which decrease the infection rate. The conclusion is thus that heterogeneous variants are particularly vulnerable to lockdown-type interventions, where social network size is reduced, while more homogeneous variants are more susceptible to interventions which reduce infectiousness during each encounter.

In Fig. 2 we show the absolute incidence of the two variants in another simulation experiment where the two types of interventions are in force simultaneously. Initially, only the (partial) lockdown is in force, with social connectivity restricted to 10 persons. At $t = 35$ days, another non-pharmaceutical intervention which reduces the transmission risk per encounter by 20% is put in force. This is seen to reduce the more homogeneous emerging variant to marginal spread while the ancestral strain starts to decline.

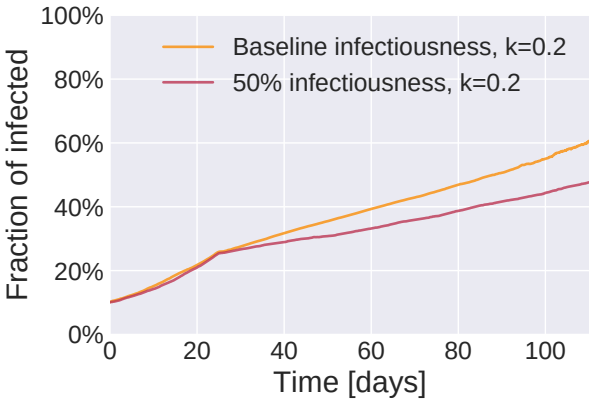


FIG. 1. **Re-opening society while putting other restrictions in place.** The plot shows the relative abundance of a new, more homogeneous variant ($k = 0.2$) as a function of time. At time $t = 0$, the new variant makes up 10% of infections, while the rest owe to the more heterogeneous "old" variant ($k = 0.1$). Until time $t = 25$ a partial lockdown is in place, modeled by restricting personal contact networks to only 10 persons on average. At $t = 25$days, society is partially opened, simulated by allowing contact with 50 different persons. The red line represents a scenario where the opening of society is accompanied by other restrictions, reducing the infection risk *per encounter* by half. As such, the diversity of contacts encountered is increased, but the infection risk per encounter is decreased. In the scenario shown by the yellow line, infection risk per encounter is unaltered. Clearly, these interventions negatively affect the competitive advantage of the more homogeneous variant.

* bjarkenielsen@nbi.ku.dk
† andreaseilersen@nbi.ku.dk
‡ lonesimo@ruc.dk
§ sneppen@nbi.dk

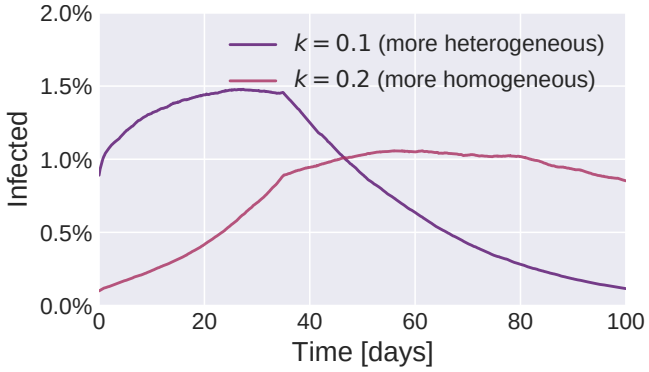


FIG. 2. Two variants with dispersion parameters $k = 0.1$ and $k = 0.2$ spread in a socially restricted society with a personal contact network size of 10. From time $t = 35$days, non-pharmaceutical interventions are introduced which reduce the rate of transmission by 20%. This is enough to reduce the more homogeneous ($k = 0.2$) variant to marginal growth, while the heterogeneous variant dies out.

Evolutionarily Stable Strategies (ESS) in a socially age-structured epidemic model

Authors: Andreas Eilersen*, Ruiyun Li**, Nils Christian Stenseth**

*The Niels Bohr Institute, University of Copenhagen, Copenhagen, Denmark

**Centre for Ecological and Evolutionary Synthesis, University of Oslo, Oslo, Norway

My contributions: I contributed to developing the model, to producing figures and deriving expressions, and to the writing of the article.

Publication status: In preparation. A draft of the manuscript is attached.

Evolutionarily Stable Strategies (ESS) in a socially age-structured epidemic model

Andreas Eilersen [1,] *, Ruiyun Li **, Nils Chr. Stenseth [2,] **

*Niels Bohr Institute, University of Copenhagen, Denmark
** Centre for Ecological and Evolutionary Synthesis (CEES), Department of Biosciences, University of Oslo, Norway
[1] Corresponding author e-mail: andreaseilersen@nbi.ku.dk
[2] Corresponding author e-mail: n.c.stenseth@mn.uio.no

ABSTRACT

To understand the dynamics of a pathogen we need to understand both its ecology and evolution. Ecological dynamics of many diseases have been investigated in socially age-structured models, not the least COVID-19. However, limited effort has been made to ask evolutionary questions, in particular questions regarding what the Evolutionarily Stable Strategy (or ESS) might be like. Here we ask what the ESS of the infection rate in an age-structured model might be. We demonstrate that the ESS of a pathogen is highly dependent upon the background age-dependent mortality rates but is independent of the demography of the host population. Our findings apply to a general epidemic pathogen in a human population. More importantly, they collectively contribute to understanding the long-term fate of novel pathogens such as SARS-CoV-2 that are currently early in their evolution.

Introduction

The fate of pandemic diseases has long been considered in the field of epidemiology and evolutionary biology. Such considerations have been even higher on the agenda ever since the onset of the COVID-19 pandemic [1] [2]. The initial evolution of an emerging pathogen is far from any evolutionary equilibrium since the pathogen is not optimised for its host. However, historical evidence from other pathogens [3] [4] [5] have shown that in the long-term, less lethal variants of the pathogen will often win over more virulent variants. Investigating how pathogen evolution may modulate epidemic dynamics is therefore a highly relevant research topic. In particular, there is an urgent need to shed light on the evolution of the virulence and transmissibility of emerging pathogens such as SARS-CoV-2 from the initial pandemic phase to the possible endemism.

An Evolutionarily Stable Strategy (ESS) is a strategy where no mutant variant of a pathogen or other organism is able to invade and displace the resident variant. Such strategies have long been investigated in parasite and pathogen systems [6] [7], and ecological systems [8] [9] [10] [11].

Given the many variants of the pathogen, ESS is the key to investigate the possible future of SARS-CoV-2. For example the evolutionarily optimal length of the pre-symptomatic infectious period have been examined [12] [13]. It is worth noting that the disease burdens of many human pathogens are highly stratified with age. Pilot studies have shown that the age pyramid and social interactions over

age groups may have shaped disease dynamics among humans [14 15 16 17 18 19 20]. Therefore, it is vital to explicitly investigate the evolutionary fate of a disease in a socially age-structured population.

By using age-structured ecological epidemiology models, we investigate the expected ESS of a pathogen. We do so under the assumption of a trade-off between the infectivity and the duration of the disease caused by the pathogen. This is a common assumption in ecology [3] and is necessary if we want the pathogen evolution to have a stable state.

Methods and Materials

Model framework

1. The ecological epidemiology model (the Eco-Epi model)

Based on our previous model [21], we develop a susceptible-infected-recovered-susceptible (SIRS) model (fig. 1a) partitioning the population into *n* age classes to simulate the epidemic dynamics in an age-structured population. In the following, we only consider a single strain (hereafter "resident strain").

$$\frac{dS_i}{dt} = \underbrace{a_{i-1}S_{i-1}}_{aging} - \underbrace{a_iS_i}_{aging\ out} - \underbrace{\delta_iS_i}_{non-disease\ deaths} + \underbrace{\omega R_i}_{loss\ of\ immunity} - \underbrace{\sum_j \beta_{ij}S_iI_j}_{infection}$$

$$\frac{dI_i}{dt} = \underbrace{a_{i-1}I_{i-1}}_{aging} + \underbrace{\sum_j \beta_{ij}S_iI_j}_{infection} - \underbrace{a_iI_i}_{aging} - \underbrace{\gamma_iI_i}_{recovery} + \underbrace{d_iI_i}_{disease\ deaths} + \underbrace{\delta_iI_i}_{non-disease\ deaths} \quad (1)$$

$$\frac{dR_{I,i}}{dt} = \underbrace{a_{i-1}R_{i-1}}_{aging} + \underbrace{\gamma_iI_i}_{recovery} - \underbrace{a_iR_i}_{aging} + \underbrace{\delta_iR_i}_{non-disease\ deaths} + \omega R_i$$

Here, a_i is the aging rate for age class i, d_i and δ_i is the disease-induced and background (or non-disease induced) death rate, respectively, γ_i is the recovery rate. ω is the uniform rate of loss of immunity over age groups, and S_0 is the total population size. β_{ij} are infection rates from age class j to i, depending on the rate of interaction between the relevant age classes and on an inherent infectivity of the infected member of age class j. The equation for the first age group needs to be modified to account for new births:

$$\frac{dS_1}{dt} = \sum_{i=1}^{n}(d_i(S_i + I_i) + \delta_i) - a_1S_1 + \omega R_1 - \sum_{i=1}^{n}\beta_{1i}S_1I_i \quad (2)$$

2. The extended Eco-Epi model

To study potential invasions from new mutants, we extend the above model by including an additional strain (hereafter "mutant strain") (fig. 1b). Accordingly, $\widetilde{I}$ is the number of infections that are induced by the mutant strain in age group i. Of note, the mutant strain may have different infectivities b_{ij} and recovery rates g_i as compared with the resident strain. We assume that the disease mortality, loss of immunity, and aging rate are the same for both the resident and mutant strains:

$$\frac{dS_i}{dt} = a_{i-1}S_{i-1} - a_iS_i - \delta_iS_i + \omega(R_i + \widetilde{R}_\iota) - \sum_j \beta_{ij}S_iI_j - \sum_j b_{ij}S_i\widetilde{I}_J$$

$$\frac{dI_i}{dt} = a_{i-1}I_{i-1} + \sum_j \beta_{ij}S_iI_j - (a_i + \gamma_i + d_i + \delta_i)I_i$$

$$\frac{dR_i}{dt} = a_{i-1}R_{i-1} + \gamma_iI_i - (a_i + \delta_i + \omega)R_i \qquad (3)$$

$$\frac{d\widetilde{I}_\iota}{dt} = a_{i-1}\widetilde{I}_{i-1} + \underbrace{\sum_j b_{ij}S_i\widetilde{I}_J}_{mutant\ infection} - (a_i + g_i + d_i + \delta_i)\widetilde{I}_\iota$$

$$\frac{d\widetilde{R}_\iota}{dt} = a_{i-1}\widetilde{R}_{i-1} + \underbrace{g_i\widetilde{I}_\iota}_{mutant\ recovery} - (a_i + \delta_i + \omega)\widetilde{R}_\iota$$

With the extended eco-epi model, we investigate the conditions for when a mutant strain will be unable to invade a population and thus when the ESS will occur.

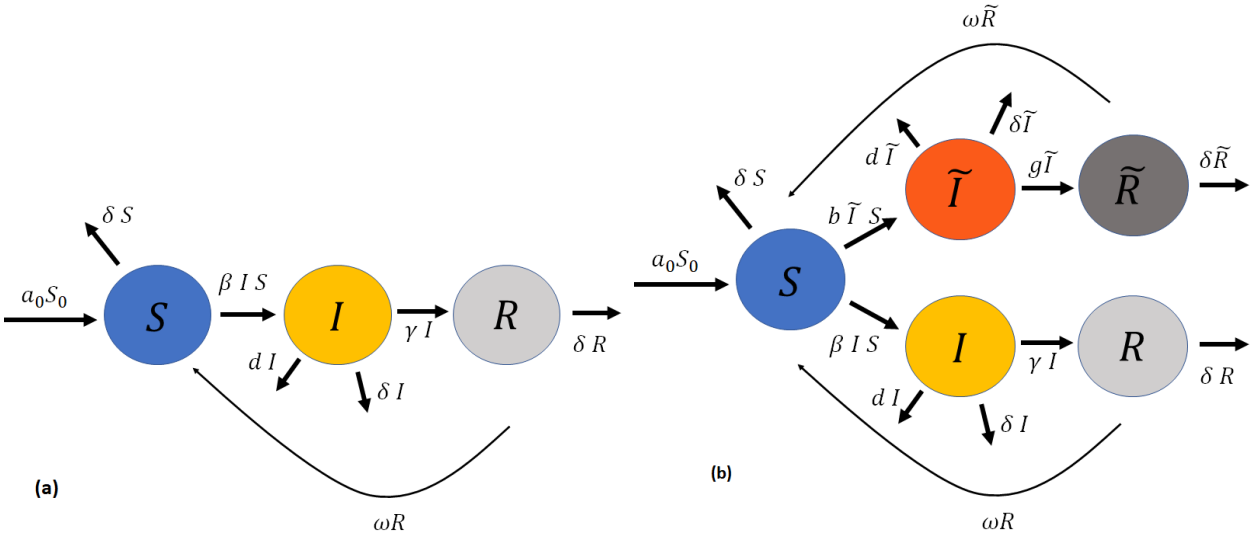


Figure 1. *An illustration of the models and parameters. (a) shows the eco-epi model with a single strain, while (b) shows the extended eco-epi model with resident and mutant strains. The age structure is not shown in this illustration.*

The analysis – Deriving the evolutionarily stable strategy (ESS)

We define an ESS as a strategy where the resident strain I dominates the epidemic dynamics in a population such that a new mutant strain $\widetilde{I}$ fails to invade and induce an outbreak regardless of its biological characteristics [10]. Mathematically, this means that at the ESS, the equilibrium disease state $\left(S, I, R, \widetilde{I}, \widetilde{R}\right) = (S, I, R, 0, 0)$ will be linearly stable with the effective reproduction number for the resident strain $R_{eff}(I) = 1$ and that for the mutant strain $R_{eff}\left(\widetilde{I}\right) < 1$.

The effective reproductive number for each of the variants can be calculated as the largest eigenvalue of their next-generation matrices. For simplicity, we assume that age-specific characteristics of the susceptibles and the infected individuals contribute multiplicatively to the infection rates, i.e., $\beta_{ij} = x_i y_j$. With this assumption, the reproductive number is the trace of the next generation matrix:

$$R_{eff} = \sum_{j=1}^{n} \sum_{i=1}^{j} \frac{\beta_{ij} \widehat{S}_\iota \prod_{k=i}^{j-1} a_k}{\prod_{k=i}^{j} (a_k + \gamma_k + d_k + \delta_k)} \ , \qquad (4)$$

Here, $\widehat{S}_\iota$ denotes the disease-free equilibrium populations given by

$$\widehat{S_1} = \frac{S_0}{1 + \sum_{k=2}^{n} \alpha_k}$$

$$\widehat{S_\iota} = \frac{\alpha_i S_0}{1 + \sum_{k=2}^{n} \alpha_k}$$

where

$$\alpha_i = \frac{\prod_{k=2}^{i} a_{k-1}}{a_k + \delta_k}$$

Let us express the reproductive number of the invader using the above equation:

$$R_{inv} = \sum_{j=1}^{n} \sum_{i=1}^{j} \frac{\beta_{ij}(\widetilde{\gamma_j}) \widehat{S}_\iota(\gamma_i) \prod_{k=i}^{j-1} a_k}{\prod_{k=i}^{j} (a_k + \tilde{\gamma}_k + d_k + \delta_k)} \ . \qquad (5)$$

Now, the ESS occurs when $R_{inv} < 1$ for all $\tilde{\gamma} \neq \gamma$. At the ESS the invader cannot gain an advantage over the resident strain by changing γ. It is therefore a necessary condition that

$$\nabla_{\tilde{\gamma}} R_{inv}|_{\tilde{\gamma}=\gamma} = 0 \, .$$

All terms that do not depend on $\tilde{\gamma}_l$ vanish when taking the derivative with respect to $\tilde{\gamma}_l$, meaning that only terms where $j > l$ remain. We can now write up n equations

$$\frac{\partial}{\partial \widetilde{\gamma_l}} \sum_{j=l}^{n} \sum_{i=1}^{j} \frac{\beta_{ij}(\widetilde{\gamma_j}) \widehat{S}_\iota(\gamma_i) \prod_{k=i}^{j-1} a_k}{\prod_{k=i}^{j} (a_k + \tilde{\gamma}_k + d_k + \delta_k)} = 0 \, . \qquad (6)$$

Infectivity ($\beta = \beta(\gamma)$and $b = \beta(\tilde{\gamma})$) and recovery rate ($\gamma$ and $g = \tilde{\gamma}$) differ among variants. The terms of this sum all contain a number of factors of the form $\frac{1}{a_j+\gamma_j+d_j+\delta_j} \cdot \frac{a_{j-1}}{a_{j-1}+\gamma_{j-1}+d_{j-1}+\delta_{j-1}} \cdots \frac{a_i}{a_i+\gamma_i+d_i+\delta_i}$. If we assume that the disease dynamics happen on a much faster timescale than aging, any term with $i \neq j$ – which contains multiple such factors – will be much smaller than 1. We can therefore neglect these terms and write the above equations as

$$\frac{\partial}{\partial\tilde{\gamma}_l}\frac{\beta_{ll}(\tilde{\gamma}_l)\widehat{S}_l(\gamma_l)}{(a_l+\tilde{\gamma}_l+d_l+\delta_l)}|_{\tilde{\gamma}=\gamma} = 0\,. \qquad (7)$$

Carrying out the derivative, we end up with

$$\beta'_{ii}(\gamma_i)\cdot\widehat{S}_i(\gamma_i)\cdot(a_i+\gamma_i+d_i+\delta_i) = \beta_{ii}(\gamma_i) \qquad (8)$$

For the concept of an ESS to be meaningful here, we have to assume that the infectivity is a function of some evolutionary parameter(s). For simplicity, we apply the trade-off hypothesis [4] and thus assume a trade-off between the infectivity of the disease (β) and its recovery rate (γ). With a concave function $\beta(\gamma)$, we can now find the ESS by solving the equation (8) for each of the age groups. This is a generalisation of a well-known result from evolutionary ecology to a model with multiple age classes [1].

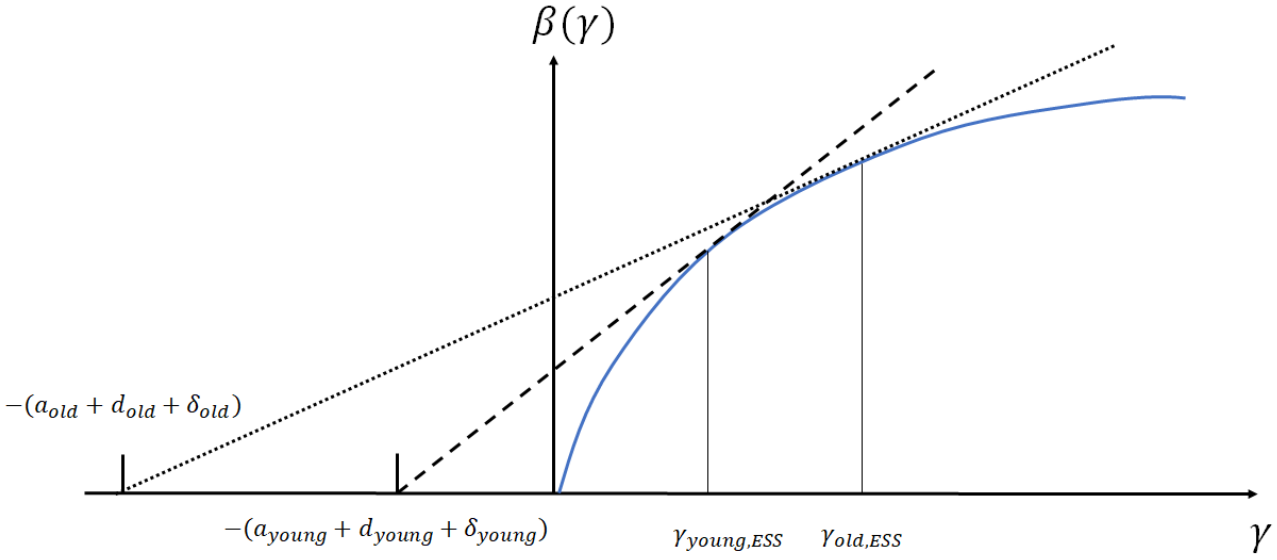


Figure 2. *An illustration of eq. 8. Solving the set of equations is equivalent to finding the point where a line through $-(a_i+\delta_i+d_i)$ is tangent to the trade-off function $\beta(\gamma_i)$. The higher the background and disease mortalities are, the higher the value of γ_i will be at the ESS.*

Assuming that the infection rate from group i to j is proportional to some pairwise contact rate between these groups, multiplied by an infectivity function $\beta(\gamma_i)$, equation (8) becomes

$$c_{ii}\beta'(\gamma_i)\cdot\widehat{S}_i(\gamma_i)\cdot(a_i+\gamma_i+d_i+\delta_i) = c_{ii}\beta(\gamma_i) \qquad (9)$$

where c_{ii} is the contact rate within group i. We see that the contact rate, c_{ii}, cancels out. This means that the ESS of the disease does not vary with the assumed form of β_{ij} and association between the rate of infection and contact.

Results and discussion

The age-specific values of β and γ at the ESS are associated with the corresponding disease fatality rate and the background mortality as demonstrated in eq. (8). By characterizing the ESS using age-specific $\beta_{ii}(\gamma_i)$, our study indicates that an ESS will only exist if $\beta_{ii}(\gamma_i)$ are concave functions since $\gamma = 0$ or $\gamma = \infty$ would otherwise always be optimal for the disease. Thus, a higher background or disease-related mortality rate implies a higher β and γ at the ESS in a given age group. This means that evolution would inflate the infection rate for age groups with fatal disease or higher background mortality, whereas evolution would lower the infection rate for age groups with a low mortality (fig. 2).

Our results imply that the ideal state of a pathogen is to be very infectious for a short period of time in population groups with a high background and disease-related mortality, and less infectious for a long time in other groups. Consistent with our findings, previous studies have shown substantial numbers of asymptomatic (but still infectious) COVID-19 cases in young people [22] [23]. A study has also suggested that COVID-19 patients with risk factors such as high age or obesity shed more virus aerosols and are thus more likely to be superspreaders [24].

We show that at the ESS, the intensity of the infectious period should increase with background mortality. This finding is, from the perspective of evolutionary strategy, consistent with the severity of the disease increasing in proportion with the age of the patient [25]. If the disease is shorter but more intense among people with a higher background mortality, we might expect the risk of death from the disease to scale with background mortality. More importantly, the already highly variable fatality rate of COVID-19 leads us to predict that it might end up having disease progressions that are particularly stratified by age, more so than other diseases (see fig. 2). On the other hand, the maximum R will for some realistic assumptions about β be higher among young people due to their lower background mortality. If $\beta(\gamma)$ takes a form as shown in fig. 2, with $\beta(\gamma) \propto \gamma^p$ where $p < 1$, $R = \frac{\beta(\gamma)}{\gamma}$ will be a decreasing function of γ. This would suggest that young people might inherently be better drivers of an epidemic like the present one than older people, at least as it evolves to become endemic.

So far, at least one emerging variant of SARS-CoV-2, namely B.1.1.7, has been shown to have a higher effective reproduction number among younger people [26]. This may support our hypothesis. However, whether this truly agrees with our predictions will depend on how the new, more infectious variants affect the duration of the disease in different age groups.

Conclusion

In this contribution we have derived conditions for an evolutionarily stable strategy for a pathogen spreading in an age-structured population. We find that these conditions vary with the background

mortality at the given age group, and with the fatality rate of the disease. Groups with higher mortalities are predicted to end up with shorter, more intense infectious periods given our assumption of a trade-off between disease infectivity and duration. The social structure and interaction rates of society are found to have little influence on the final ESS, though they may still influence the path the pathogen takes to get there.

Finally, our results are only specific to age-based social structure because we include aging in our model. If we drop aging, our results can be generalised to any other kind of socially stratified society. Our results are therefore sufficiently general to be applicable to a variety of pathogens that spread in a society with a rich social structure. These findings may help shed some light on the eventual fate of pandemic diseases, and thus what direction we should expect to see a new disease evolve.

Acknowledgements

We would like to thank Yihan Cao, Ottar N. Bjørnstad, Bruce Levin, Jason D. Whittington, and Boris V. Schmid for enlightening discussions, and Sebastian Schreiber for assistance in the analytical work presented here. This study received funding from the Research Council of Norway through the COVID-19 Seasonality Project (Reference No. 312740), and from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (Grant Agreement No. 740704).

Author contributions
NCS and AE designed the study. RL built the epidemiological model. AE carried out the evolutionary analysis. All authors edited and approved the manuscript.

---

[1] Wu, A., Wang, L., Zhou, H. Y., Ji, C. Y., Xia, S. Z., Cao, Y., ... & Cheng, G. (2021). One year of SARS-CoV-2 evolution. *Cell Host & Microbe*, *29*(4), 503-507.

[2] Saad-Roy, C. M., Morris, S. E., Metcalf, C. J. E., Mina, M. J., Baker, R. E., Farrar, J., ... & Wagner, C. E. (2021). Epidemiological and evolutionary considerations of SARS-CoV-2 vaccine dosing regimes. *Science*, *372*(6540), 363-370.

[3] Alizon, S., Hurford, A., Mideo, N., & Van Baalen, M. (2009). Virulence evolution and the trade-off hypothesis: history, current state of affairs and the future. *Journal of evolutionary biology*, *22*(2), 245-259.

[4] Acevedo, M. A., Dillemuth, F. P., Flick, A. J., Faldyn, M. J., & Elderd, B. D. (2019). Virulence-driven trade-offs in disease transmission: A meta-analysis. *Evolution*, *73*(4), 636-647.

[5] Knell, R. J. (2004). Syphilis in Renaissance Europe: rapid evolution of an introduced sexually transmitted disease? *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *271*(suppl_4), S174-S176.

[6] Anderson, R. M., & May, R. M. (1982). Coevolution of hosts and parasites. *Parasitology*, *85*(Pt 2), 411-426.

[7] Cortez, M. H. (2013). When does pathogen evolution maximize the basic reproductive number in well-mixed host–pathogen systems?. *Journal of Mathematical Biology*, *67*(6), 1533-1585.

[8] Maynard Smith, J., *On Evolution* (Edinburgh University Press, 1972), p. 125.

[9] Maynard Smith, J., & Price, G. R. (1973). The logic of animal conflict. *Nature*, *246*(5427), 15-18.

[10] Reed, J., & Stenseth, N. C. (1984). On evolutionarily stable strategies. *Journal of theoretical biology*, *108*(4), 491-508.

[11] Grunert, K., Holden, H., Jakobsen, E. R., & Stenseth, N. C. (2021). Evolutionarily stable strategies in stable and periodically fluctuating populations: The Rosenzweig–MacArthur predator–prey model. *Proceedings of the National Academy of Sciences*, *118*(4).

[12] Saad-Roy, C. M., Grenfell, B. T., Levin, S. A., Pellis, L., Stage, H. B., Van Den Driessche, P., & Wingreen, N. S. (2021). Superinfection and the evolution of an initial asymptomatic stage. *Royal Society open science*, *8*(1), 202212.

[13] Saad-Roy, C. M., Wingreen, N. S., Levin, S. A., & Grenfell, B. T. (2020). Dynamics in a simple evolutionary-epidemiological model for the evolution of an initial asymptomatic infection stage. *Proceedings of the national academy of sciences*, *117*(21), 11541-11550.

[14] Kuchler, T., Russel, D., & Stroebel, J. (2020). *The geographic spread of COVID-19 correlates with the structure of social networks as measured by Facebook* (No. w26990). National Bureau of Economic Research.

[15] Nande, A., Adlam, B., Sheen, J., Levy, M. Z., & Hill, A. L. (2021). Dynamics of COVID-19 under social distancing measures are driven by transmission network structure. *PLoS computational biology*, *17*(2), e1008684.

[16] Grabowski, A., & Kosiński, R. A. (2004). Epidemic spreading in a hierarchical social network. *Physical Review E*, *70*(3), 031908.

[17] Frias-Martinez, E., Williamson, G., & Frias-Martinez, V. (2011, October). An agent-based model of epidemic spread using human mobility and social network information. In *2011 IEEE third international conference on privacy, security, risk and trust and 2011 IEEE third international conference on social computing* (pp. 57-64). IEEE.

[18] Klepac, P., Kissler, S., & Gog, J. (2018). Contagion! the bbc four pandemic–the model behind the documentary. *Epidemics*, *24*, 49-59.

[19] Nielsen, B. F., Simonsen, L., & Sneppen, K. (2021). COVID-19 superspreading suggests mitigation by social network modulation. *Physical Review Letters*, *126*(11), 118301.

[20] Block, P., Hoffman, M., Raabe, I. J., Dowd, J. B., Rahal, C., Kashyap, R., & Mills, M. C. (2020). Social network-based distancing strategies to flatten the COVID-19 curve in a post-lockdown world. *Nature Human Behaviour*, *4*(6), 588-596.

[21] R. Li, N. Stenseth, In press, *Science Advances*

[22] Y. Wang, Y. Liu, L. Liu, X. Wang, N. Luo, L. Ling Clinical outcome of 55 asymptomatic cases at the time of hospital admission infected with SARS-Coronavirus-2 in Shenzhen, China, J Infect Dis, 221 (11) (2020), pp. 1770-1774

[23] Z. Hu, C. Song, C. Xu, G. Jin, Y. Chen, X. Xu, *et al.* Clinical characteristics of 24 asymptomatic infections with COVID-19 screened among close contacts in Nanjing, China, Sci China Life Sci, 63 (5) (2020), pp. 706-711

[24] Edwards, D. A., Ausiello, D., Salzman, J., Devlin, T., Langer, R., Beddingfield, B. J., ... & Roy, C. J. (2021). Exhaled aerosol increases with COVID-19 infection, age, and obesity. *Proceedings of the National Academy of Sciences*, *118*(8).

[25] Onder, G., Rezza, G., & Brusaferro, S. (2020). Case-fatality rate and characteristics of patients dying in relation to COVID-19 in Italy. *Jama*, *323*(18), 1775-1776.

[26] Volz, E., Mishra, S., Chand, M., Barrett, J. C., Johnson, R., Geidelberg, L., ... & Ferguson, N. M. (2021). Assessing transmissibility of SARS-CoV-2 lineage B. 1.1. 7 in England. *Nature*, 1-17.

Supplementary figure to the article “Evolutionarily Stable Strategies (ESS) in a socially age-structured epidemic model”

The figure below shows the trajectories of the evolutionary parameters of different age groups as we algorithmically maximise the basic reproductive number of a disease in age-structured populations. Contact rates between different age classes are based on contact data also used in (Stenseth & Li, 2021) from various countries. We take the inherent infectivity β_j of infected individuals of class j to be given by the tradeoff function $\beta_j = A\gamma_j^p$, where we here set $A = 2$ and $p = 0.9$. The infection rate between groups i and j thus becomes $\beta_{ij} = Ac_{ij}\gamma_j^p$ where c_{ij} is the contact rate. The figure shows that the endpoint of R_0 maximisation, here taken to be the ESS, is highly similar, even in countries with wildly different demographies and social structures.

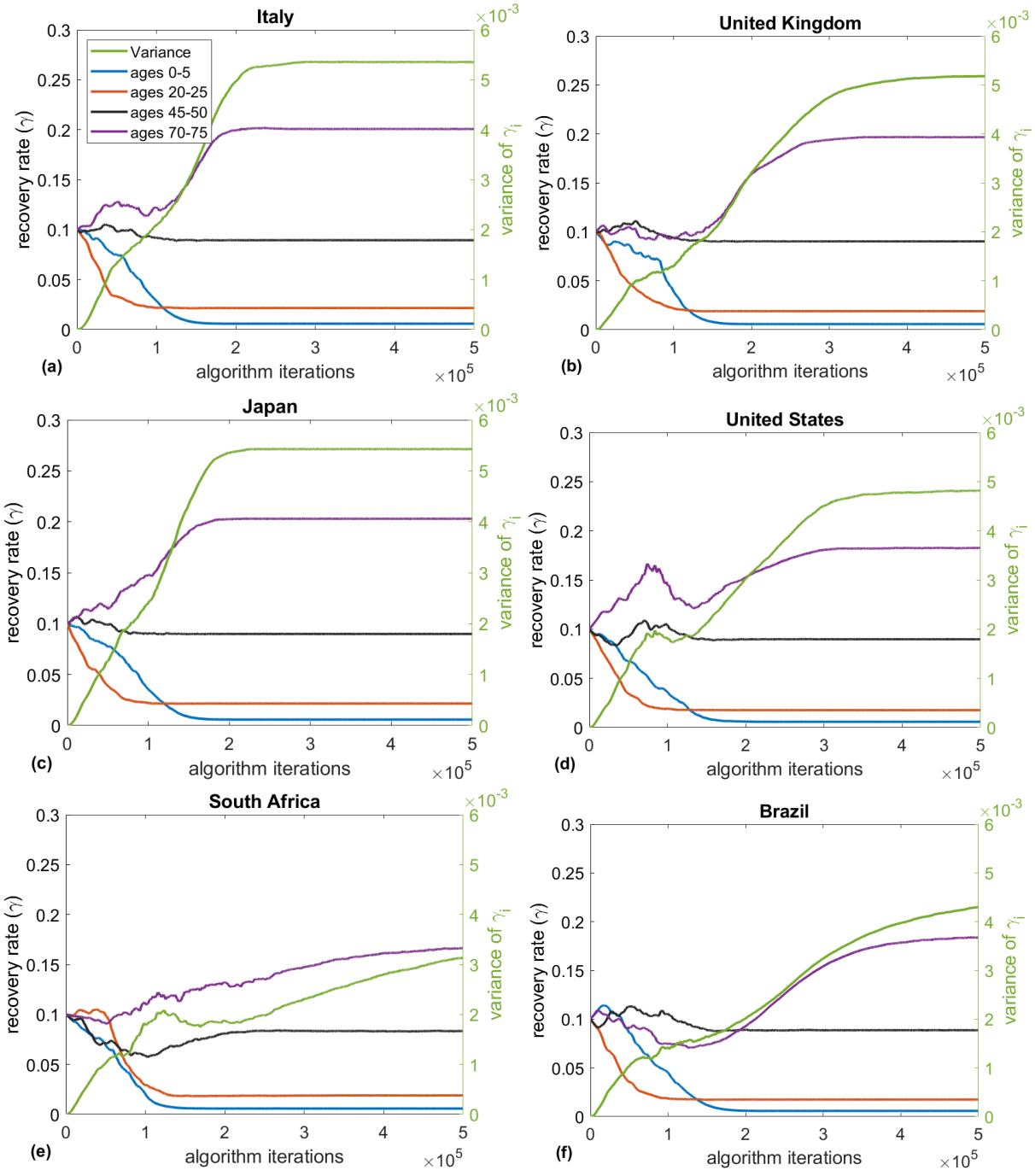


Figure S1. *Trajectories of the pathogen as we algorithmically determine the ESS in six countries. The R_{eff}-maximisation algorithm is applied to countries with older populations (Italy, UK and Japan) and younger populations (South Africa, Brazil and the US). The assumed trade-off function is $\beta = 2 \cdot \gamma^{0.9}$. This very slightly concave function was chosen to for illustrative purposes, to let the values of γ_i be obviously different at the ESS. The background and disease related mortality rates are set to $\delta_i + d_i = 10^{-4} \cdot i^2$. The graphs are averages of 100 runs of the algorithm.*

Tradeoff between speed and infectivity in pathogen evolution

Authors: Andreas Eilersen* and Kim Sneppen*

*The Niels Bohr Institute, University of Copenhagen, Copenhagen, Denmark

My contributions: I contributed to developing the model and writing associated programs, to producing figures and deriving expressions, and to the writing of the article.

Publication status: In preparation. A draft of the manuscript is attached.

Tradeoff between speed and infectivity in pathogen evolution - DRAFT

Andreas Eilersen[1,*] **and Kim Sneppen**[1]

[1]Niels Bohr Institute, University of Copenhagen, Copenhagen, Denmark
[*]andreaseilersen@nbi.ku.dk

ABSTRACT

Given the present pandemic and the constantly arising new variants of SARS-CoV-2, there is an urgent need to understand the factors driving disease evolution. In this paper, we will investigate the tradeoff between the speed at which a disease progresses and its infectivity. Using SIR and agent-based models, we show that in the exponential growth phase of an epidemic, there will be an optimal duration of new disease variants, balancing the advantage of developing fast with the advantage of infecting many new people. In the endemic state, this optimum disappears and lasting longer is always advantageous for the disease. However, if we take into account the possibility of quarantining the infected, this leads to a new optimum disease duration. This work provides some much needed understanding of the factors that affect the evolution of new SARS-CoV-2 variants and other emerging pathogens.

Introduction

Since the emergence of SARS-CoV-2, multiple variants of the virus with significantly different dynamics have arisen. The variants have supplanted each other in successive waves, letting us observe as the disease has gradually evolved. Since we are in the midst of this evolutionary race towards more infectious, less immunogenic, and faster spreading variants, it is highly necessary to understand what drives new variants to be successful. In the present article, we will focus on the tradeoff between the duration of the infectious period and latency time of a disease, and the number of secondary cases each infected individual generates. We assume that each infected individual transmits the disease with a roughly constant rate for each day of the infectious period. This means that a long disease duration should lead to a higher effective reproductive number, R_{eff}, that is, to more secondary infections. However, a long disease duration might also eventually become a disadvantage to the disease, as it makes it slow to develop. This is exacerbated if there is a link between the duration of latency time and infectious period. We will derive relations and create an agent-based model to show when an optimum disease duration exists.

Some work has already been done on modelling the evolution of SARS-CoV-2 and other similar pathogens. Saad-Roy *et al.* studied the evolution of a presymptomatic infectious state under the assumption that such a state is less infectious[1], or in the context of superinfection and within-host competition[2]. In addition, the relationship between the duration of a disease or parasite infection and the infection rate has been studied under the assumption of a tradeoff function between the two[3,4]. These findings have analogues in other ecological relationships, such as predation[5,6]. However, the possibility that having a long infectious period might simultaneously be an advantage and a disadvantage for a disease has not been studied in detail.

Model setup

We assume that the infection rate of a disease is roughly constant for the duration of the infectious period, giving a linear relationship between disease duration and number of secondary cases. This is likely not entirely naturalistic[7,8], but since a longer duration of the infectious period leads to more opportunities of passing on the infection, there must be some positive relationship between the two.

Analytically, we will exclusively look at the initial, exponential growth phase of the epidemic. In this phase, it will be an advantage of a disease to be fast-growing, whereas in a full SIR simulation of an epidemic, the disease with the highest R_{eff} will win. However, in the case when a society is far from herd immunity, exponential growth will occur when a disease starts spreading. This has for example been the case for most of the world for most of the COVID-19 pandemic. Mitigation efforts has artificially kept society far from herd immunity, meaning that whenever R_{eff} grew above 1 and the disease started spreading, we saw exponential growth. The scenario we look at is thus not as limited as one might think.

When we wish to study the evolution of the disease in an endemic state with a high degree of existing immunity in the population, we instead use an agent-based model. In this model, we let agents randomly infect each other, with some small probability of producing a mutant strain for each infection. A new mutant will have a duration of infectiousness that is

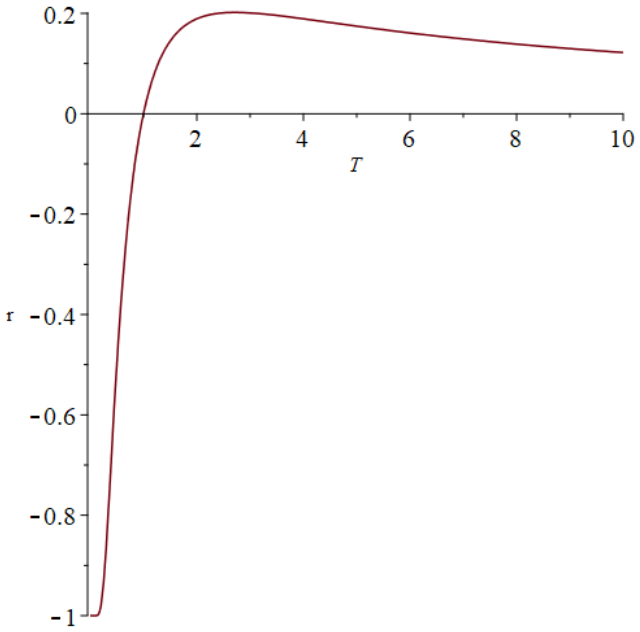


Figure 1. A plot of growth rate as a function of disease duration T as derived in eq. 3. Here, we set $c = \beta = 1$. We see clearly that the growth rate has an optimum.

slightly different from its parent strain. We will then run the simulation over a long period of time to see which strains end up dominating.

Finally, we add a quarantine rate to the agent-based model. This is supposed to represent how individuals have some chance of becoming symptomatic, being contact traced, or otherwise being diagnosed for each day of illness. We should therefore expect that people suffering from a very long-lasting infectious disease will eventually self-quarantine. This quarantine probability might change the evolutionary dynamics of the disease.

Results

Optimum disease duration for exponential growth

In the exponential growth phase of an epidemic, the number of infected individuals is approximately

$$I(t) \approx R_0^{\frac{t}{\tau+T}} = (1+r)^t \tag{1}$$

where R_0 is the basic reproductive number of the disease, τ is the latency time, T is the duration of the infectious period, and r is the epidemic growth rate. We now want to calculate the growth rate as a function of these parameters:

$$r = R_0^{\frac{1}{\tau+T}} - 1 \tag{2}$$

In the exponential growth phase, the variant with the highest growth rate will quickly come to dominate. If each infected individual transmits the disease at a constant daily rate β, and we assume some constant relationship c between the latency time and duration of infectiousness, we can write

$$r = (\beta T)^{\frac{1}{(1+c)T}} - 1. \tag{3}$$

This function has a local maximum for $T = e/\beta$. A plot of the growth rate as functions of T for various β and c can be seen in fig. 1. An illustration of the initial exponential growth phase for different disease durations can be seen in fig. 2. In a situation where the disease is growing exponentially, e.g., when an epidemic is breaking out or control measures are failing, the variants that balance the need to be fast with the need to be highly contagious will win. This is, however, not always the case as we shall see in the following.

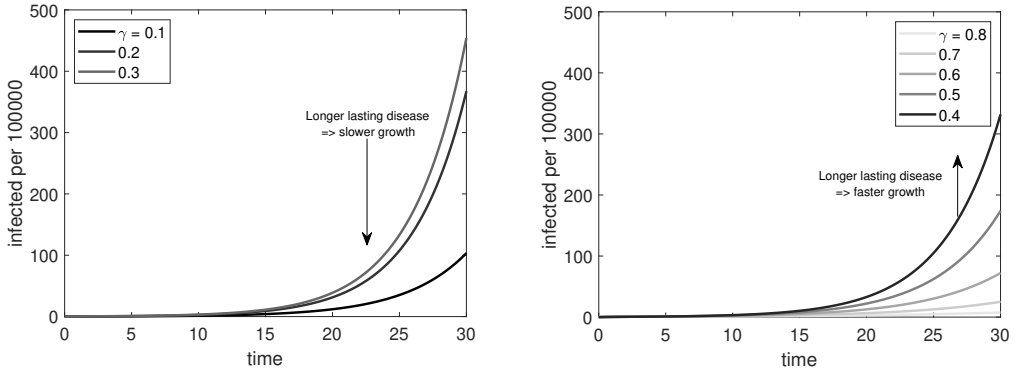


Figure 2. A simple simulation of the initial exponential phase of an epidemic given different values of the recovery rate γ, i.e., different disease durations. We see that up to a certain optimum, the disease grows faster for larger values of γ, but afterwards a larger γ leads to slower exponential growth. This illustrates the analytically derived growth rate shown in fig. 1.

The unmitigated endemic state

In the case where the epidemic is completely unmitigated and infectious individuals are never quarantined, it will always be possible to increase the infectivity of the disease by lengthening disease duration. From our agent-based simulations (fig. 3 (a)) we see that the successfully invading variants will continuously develop lower and lower recovery rates, that is, longer and longer disease duration. This happens regardless of the initial disease duration. Therefore, we conclude that a pathogen in an unmitigated endemic state will always evolve to last longer. In practice, there will always be some form of mitigation, which is what we investigate below.

Endemic state with quarantine

Finally, we consider the case when each infectious, symptomatic individual has some probability p of isolating themselves for each day of the infectious period. In this case, we see from simulations (fig. 3(b)) that there will be an optimum disease duration which new variants approach over time. If we start out with a pathogen with a longer disease duration, it will develop towards shorter disease durations and vice versa, as opposed to the case above. Therefore, we can conclude that in the case where infected individuals sooner or later will isolate themselves and stop infecting, there will again be some optimal disease duration.

Discussion

The results of this analysis show that in some cases, being fast-acting can be an evolutionary advantage for a pathogen, even if it comes at the cost of a lower reproduction number. This includes the case when the population is far from herd immunity and the number of infected is growing exponentially, or if the infected have some probability each day of self-quarantining. These situations are expected to occur frequently in real life. For example, during the COVID-19 pandemic mitigation efforts in various locations often kept the local R_{eff} at or below 1. When such efforts failed or were relaxed, local epidemics entered a new exponential growth phase. In the case of most infections, we would also expect the onset of symptoms to increase the chance that individuals stay home or are bedridden, effectively self-quarantining.

These results are particularly interesting in the context of the pandemic, especially as they may help explain the swift takeover and large impact of the delta variant. The Delta variant has been shown by some studies to have a somewhat shorter incubation time and significantly shorter generation time[9] than the wild-type virus[10,11]. Other studies have, however, questioned whether the difference in incubation time is significant[12]. If the new variant is truly faster than the wild-type, it may help explain its rapid takeover, although its higher infectiousness is of course the primary reason. More importantly it may help explain why countries that had previously successfully pursued a zero-COVID strategy came to struggle containing the new variant. If a variant is faster, it will according to this model do much better than a slower variant in a situation with a high probability of quarantine for each day of infection.

The direction in which the evolution of SARS-CoV-2 is moving is a question which urgently needs to be answered to understand the future of the disease and its mitigation and the end of the pandemic. Here, we have examined the specific topic of the duration of the incubation and infectious periods of COVID-19. We show that under a couple of realistic assumptions,

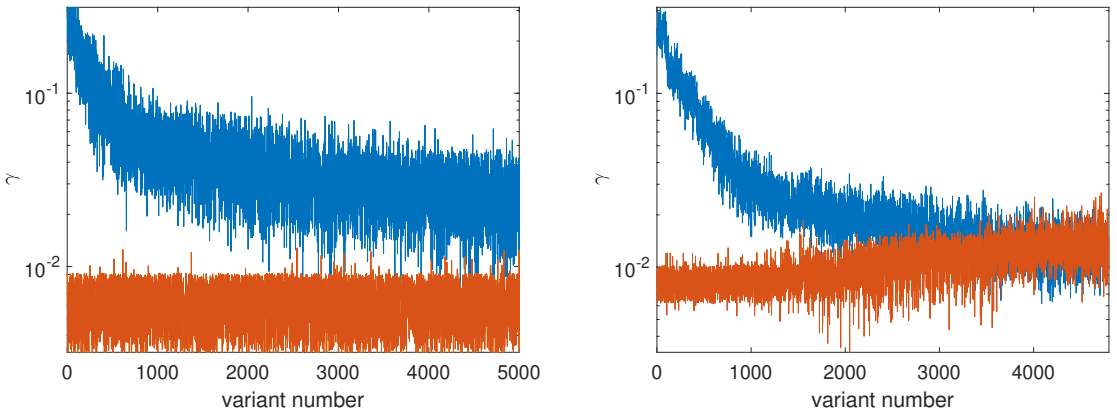


Figure 3. Evolution of variants over time. The figure shows the recovery rate γ (i.e., the inverse disease duration) of new variants (a) in the system with no mitigation at all and (b) in the system where symptomatic individuals self-quarantine with some rate $p = 0.1/day$. β is here set to 0.5. We see that when people are able to self-quarantine, the variants gradually evolve towards some evolutionarily stable state, whereas this does not appear to be the case if there is no quarantine.

there is a tradeoff between disease infectivity and speed. A fast disease may be able to outcompete a slower, more infectious disease in the case with exponential growth in number of infected or with quarantine of symptomatic individuals. Previously, it has been shown that mitigation strategies may interact with pathogen evolution by disproportionately affecting superspreaders, benefitting homogeneously spreading diseases[13]. The present work shows that mitigation efforts may also drive the pathogen to evolve towards a shorter disease duration.

References

1. Saad-Roy, C. M., Wingreen, N. S., Levin, S. A. & Grenfell, B. T. Dynamics in a simple evolutionary-epidemiological model for the evolution of an initial asymptomatic infection stage. *Proc. national academy sciences* **117**, 11541–11550 (2020).
2. Saad-Roy, C. M. *et al.* Superinfection and the evolution of an initial asymptomatic stage. *Royal Soc. open science* **8**, 202212 (2021).
3. Anderson, R. M. & May, R. M. Coevolution of hosts and parasites. *Parasitology* **85**, 411–426 (1982).
4. Alizon, S., Hurford, A., Mideo, N. & Van Baalen, M. Virulence evolution and the trade-off hypothesis: history, current state of affairs and the future. *J. evolutionary biology* **22**, 245–259 (2009).
5. Reed, J. & Stenseth, N. C. On evolutionarily stable strategies. *J. theoretical biology* **108**, 491–508 (1984).
6. Grunert, K., Holden, H., Jakobsen, E. R. & Stenseth, N. C. Evolutionarily stable strategies in stable and periodically fluctuating populations: The rosenzweig–macarthur predator–prey model. *Proc. Natl. Acad. Sci.* **118** (2021).
7. Lehtinen, S., Ashcroft, P. & Bonhoeffer, S. On the relationship between serial interval, infectiousness profile and generation time. *J. Royal Soc. Interface* **18**, 20200756 (2021).
8. van Kampen, J. J. *et al.* Duration and key determinants of infectious virus shedding in hospitalized patients with coronavirus disease-2019 (covid-19). *Nat. communications* **12**, 1–6 (2021).
9. Zhang, M. *et al.* Transmission dynamics of an outbreak of the covid-19 delta variant b. 1.617. 2—guangdong province, china, may–june 2021. *China CDC weekly* **3**, 584 (2021).
10. Linton, N. M. *et al.* Incubation period and other epidemiological characteristics of 2019 novel coronavirus infections with right truncation: a statistical analysis of publicly available case data. *J. clinical medicine* **9**, 538 (2020).
11. Griffin, J. *et al.* Rapid review of available evidence on the serial interval and generation time of covid-19. *BMJ open* **10**, e040263 (2020).
12. Pung, R., Mak, T. M., Kucharski, A. J. & Lee, V. J. Serial intervals in sars-cov-2 b. 1.617. 2 variant cases. *The Lancet* **398**, 837–838 (2021).

13. Nielsen, B. F., Eilersen, A., Simonsen, L. & Sneppen, K. Lockdowns exert selection pressure on overdispersion of sars-cov-2 variants. *medRxiv* (2021).

Acknowledgements

Our research has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme under Grant Agreement No. [740704].

Author contributions statement

KS proposed the models. AE derived the expressions and wrote the initial draft of the manuscript. Both KS and AE edited the manuscript.

Bibliography

[1] S. Alizon, A. Hurford, N. Mideo, and M. Van Baalen. Virulence evolution and the trade-off hypothesis: history, current state of affairs and the future. *Journal of evolutionary biology*, 22 (2):245–259, 2009.

[2] R. M. Anderson and R. M. May. Coevolution of hosts and parasites. *Parasitology*, 85(2):411–426, 1982.

[3] R. M. Anderson and R. M. May. *Infectious diseases of humans: dynamics and control*. Oxford university press, 1992.

[4] V. Andreasen, C. Berrig, S. Ørskov, R. K. Pedersen, K. A. Krogfelt, L. Simonsen, A. Eilersen, B. F. Nielsen, K. Sneppen, and C. S. Benn. Muligheder for gentagen anvendelse af antigentests til reduktion af smitte på skoler og arbejdspladser. 2021.

[5] J. M. Cable, B. J. Enquist, and M. E. Moses. The allometry of host-pathogen interactions. *PloS one*, 2(11):e1130, 2007.

[6] C. H. Calisher, J. E. Childs, H. E. Field, K. V. Holmes, and T. Schountz. Bats: important reservoir hosts of emerging viruses. *Clinical microbiology reviews*, 19(3):531–545, 2006.

[7] M. Choisy, S. P. Brown, K. D. Lafferty, and F. Thomas. Evolution of trophic transmission in parasites: why add intermediate hosts? *The American Naturalist*, 162(2):172–181, 2003.

[8] A. Dobson. Population dynamics of pathogens with multiple host species. *the american naturalist*, 164(S5):S64–S78, 2004.

[9] A. Eilersen. *Disease in predator-prey systems*. Niels Bohr Institute, Copenhagen University, 2018.

[10] A. Eilersen and K. Sneppen. Applying allometric scaling to predator-prey systems. *Physical Review E*, 99(2):022405, 2019.

[11] A. Eilersen and K. Sneppen. Cost–benefit of limited isolation and testing in covid-19 mitigation. *Scientific reports*, 10(1):1–7, 2020.

[12] A. Eilersen and K. Sneppen. The uneasy coexistence of predators and pathogens. *The European Physical Journal E*, 43(7):1–7, 2020.

[13] A. Eilersen and K. Sneppen. Sars-cov-2 superspreading in cities vs the countryside. *Apmis*, 2021.

[14] A. Eilersen and K. Sneppen. Tradeoff between speed and infectivity in pathogen evolution. *In preparation*, 2021.

[15] A. Eilersen, M. H. Jensen, and K. Sneppen. Chaos in disease outbreaks among prey. *Scientific reports*, 10(1):1–7, 2020.

[16] A. Eilersen, R. Li, N. C. Stenseth, S. Schreiber, O. Bjørnstad, J. Wittington, Y. Cao, and B. Schmid. Evolutionarily stable strategies (ess) in a socially age-structured epidemic model. *In preparation*, 2021.

[17] A. Elbanna, G. N. Wong, Z. J. Weiner, T. Wang, H. Zhang, Z. Liu, A. V. Tkachenko, S. Maslov, and N. Goldenfeld. Entry screening and multi-layer mitigation of covid-19 cases for a safe university reopening. *MedRxiv*, 2020.

[18] A. Endo et al. Estimating the overdispersion in covid-19 transmission using outbreak sizes outside china. *Wellcome open research*, 5, 2020.

[19] N. Ferguson, D. Laydon, G. Nedjati-Gilani, N. Imai, K. Ainslie, M. Baguelin, S. Bhatia, A. Boonyasiri, Z. Cucunubá, G. Cuomo-Dannenburg, et al. Report 9: Impact of non-pharmaceutical interventions (npis) to reduce covid19 mortality and healthcare demand. *Imperial College London*, 10(77482):491–497, 2020.

[20] C. Fraser, D. A. Cummings, D. Klinkenberg, D. S. Burke, and N. M. Ferguson. Influenza transmission in households during the 1918 pandemic. *American journal of epidemiology*, 174 (5):505–514, 2011.

[21] E. Frias-Martinez, G. Williamson, and V. Frias-Martinez. An agent-based model of epidemic spread using human mobility and social network information. In *2011 IEEE third international conference on privacy, security, risk and trust and 2011 IEEE third international conference on social computing*, pages 57–64. IEEE, 2011.

[22] G. F. Gause. Experimental analysis of vito volterra's mathematical theory of the struggle for existence. *Science*, 79(2036):16–17, 1934.

[23] M. E. Gilpin. Spiral chaos in a predator-prey model. *The American Naturalist*, 113(2):306–308, 1979.

[24] L. R. Ginzburg, O. Burger, and J. Damuth. The may threshold and life-history allometry. *Biology Letters*, 6(6):850–853, 2010. ISSN 1744-9561.

[25] A. E. Goodwin, J. E. Peterson, T. R. Meyers, and D. J. Money. Transmission of exotic fish viruses: the relative risks of wild and cultured bait. *Fisheries*, 29(5):19–23, 2004.

[26] A. Hastings, C. L. Hom, S. Ellner, P. Turchin, and H. C. J. Godfray. Chaos in ecology: is mother nature a strange attractor? *Annual review of ecology and systematics*, 24(1):1–33, 1993.

[27] H. W. Hethcote, W. Wang, L. Han, and Z. Ma. A predator–prey model with infected prey. *Theoretical population biology*, 66(3):259–268, 2004.

[28] Y.-H. Hsieh and C.-K. Hsiao. Predator–prey model with disease infection in both populations. *Mathematical medicine and biology: a journal of the IMA*, 25(3):247–266, 2008.

[29] W. O. Kermack and A. G. McKendrick. A contribution to the mathematical theory of epidemics. *Proceedings of the royal society of london. Series A, Containing papers of a mathematical and physical character*, 115(772):700–721, 1927.

[30] M. Kidd, A. Richter, A. Best, N. Cumley, J. Mirza, B. Percival, M. Mayhew, O. Megram, F. Ashford, T. White, et al. S-variant sars-cov-2 lineage b1. 1.7 is associated with significantly higher viral load in samples tested by taqpath polymerase chain reaction. *The Journal of infectious diseases*, 223(10):1666–1670, 2021.

[31] A. Klebanoff and A. Hastings. Chaos in one-predator, two-prey models: cgeneral results from bifurcation theory. *Mathematical biosciences*, 122(2):221–233, 1994.

[32] P. Klepac, S. Kissler, and J. Gog. Contagion! the bbc four pandemic–the model behind the documentary. *Epidemics*, 24:49–59, 2018.

[33] K. D. Lafferty. The evolution of trophic transmission. *Parasitology Today*, 15(3):111–115, 1999.

[34] K. D. Lafferty, G. DeLeo, C. J. Briggs, A. P. Dobson, T. Gross, and A. M. Kuris. A general consumer-resource population model. *Science*, 349(6250):854–857, 2015.

[35] D. B. Larremore, B. Wilder, E. Lester, S. Shehata, J. M. Burke, J. A. Hay, M. Tambe, M. J. Mina, and R. Parker. Test sensitivity is secondary to frequency and turnaround time for covid-19 screening. *Science advances*, 7(1):eabd5393, 2021.

[36] Q. Li, X. Guan, P. Wu, X. Wang, L. Zhou, Y. Tong, R. Ren, K. S. Leung, E. H. Lau, J. Y. Wong, et al. Early transmission dynamics in wuhan, china, of novel coronavirus–infected pneumonia. *New England journal of medicine*, 2020.

[37] R. Li, C. J. E. Metcalf, N. C. Stenseth, and O. N. Bjørnstad. A general model for the demographic signatures of the transition from pandemic emergence to endemicity. *Science Advances*, 7(33):eabf9040, 2021.

[38] R. L. Lindeman. The trophic-dynamic aspect of ecology. *Ecology*, 23(4):399–417, 1942.

[39] J. O. Lloyd-Smith, S. J. Schreiber, P. E. Kopp, and W. M. Getz. Superspreading and the effect of individual variation on disease emergence. *Nature*, 438(7066):355–359, 2005.

[40] A. J. Lotka. Analytical note on certain rhythmic relations in organic systems. *Proceedings of the National Academy of Sciences*, 6(7):410–415, 1920.

[41] S. Maslov and K. Sneppen. Severe population collapses and species extinctions in multihost epidemic dynamics. *Physical Review E*, 96(2):022412, 2017.

[42] R. M. May. Simple mathematical models with very complicated dynamics. *Nature*, 261(5560): 459–467, 1976.

[43] A. Mouinga-Ondémé, M. Caron, D. Nkoghé, P. Telfer, P. Marx, A. Saïb, E. Leroy, J.-P. Gonzalez, A. Gessain, and M. Kazanji. Cross-species transmission of simian foamy virus to humans in rural gabon, central africa. *Journal of virology*, 86(2):1255–1260, 2012.

[44] W. W. Murdoch, C. J. Briggs, and R. M. Nisbet. *Consumer-resource dynamics (MPB-36)*. Princeton University Press, 2013.

[45] D. Musso, A. J. Rodriguez-Morales, J. E. Levi, V.-M. Cao-Lormeau, and D. J. Gubler. Unexpected outbreaks of arbovirus infections: lessons learned from the pacific and tropical america. *The Lancet Infectious diseases*, 18(11):e355–e361, 2018.

[46] M. Newman and B. W. Roberts. Mass extinction: Evolution and the effects of external influences on unfit species. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 260 (1357):31–37, 1995.

[47] M. E. Newman. A model of mass extinction. *Journal of theoretical Biology*, 189(3):235–252, 1997.

[48] B. F. Nielsen, A. Eilersen, L. Simonsen, and K. Sneppen. Lockdowns exert selection pressure on overdispersion of sars-cov-2 variants. *medRxiv*, 2021.

[49] C. Packer, R. D. Holt, P. J. Hudson, K. D. Lafferty, and A. P. Dobson. Keeping the herds healthy and alert: implications of predator control for infectious disease. *Ecology Letters*, 6(9): 797–802, 2003.

[50] L. Pappalardo, S. Rinzivillo, Z. Qu, D. Pedreschi, and F. Giannotti. Understanding the patterns of car travel. *The European Physical Journal Special Topics*, 215(1):61–73, 2013.

[51] L. Perez and S. Dragicevic. An agent-based approach for modeling dynamics of contagious disease spread. *International journal of health geographics*, 8(1):1–17, 2009.

[52] R. H. Peters. *The ecological implications of body size*, volume 2. Cambridge university press, 1986.

[53] R. Peterson, R. Page, and K. Dodge. Wolves, moose, and the allometry of population cycles. *Science*, 224(4655):1350–1352, 1984.

[54] J. Reed and N. C. Stenseth. On evolutionarily stable strategies. *Journal of theoretical biology*, 108(4):491–508, 1984.

[55] M. L. Rosenzweig and R. H. MacArthur. Graphical representation and stability conditions of predator-prey interactions. *The American Naturalist*, 97(895):209–223, 1963.

[56] J. M. Smith and G. R. Price. The logic of animal conflict. *Nature*, 246(5427):15–18, 1973.

[57] K. Sneppen, P. Bak, H. Flyvbjerg, and M. H. Jensen. Evolution as a self-organized critical phenomenon. *Proceedings of the National Academy of Sciences*, 92(11):5209–5213, 1995.

[58] K. Sneppen, B. F. Nielsen, R. J. Taylor, and L. Simonsen. Overdispersion in covid-19 increases the effectiveness of limiting nonrepetitive contacts for transmission control. *Proceedings of the National Academy of Sciences*, 118(14), 2021.

[59] Statistics Denmark. Fam44n: Families 1. january by municipality, type of family, size of family and number of children. *Statistics Denmark*, 2020. URL https://www.statistikbanken.dk/statbank5a/SelectVarVal/Define.asp?Maintable=FAM44N&PLanguage=1.

[60] N. C. Stenseth, W. Falck, O. N. Bjørnstad, and C. J. Krebs. Population regulation in snowshoe hare and canadian lynx: asymmetric food web configurations between hare and lynx. *Proceedings of the National Academy of Sciences*, 94(10):5147–5152, 1997.

[61] J. Sundell, J. A. Eccard, R. Tiilikainen, and H. Ylönen. Predation rate, prey preference and predator switching: experiments on voles and weasels. *Oikos*, 101(3):615–623, 2003.

[62] Y. Takeuchi and N. Adachi. Existence and bifurcation of stable equilibrium in two-prey, one-predator communities. *Bulletin of mathematical Biology*, 45(6):877–900, 1983.

[63] V. Volterra. Fluctuations in the abundance of a species considered mathematically. *Nature*, 119 (2983):12–13, 1927.

[64] W. E. Wei, Z. Li, C. J. Chiew, S. E. Yong, M. P. Toh, and V. J. Lee. Presymptomatic transmission of sars-cov-2—singapore, january 23–march 16, 2020. *Morbidity and Mortality Weekly Report*, 69(14):411, 2020.

[65] J. S. Weitz and S. A. Levin. Size and scaling of predator–prey dynamics. *Ecology letters*, 9(5): 548–557, 2006.

[66] N. D. Wolfe, C. P. Dunavan, and J. Diamond. Origins of major human infectious diseases. *Nature*, 447(7142):279–283, 2007.

[67] P. Yodzis and S. Innes. Body size and consumer-resource dynamics. *The American Naturalist*, 139(6):1151–1175, 1992.