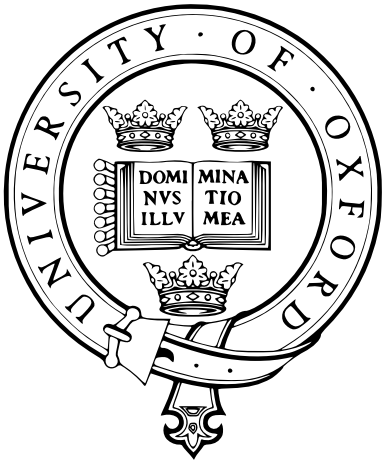
# Metapopulation models of pandemic risk and control

Benjamin John Singer



Brasenose College




A thesis submitted in partial fulfilment
of the requirements for the degree of

*Doctor of Philosophy*



Michaelmas Term 2021

# Metapopulation models of pandemic risk and control

## *Abstract*

Benjamin John Singer ⌬ Brasenose College ⌬ Michaelmas Term 2021

A pandemic is a global outbreak of an infectious disease, which can increase rates of death and disability worldwide, and impact economies and civilisations more generally. Ascertaining the risk of large pandemics with high mortality is challenging, due to their rarity. We can understand pandemics better by using the methods of theoretical epidemiology. In this thesis, I explore what one of these methods—metapopulation models—can tell us about pandemic risk and control. Metapopulation models can represent the spatial spread of pathogens, and thereby capture the global nature of pandemics.

I begin by exploring the effect on pandemic risk of cross-immunity between strains of a pathogen. Cross-immunity occurs when infection and subsequent recovery with one strain confers ongoing protection against another strain. I develop a novel Markov chain method to explore the effects of cross-immunity in a metapopulation model. I find that cross-immunity could help to explain why some pandemics cause more mortality than others, and that its effects are likely to change non-linearly as global travel increases.

Using mathematical models to study pandemic risk naturally leads to questions about how to define a pandemic in quantitative terms. I find that quantitative changes in definition can lead to qualitative changes in the interpretation of model results. I use my novel Markov chain method to explore the effects of different pandemic definitions in interpreting metapopulation models, underlining the importance of precise language when discussing pandemic risk.

Metapopulation models can be used to generate recommendations for pandemic control. I evaluate spatial control strategies that could help prevent or ameliorate pandemics, comparing risk-targeting and centrality-targeting strategies for regional allocation of prophylaxis. Risk-targeting strategies use simulation to allocate prophylaxis to regions at the highest risk of epidemics. Centrality-targeting strategies use properties of the travel network to detect regions that contribute most to a pathogen's circulation. Risk-targeting strategies are more commonly discussed in the literature on spatial pathogen control, but I find that centrality-targeting strategies can reduce infections more effectively when resources for control are limited and the pathogen has the potential to spread widely.

A large quantity of data has been gathered on the spread of SARS-CoV-2 in England during the COVID-19 pandemic. I use these data to parameterise a metapopulation model to evaluate prophylactic drug stockpiling for pandemic preparedness. I find that prophylactic drugs can be particularly effective at containing or slowing an outbreak when used alongside non-pharmaceutical interventions.

# Publications

Some of the work presented in this thesis has also been published in article format, as listed below.

- Singer, B. J., Thompson, R. N. & Bonsall M. B. The effect of the definition of 'pandemic' on quantitative assessments of infectious disease outbreak risk. *Scientific Reports* **11**, 2547 (2021). DOI: 10.1038/s41598-021-81814-3. This publication discusses some of the work presented in chapters 2 and 3.

- Singer, B. J., Thompson, R. N. & Bonsall M. B. Evaluating strategies for spatial allocation of vaccines based on risk and centrality. In review: *Journal of the Royal Society Interface*. Pre-print DOI: 10.1101/2021.09.07.21263209. This publication discusses some of the work presented in chapter 4.

I have contributed to the following additional articles outside the scope of this thesis, while supported by this studentship.

- Ferretti L., Di Nardo A., Singer B. J., Lasecka L., Logan G., Wright C. F., Pérez-Martín E., King D. P., Tuthill T. J. & Ribeca P. Within-Host Recombination in the Foot-and-Mouth Disease Virus Genome. *Viruses* **10**, 221 (2018). DOI: 10.3390/v10050221.

- Singer B. J., Di Nardo A., Hein J. & Ferretti L. Comparing Phylogeographies: Incompatible Geographical Histories in Pathogens' Genomes. Working paper. Pre-print DOI: 10.1101/2020.01.10.902460.

An implementation in Python of the Markov chain model used in chapters 2 and 3 is available at `https://github.com/BenSinger01/Metapopulation-Markov-Chain/`, and the Python code used to generate the figures in chapter 4 is available at `https://github.com/BenSinger01/Evaluating-Strategies/`.

# Acknowledgements

I would like to thank my supervisors Michael Bonsall, for giving me the freedom to pursue the niche topic that was pandemic risk; and Robin Thompson, for continuing to send me back to the drawing board until I figured out how to write a decent paper. I am grateful also to the supervisors of my first rotation project, Luca Ferretti and Jotun Hein, for their enthusiasm and for all the long conversations over a whiteboard, commute, or glass of wine. The past and present members of both the Mathematical Ecology Research Group and the Statistical Genetics Group have all had a hand in this thesis, through their feedback, advice, and second hand textbooks. My college advisor, Elspeth Garman, has been instrumental in guiding me through the entire process, from nurturing my first sparks of interest in a life sciences DPhil, to helping me redraft my writing schedule when projects started getting out of hand.

Many thanks to all the infectious disease modellers who have had a moment to spare at a conference or over an email to offer a bit of advice or a data set, including Ben Althouse, José Lourenço, Roz Eggo, and Hamish Gibbs. If this thesis has any relevance to public policy, it due to many fascinating conversations with Jasdeep Sandhu, Alastair Ager, and Charlotte Watts at the Department for International Development (long may its memory continue). I'd also like to acknowledge the works of science fiction authors Ada Palmer and the late Ursula K. Le Guin, as well as conversations with Greg Lewis and Andrew Snyder-Beattie, for helping me to imagine the world that I hope my work has some small hand in creating.

Credit for keeping me afloat during these four years goes to Leila, for collegial-uxorial succour since our arrival in Oxford; Matteo, for my first lessons in Italian, consciousness, and writing a thesis; Tom Fay, for all his nonsense, gentle and otherwise; Ariane, for listening to my confessions; Eric, for teaching me so much about so many things, and for plying me with delicacies while he does it; Matthew, for his tear-absorbent shoulders and long, hug-giving arms; and Steph, for lockdown walks and for always being in my corner. I couldn't have got through the experience of finishing a DPhil during a pandemic if not for Thomas's support. Thank you for never being further than a phone call away, and for travelling down many roads—and back again!

Thanks above all to my mother, Katherine, who has been listening to me ramble about my strange interests since I learned to speak. None of this could have happened without her encouragement from the very start.

# Contents


**1 Introduction** **1**

1.1 Theoretical epidemiology . . . . . 3

1.2 The SIR model . . . . . 7

1.3 The stochastic SIR model . . . . . 11

1.4 Metapopulation models . . . . . 16

1.5 Pandemics and policy . . . . . 20

1.6 Thesis outline . . . . . 21

**2 Cross-immunity and pandemic risk** **25**

2.1 Cross-immunity and epidemic risk . . . . . 28

2.2 Pandemics on two regions . . . . . 35

2.3 Pandemics on networks . . . . . 41

2.4 Efficiency of the Markov chain . . . . . 54

2.5 Conclusions . . . . . 56

**3 Pandemic definitions** **61**

3.1 Between-region travel . . . . . 64

3.2 Cross-immunity . . . . . 70

3.3 Heterogeneous transmission . . . . . 74

3.4 Conclusions . . . . . 79

**4 Spatial control of pathogens** **85**

4.1 Centrality . . . 89
4.2 A simple metapopulation model . . . 94
4.3 Prophylaxis strategies . . . 97
4.4 Artificial networks . . . 98
4.5 Empirical networks . . . 104
4.6 Non-pharmaceutical interventions . . . 110
4.7 Conclusions . . . 111

**5 Prophylaxis against SARS-CoV-2 in England 117**
5.1 Modelling travel in England . . . 119
5.2 Pathogen spread . . . 124
5.3 Parameterisation . . . 125
5.4 Prophylaxis . . . 127
5.5 Early detection . . . 129
5.6 Late detection . . . 132
5.7 Conclusions . . . 135

**6 Conclusion 141**
6.1 Risk . . . 141
6.2 Control . . . 143
6.3 Spatial models and pandemics . . . 144
6.4 Future work . . . 148

**Appendices 151**

**A Heat maps with a single scale 153**

**B Spatial control in network variations 163**
B.1 Long-distance travel . . . 163
B.2 Joined grids variations . . . 167
B.3 Non-pharmaceutical interventions . . . 173

**C Statistical models of the spread of SARS-CoV-2 in England** **177**

C.1 Modelling weekly cases . . . . . 177

C.2 The Poisson model . . . . . 179

C.3 The negative binomial model . . . . . 182

C.4 Discussion . . . . . 183

**Bibliography** **187**

*I simply wish that, in a matter which so closely concerns the well-being of mankind, no decision shall be made without all the knowledge which a little analysis and calculation can provide.*



Daniel Bernoulli, on smallpox inoculation, 1766

# 1 Introduction

A pandemic is a global outbreak of an infectious disease [1–3]. Infectious agents (pathogens) that spread to all parts of the globe have the opportunity to cause extremely large numbers of infections. If the disease caused by a pandemic pathogen can result in the death of the host, then a pandemic can result in a marked increase in global mortality.

The 1918 strain of pandemic influenza caused around 50 million deaths, and the HIV/AIDS pandemic has caused around 35 million deaths as of late 2021 [4, 5]. These figures are several times the typical annual mortality due to infectious diseases [6, 7]. Pandemics can also have severe effects on human societies beyond increased mortality. Outbreaks of smallpox and other infectious diseases in the Columbian exchange, combined with warfare and colonisation, led to the loss of many indigenous American languages, civilisations, and cultural practices [8–11]. Economies in sub-Saharan Africa have suffered reduced growth due to the HIV/AIDS pandemic [12, 13]. In general, pandemics can result in stigmatisation and economic disadvantage of the affected populations, nations, or communities [14–16].

A pathogen with characteristics that may allow it to spread globally is known as a *pathogen with pandemic potential* (PPP). Only rarely have PPPs spread to truly global scales and caused millions of deaths. This makes it difficult to determine the year-on-year likelihood of such a pandemic, and the relative probability of pandemics of different scales. There are, however, reasons to believe that a future

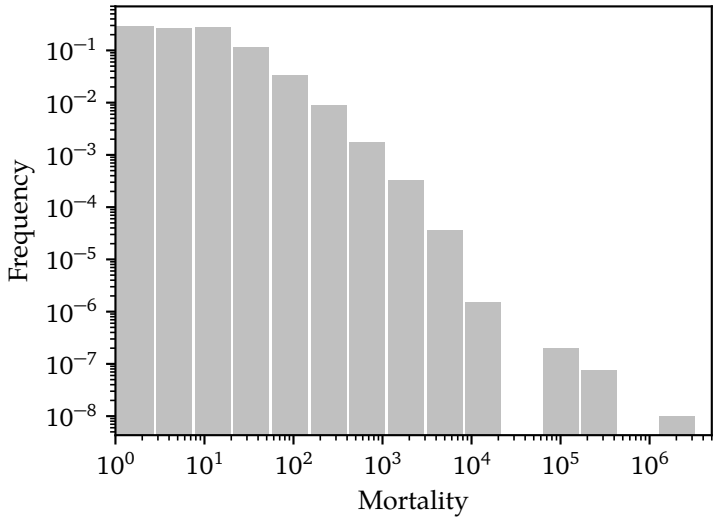



**Figure 1.1** A histogram of the mortality caused by infectious disease outbreaks in the Emergency Events Database [17]. Both the y-axis and the x-axis are on a logarithmic scale. The shape of this histogram suggests that the mortality due to infectious disease outbreaks follows a heavy-tailed distribution.

pandemic could exceed the scale of any past outbreak.

Figure 1.1 shows a histogram of the total mortality recorded from outbreaks of human infectious diseases in the Emergency Events Database, compiled by the Centre for Research on the Epidemiology of Disasters [17]. The plotted portion of the database contains entries from Africa, Europe, and Oceania from 1900 to 2018. These data seem to follow a heavy-tailed distribution, meaning that, although higher-mortality events have lower frequencies, extreme events happen sufficiently often that their contribution to the overall mortality is non-negligible. In particular, the data for high-mortality outbreaks is consistent with a power-law distribution, in which the frequency of outbreaks of size $x$ is proportional to $x^{-\alpha}$, with $\alpha \approx 1.4$. Power-law distributions are thought to describe the mortality from some other kinds of disasters, such as wars and terrorism [18–20]. For such a distribution, the mean mortality from past events is not a good predictor of the mean mortality from future events, due to the large contribution of rare but extremely high-mortality events.

Preparing for rare, extreme events is a difficult task for the global community.

Many years before the onset of the COVID-19 pandemic, epidemiologists warned of the risk of a pandemic of so-called *disease X*, a disease caused by a pathogen so far unknown to infect humans (though this particular terminology only became widespread in 2018) [21–23]. Despite these warnings, a risk assessment from the government of the United Kingdom in 2017 placed an upper bound of 100 on the estimated mortality that could occur in the UK due to an emerging infectious disease [24]. As of late 2021, over 137,000 deaths have been attributed to COVID-19 in the UK [25]. This underestimate, covering three orders of magnitude, illustrates the difficulty of reasoning about events for which there is very little historical precedent. One hundred deaths was a reasonable upper bound on the mortality due to emerging diseases in the UK in recent history, but the nature of infectious disease outbreaks made this a poor estimate for the upper bound on future mortality.

The scientific community faces a great challenge in preparing for severe pandemics that may occur less often than once in a century. Due to the rarity of these events, empirical evidence on their emergence and spread is scarce. Future pandemic pathogens are not guaranteed to resemble past ones, and human activities are constantly reshaping the global dynamics of infectious disease [21–23]. We must therefore rely on methods that allow us to generalise from our knowledge of other kinds of outbreaks, and from the sources of data that we do have on global travel, trade, and agriculture. One source of these methods is theoretical epidemiology, a field that uses mathematics to describe and forecast the dynamics of infectious disease.

## Theoretical epidemiology 1.1

Epidemiology is the study of the occurrence and spread of disease. Theoretical or mathematical epidemiology is a subfield of epidemiology that uses mathematical

and statistical techniques to understand and predict the occurrence of disease, including infectious diseases. It is a highly interdisciplinary field, making use of results from physics [26–28], ecology [29–32], computer science [33–36], applied mathematics [37–41], and statistics [42–44].

An important research activity within theoretical epidemiology is the development and analysis of mathematical models of the spread of pathogens. A mathematical model, in science, consists of a mathematical structure that represents or recreates entities and phenomena of interest [45]. Scientific models allow for a style of inquiry called *surrogative reasoning*, where, instead of examining a system directly, we investigate the properties of a model of that system, and use this investigation to come to conclusions about the system itself. Mathematical models can be used to understand the causes of observed phenomena [46, 47], and to support claims about the potential behaviours of the system of interest [48–50].

Forecasting is a particular use of mathematical models in order to predict the future values of time-dependent variables, such as the number of deaths caused by a disease [51]. An example is the prediction by the Imperial College COVID-19 Response Team in March 2020 that an unmitigated epidemic of COVID-19 in the UK would cause 510,000 deaths by October 2020 [52]. Forecasts that are conditional on a particular circumstance—here, the circumstance of no mitigation being employed—are sometimes known as projections. Both forecasts and projections can be valuable for policy-makers. However, the importance of random, chaotic, or simply unpredictable events in the spread of diseases can make epidemiological forecasting unreliable [53–55]. In these circumstances, epidemiologists favour other kinds of predictions. Another prediction made by the Imperial College team in March 2020 was that school and university closure, case isolation, and general social distancing would prevent the UK National Health Service's surge critical care bed capacity from being overwhelmed. Rather than giving the value of a variable at a particular time, this is a prediction of the effect that one set of

variables will have on another—a different kind of prediction from a forecast, but nonetheless one that can help inform public health policy.

The specific objectives of the modeller will affect which variables are targeted for prediction in a mathematical model. Policy-relevant epidemiological variables can include the cumulative number of infected individuals, the maximum concurrent hospitalisations, or the duration of an epidemic [56]. These variables relate to different public health policy objectives, such as reducing the total morbidity, likelihood of overwhelming health services, or duration of non-pharmaceutical interventions. Further understanding of the effects of an outbreak on society may be gained by coupling mathematical models of infectious disease to econometric models [13, 57].

Some studies of mathematical models are not expected to generate accurate predictions at all. Theoretical epidemiologists often create and examine models primarily to better understand the behaviour of the model itself, rather than the system being modelled. This can produce important results regarding how scientists can make better use of a particular kind of model, including how it can be used to generate predictions, or to understand the causes of observed phenomena. For instance, an article by Kermack and McKendrick in 1927 used a model formulated by McKendrick in 1912 to derive a threshold for how transmissible a pathogen must be in order to cause a large outbreak [38, 58]. The model analysed by the authors is not a model of a specific disease, and the article contains no predictions about whether any particular pathogen will cause an outbreak. However, it provided a theoretical account of the phenomenon known as *herd immunity*, a concept that has proved greatly important to epidemiology and public health policy making [59–63]. I discuss herd immunity further in section 1.2.

Mathematical models consist of parameters and variables that are related through equations. The complexity of a model is typically judged in terms of the number of parameters, variables, and equations needed to describe it. Simpler

models can be interpreted and manipulated more easily, while more complicated models can represent more parts of a system and display a greater range of behaviours. Exactly how complicated a model ought to be depends on the purpose for which it will be put to use.

Some models used for forecasting are very complicated, the hope being that the many parameters of the model will allow insight into many parameters of the system being modelled [64, 65]. However, when a model is used for prediction, the values of its parameters must be estimated from data, and the more parameters a model has, the more (high-quality) data are required for this process to be successful. Therefore, when data are scarce, simple models may be more useful for forecasting than complicated ones.

When a model is formulated to achieve more general scientific insight, it is often sensible to restrict the parameters included in the model to the smallest set that will reproduce the phenomenon of interest. This type of model is sometimes known as a *minimal* model. Using a small set of parameters makes the manipulation of the model easier—both in the derivation of analytical results, and numerical treatments. However, different parameters and phenomena can interact, meaning that a more complicated model can sometimes reveal important effects that cannot be reproduced in simpler models.

In practice, theoretical epidemiologists use a wide range of models with varying complexity [66]. In models of infectious disease, the most essential phenomenon to reproduce is the spread of a pathogen from one individual to another. In the next section, I introduce one of the most basic and ubiquitous models of pathogen spread.

# The SIR model ⬡ 1.2

Suppose that there exists some population of $N$ individuals, some of whom are infected by a pathogen. We denote with $I$, for infective, the number of individuals who are infected by the pathogen and may pass it on to others; and denote with $S$, for susceptible, the number of individuals who have not yet been infected (but may become so). We assume that individuals who are infected either die, or recover from the infection and become immune to it. In either case, once an individual is no longer infective, they count towards the number $R$, for removed.

To model how the pathogen spreads through the population, we must account for how it transmits between individuals. Suppose that each infective individual encounters susceptible individuals at a rate $\kappa$, and that a proportion $p$ of these encounters result in the transmission of the pathogen. We assume that the population is *panmictic*, meaning that each individual is equally likely to encounter each other individual. In this case, the principle of mass action applies, and transmission events occur at the rate $p\kappa SI$. We can write this as $(\beta/N)SI$, where $\beta = Np\kappa$ is the *transmission rate*, with dimensions of inverse time. Once an individual is infected, we assume that they recover at a rate $\mu_\mathrm{r}$, or die at a rate $\mu_\mathrm{d}$. The combined rate of removal from the infective population is therefore $\mu = \mu_\mathrm{r} + \mu_\mathrm{d}$.

If the infectiousness and recovery rate of an infected individual do not change over the course of their infection, then the assumptions given above are described by the following set of differential equations:

$$\frac{\mathrm{d}S}{\mathrm{d}t} = -\frac{\beta}{N}SI \tag{1.1}$$

$$\frac{\mathrm{d}I}{\mathrm{d}t} = \frac{\beta}{N}SI - \mu I \tag{1.2}$$

$$\frac{\mathrm{d}R}{\mathrm{d}t} = \mu I, \tag{1.3}$$

where $S$, $I$, and $R$ are functions of the time $t$. Here we assume that the population

size is very large, so that $S$, $I$, and $R$ can be approximated as continuous rather than discrete quantities. This set of equations, relating the parameters $\beta$, $N$, and $\mu$, and the variables $S$, $I$, $R$, and $t$, make up the *susceptible-infective-removed* (SIR) model of pathogen spread. This model was developed in the early 20$^{\text{th}}$ century, and is one of the simplest possible models of disease spread [38, 58]. Note that the variable $R$ is redundant, since $N$ is constant and $R = N - S - I$, so the entire model can be described using only three parameters and two time-dependent variables, related by two equations.

Within this formalism, it is possible to derive a number of important results concerning the spread of a pathogen. First, we can consider the rate at which an emerging outbreak grows in its early stages, when the number of infective individuals is small. When $I$ and $R$ are small relative to $N$, then $S \approx N$. Then $(\beta/N)SI$ simplifies to $\beta I$, and equation 1.2 can be integrated to give $I$ as a function of time:

$$I(t) \approx I(0)e^{(\beta-\mu)t}. \tag{1.4}$$

This is an equation of exponential growth (if $\beta > \mu$) or decay (if $\beta < \mu$). For this reason, the early stage of an outbreak is sometimes called the exponential phase. The growth rate of an outbreak in the exponential phase is given by $\beta - \mu$.

The insight from this simple model, that the early growth of epidemics is approximately exponential, has important implications for outbreak response and control. A small outbreak that grows exponentially can become very large in a small amount of time, compared to one that grows linearly [67, 68]. Because of this, an important real-time metric of success for epidemic control in the early stages is the *growth rate*, with the number of infected individuals at a given point in time being of secondary importance [69, 70].

If $\beta > \mu$, then eventually, as more susceptibles become infected (and subsequently recover or die), the approximation $S \approx N$ ceases to be valid. At this point, the growth of the epidemic slows. Suppose we wish to determine whether the

epidemic is growing or shrinking at a particular time. The sign of $dI/dt$ is given by the sign of $(\beta/N)S-\mu$, since $I$ cannot be negative. So if $(\beta/N)S(t')-\mu>0$, then the epidemic is growing at time $t'$, and if $(\beta/N)S(t')-\mu<0$, then the epidemic is shrinking at time $t'$. This is the basis for the definition of the time-dependent reproduction number $R_t$:

$$R_t(t)=\frac{\beta}{\mu}\frac{S(t)}{N}. \tag{1.5}$$

When $R_t>1$ the epidemic is growing, and when $R_t<1$ the epidemic is shrinking. $R_t$ is a dimensionless quantity—it is not a rate, and so it cannot tell us how *quickly* the epidemic is growing or shrinking. $R_t$ can, however, be interpreted as the number of susceptible individuals to whom a single infective individual transmits the pathogen over the course of their infection, assuming constant $S(t)$. The infectious period lasts for an expected time $1/\mu$, during which the infective individual transmits the pathogen at a rate $(\beta/N)S$. Multiplying these two quantities yields the right hand side of equation 1.5. The conventional notation $R_t$ should not be confused with that for the number of removed individuals, $R$.

When $S(t)<(\mu/\beta)N$, enough individuals have been infected that $R_t<1$ and the epidemic begins to shrink. At this point the population is said to have acquired *herd immunity*, meaning that enough individuals are immune to re-infection that the number of infective individuals $I$ can no longer increase [62]. This concept can be extended to immunity acquired by other means, such as vaccines, to determine how many individuals must be immunised in order to prevent a pathogen from causing new outbreaks [59, 71–73]. In this case, it is important to know the *herd immunity threshold* $p_{\mathrm{H}}$, the proportion $p_{\mathrm{H}}=1-\mu/\beta$ of individuals who must be immunised in order to achieve $S<(\mu/\beta)N$.

We can investigate the behaviour of the SIR model when $t$ is large by using the chain rule and equations 1.1 and 1.3:

$$\frac{\mathrm{d}S}{\mathrm{d}R}=\frac{\mathrm{d}S}{\mathrm{d}t}\Big/\frac{\mathrm{d}R}{\mathrm{d}t}=-\frac{\beta}{\mu}\frac{S}{N}. \tag{1.6}$$

This differential equation can be solved with separation of variables, yielding:

$$S(t) = S(0)\exp\left(-\frac{\beta}{\mu}\frac{R(t)-R(0)}{N}\right). \tag{1.7}$$

The epidemic ends when $I = 0$, so in the long term $S(\infty) = N - R(\infty)$. This allows us to write the so-called *final size equation*:

$$R(\infty) = N - S(0)\exp\left(-\frac{\beta}{\mu}\frac{R(\infty)-R(0)}{N}\right). \tag{1.8}$$

$R(\infty)$ is the final size—the total number of individuals who become infected at any point during the epidemic. A proportion $\mu_d/\mu$ of the removed individuals die, so the total mortality $M$ of the epidemic is proportional to the final size: $M = (\mu_d/\mu)R(\infty)$.

Note that equation 1.8 cannot be solved analytically for $R(\infty)$, which must be found numerically. Importantly, however, we can show that $R(\infty) = N$ is *not* a solution to equation 1.8 when $R(0) < N$, since $N \neq N - S(0)\exp(-(\beta/\mu)\frac{R(\infty)-R(0)}{N})$ for all $S(0) \neq 0$. This means that no epidemic can infect every individual in the population—the population of infective individuals runs out before the population of susceptible individuals.

It is commonly assumed that $R(0) = 0$, and that the initial infective population $I(0)$ is very small, meaning that $S(0) \approx N$. This leads to the definition of $R_0 \approx R_t(0)$:

$$R_0 = \frac{\beta}{\mu}. \tag{1.9}$$

$R_0$ is known as the *basic reproduction number*, or the reproduction number in a susceptible population. When $R_0 > 1$, a pathogen introduced to a completely susceptible population will cause an epidemic, and when $R_0 \leq 1$ it will not. Many outcomes of the SIR model can be expressed in terms of $R_0$, for instance, the herd immunity threshold $p_H$ can be written in terms of $R_0$: $p_H = 1 - 1/R_0$. The final size

equation can also be written in terms of $R_0$. $R_0$ can be estimated from epidemiological data, and models allow us to connect this estimation to a range of possible outcomes and behaviours of an epidemic [49, 74–76].

The very simple SIR model of disease spread has yielded some highly influential results concerning epidemics of infectious diseases. Due to its simplicity, the model can be extended in a number of different ways, and many of the results of the basic model can be generalised to these extensions. The version of the SIR model presented in this section is sometimes known more specifically as the deterministic, panmictic SIR model. In the next section, I discuss the stochastic version of this model.

## The stochastic SIR model 1.3

Consider an entirely susceptible population, into which we introduce a single infective individual. If $\beta > \mu$, the deterministic SIR model predicts that an epidemic will inevitably occur, reaching the final size given in equation 1.8. This seems plausible when the initial population of infective individuals is very large. However, when only one individual is infective, it seems that there should be some chance for them to recover before they get a chance to pass the pathogen on, preventing an epidemic from occurring. To incorporate this idea of chance into the SIR model, and therefore find the probability that a single infective individual will cause an epidemic, we will need to introduce an element of randomness, otherwise known as *stochasticity*.

In the stochastic SIR model, we interpret $S$, $I$, and $R$ as discrete quantities—i.e. as natural numbers. Instead of infective individuals encountering susceptible individuals at a deterministic rate $\kappa$, we interpret this value as a stochastic rate, giving the expected value of the number of contacts per individual per unit time. For each contact between an infectious and a susceptible individual, we assume

that there is a probability $p$ for the susceptible individual to become infected. Assuming again that the population is panmictic, the rate at which susceptibles become infected is $(\beta/N)SI$. Infected individuals recover at a stochastic rate $\mu$. So, in a very small time $\delta t$, one of the following events can occur:

- transmission, where $S \to S-1$ and $I \to I+1$, with probability $(\beta/N)SI\,\delta t$,
- recovery, where $I \to I-1$ and $R \to R+1$, with probability $\mu I\,\delta t$,
- no event, with probability $1-((\beta/N)SI+\mu I)\,\delta t$.

The probability $P_{\mathrm{I}}$ that the next event will be an infection is:

$$P_{\mathrm{I}} = \frac{\frac{\beta}{N}S}{\frac{\beta}{N}S+\mu}, \tag{1.10}$$

and the probability $P_{\mathrm{R}}$ that it will be a recovery is:

$$P_{\mathrm{R}} = \frac{\mu}{\frac{\beta}{N}S+\mu}. \tag{1.11}$$

Note that $P_{\mathrm{R}} = 1-P_{\mathrm{I}}$, since these are the only two events that can occur. Since the rates of these events are both directly proportional to the number of infective individuals $I$, the probability of one happening over the other is not affected by $I$. Instead, $I$ affects the total rate $\zeta$ at which events happen: $\zeta = (\beta/N)SI+\mu I$.

A stochastic SIR outbreak can be efficiently simulated using a method called the *Gillespie direct algorithm* [77]. The algorithm is initialised with $t=0$, and given values of $S(0)$, $I(0)$, $\beta$, and $\mu$. At each step of the algorithm, two random numbers, $0<r_1<1$ and $0<r_2<1$, are generated. The time $t$ is incremented by $(1/\zeta)\ln(1/r_1)$, which has an exponential distribution with mean $1/\zeta$. What happens to $S$ and $I$ depends on the value of $r_2$.

- If $r_2 \le P_{\mathrm{I}}$, $S$ is decremented by one, and $I$ is incremented by one.
- If $r_2 > P_{\mathrm{I}}$, $I$ is decremented by one.

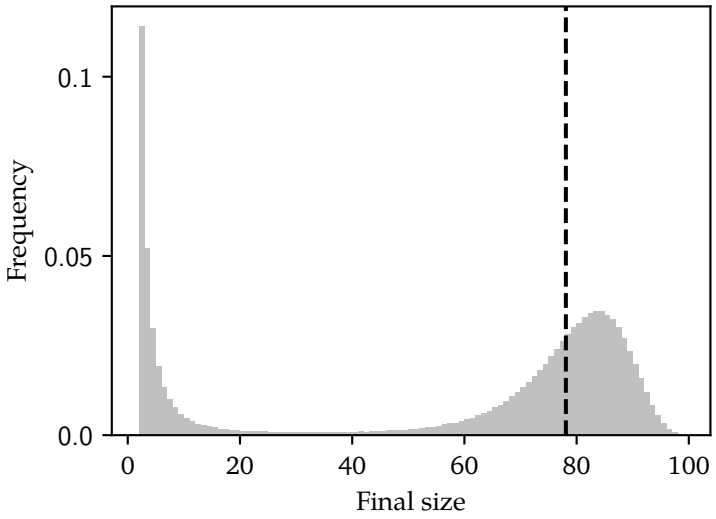



**Figure 1.2** A histogram of the final size of an epidemic in one million iterations of the stochastic SIR model with $I(0) = 2$, $S(0) = 98$, $R(0) = 0$, $\beta = 2$ and $\mu = 1$. The distribution is bimodal, with one quarter of the outbreaks failing to cause a large epidemic, and three quarters of the outbreaks causing an epidemic whose size scales with $S(0)$. The mean size of outbreaks in the latter category is equal to the final size given by equation 1.8, shown by the dashed line.

These steps are then repeated until $I = 0$, and the epidemic ends. The final size is then given by $R(\infty) = N - S$. The probability of a given final size $R(\infty) = x$ can be determined by running the algorithm many times, using a new set of random numbers each time, then counting the proportion of runs with $R(\infty) = x$. The probability of each possible final size of an outbreak, when $I(0) = 2$, $S(0) = 98$, $R(0) = 0$, $\beta = 2$, and $\mu = 1$, is illustrated in figure 1.2.

With the stochastic SIR model specified, we can return to the question of how likely a single infective individual is to cause an epidemic. Define $q$ as the probability that no epidemic will occur. When almost all the population is susceptible, $S \approx N$, then the probability for the single individual to be removed before any transmission occurs is $\mu/(\beta + \mu)$. In this case, an epidemic does not occur. However, if instead the infective individual infects at least one susceptible individual before its removal, which occurs with probability $\beta/(\beta + \mu)$, then there is still a chance for an epidemic not to occur if both infected individuals fail to cause an epidemic. If the transmission from each individual is independent, which

is true when $N$ is large and the individuals are unlikely to encounter the same susceptibles, then this probability is simply equal to $q^2$. So we can write:

$$q = \frac{\mu}{\beta + \mu} + \frac{\beta}{\beta + \mu} q^2. \tag{1.12}$$

The two solutions to this quadratic equation are $q_+ = 1$ and $q_- = \mu/\beta = 1/R_0$. We take the smaller solution [78]. If $R_0 > 1$, then $q = q_- = 1/R_0$, and an epidemic may occur with probability $1 - 1/R_0$. If $R_0 \leq 1$, then $q = q_+ = 1$, and no epidemic occurs. This is an extension of the result from the deterministic SIR model that an epidemic can only occur if $R_0 > 1$. The difference in the stochastic model is that there is still some chance for no epidemic to occur, even if $R_0 > 1$.

The above result can be extended to larger numbers $I(0)$ of initial infective individuals, so long as $I(0) \ll N$ [53, 66, 79–82]. In general, the probability $\mathbb{P}$ that $I(0)$ initial infected individuals will cause an epidemic in an otherwise completely susceptible population is:

$$\mathbb{P} = 1 - \left(\frac{1}{R_0}\right)^{I(0)}, \tag{1.13}$$

so long as $R_0 > 1$. Otherwise, $\mathbb{P} = 0$. The event in which an outbreak fails to occur, despite a value of $R_0$ greater than one, is referred to as *stochastic extinction*.

We can return to figure 1.2 to understand the meaning of equation 1.13. There are two distinct possible outcomes of the introduction of two infected individuals to the population, represented by the two peaks of the distribution. With probability $\mathbb{P} = 1 - (1/2)^2 = 3/4$, a large epidemic occurs, with the final size varying around a mean value that depends on the initial number of susceptibles $S(0)$. With probability $1/4$, no epidemic occurs, and the final size follows a highly skewed distribution with a mean close to $R_0 I(0)$, which is independent of $S(0)$.

This result acts as the foundation for a theoretical definition of the term 'epidemic', which I have so far used somewhat informally. In mathematical terms, an epidemic is an outbreak whose size depends on the number of susceptibles, rather

than on the number of initial infective individuals. This occurs when a pathogen transmits freely through a population, rather than being limited to only a small subset. The distinction between a limited outbreak and an epidemic is particularly clear in the stochastic SIR model, but similar bimodal distributions can be observed in a wide range of stochastic models of infectious disease [41, 56, 82, 83]. I will use the term 'outbreak' to refer to any event in which susceptible individuals become infected by a pathogen, and the term 'epidemic' to refer to an outbreak whose size scales with the number of susceptibles.

Many important results of the deterministic SIR model can be reproduced or extended in the stochastic case. The expected value of the number of infective individuals grows exponentially when $S \approx N$, so long as the outbreak does not experience stochastic extinction [84]. The time-dependent reproduction number can be defined in the stochastic SIR model in the same way as for the deterministic SIR model, and used to determine the herd immunity threshold $p_{\mathrm{H}} = 1 - 1/R_0$. Equation 1.8 can be solved to give the expected value of the final size of an epidemic, in the case that stochastic extinction does not occur [85].

It should be noted that neither form of the SIR model presented in this chapter can sustain an *endemic* pathogen—that is, a pathogen whose prevalence does not eventually go to zero. However, since pandemics are a type of *epidemic*, this behaviour is not necessary for modelling pandemics in the first instance. Of course, pandemic pathogens can *become* endemic, but this is not a scenario that I treat in this thesis.

The stochastic SIR model allows for some unknown variation to occur in the infection process—for instance, individuals might vary randomly in their number of contacts, giving rise to stochastic effects. However, no *systematic* variation can occur if the population is to remain panmictic. In the next section, I will relax the panmictic assumption.

## Metapopulation models 1.4

In the panmictic models described above, all epidemics are global in the sense that the size of any epidemic scales with the size of the whole population. These models are therefore inadequate to address questions about pandemics as a distinct type of epidemic. Models of pandemics must therefore include some way to assign locations to individuals and couple these locations to the transmission dynamics. Such a model has the capacity to simulate outbreaks that remain local, as well as outbreaks that spread globally.

A conceptually straightforward way to represent location is to assign each individual a position in geographical space (i.e. a latitude and a longitude) as a function of time. These locations could be coupled to transmission by classifying any individuals whose locations co-occur within some radius as potentially infectious contacts. However, such a model would involve two time-dependent continuous variables for each individual, in addition to their infection status. For any large population, such a model is unlikely to be analytically or computationally tractable. For this reason, mathematical epidemiologists tend to make simplifying assumptions about the relationship between space and transmission.

One class of spatial models, *agent-based* models, retains the description of the population at the level of individuals, while spatial location is modelled in some simplified way [86–88]. Such simplifications typically involve representing space as a set of distinct locations—such as households, workplaces, and schools—in which the pathogen can be transmitted. Each individual is then assigned to a particular set of locations, and infective individuals are assumed to infect co-located susceptible individuals at some given rate. Location is rarely represented as an explicit function of time—more often the infection rate for an individual is averaged over all locations to which that individual is assigned, weighted by the expected amount of time spent in each location.

Agent-based models can incorporate a large amount of data on commuter patterns, workplace structure, contact rates between different age groups, and other information pertinent to disease spread [89, 90]. When it is unclear exactly which properties of a population will affect important outcomes of an outbreak, it can be important to incorporate a wide range of data streams. Such models are therefore particularly valuable for forecasting, and for producing policy recommendations in response to specific outbreaks [52]. However, such usage relies on the existence of robust data on a great number of parameters of the modelled population, which cannot always be relied upon when reacting in real time.

Another challenge of agent-based models is that, despite the simplified representation of space, the number of variables is still very large. This makes simulation computationally intensive, and pen-and-paper analysis unfeasible [91, 92]. If we wish to reach general conclusions about the possible behaviours of future pandemics, we may need to rely on further simplified models of disease spread.

Spatial metapopulation models are a kind of simple spatial model, first developed in ecology, which are comparatively mathematically and computationally tractable [30, 93–99]. A metapopulation model is any model that splits a population up into smaller groups, with each individual being a member of exactly one of these groups at a given time. Given certain assumptions about the dynamics within each group, the state of the model can be described with a set of variables giving the number of individuals in each state in each group. If the possible states of an individual are susceptible, infective, and removed, as in the models above, and the number of groups is $n$, then the total number of time-dependent variables in the model is $3n$. This can be a great simplification compared to agent-based models, especially when there are few groups.

Suppose that a portion of geographic space can be sub-divided into $n$ regions. Then we can create a metapopulation model in which each member of the population is assigned to a group corresponding to the region in which they are located.

Such regions could be arbitrary divisions of space, but it is usually more useful to match them to existing geographic entities, such as cities, counties, states, or any of the other social-spatial structures into which humans organise themselves. For agricultural pathogens, the regions can instead correspond to farms.

With 'region' suitably defined, then, to achieve the simplification outlined above, we must assume that all individuals assigned to the same region are panmictic amongst themselves (but not with individuals in other regions). The movement of individuals from one region to another can then be summarised in terms of the overall migration rates between pairs of regions. Alternatively, we can assume that individuals' region memberships are fixed, but that some level of infectious contact occurs between individuals in different regions. In either framework there is a rate, $\lambda_{ij}$, that characterises either the migration rate or the contact rate between a given region $i$ and a given region $j$. We generally assume that this rate is low relative to the contact rates *within* regions, since infectious contacts mainly occur locally [100].

The set of the values $\lambda_{ij}$ forms an $n \times n$ matrix, $\Lambda$. This matrix can be interpreted as the adjacency matrix of a network. A network is a mathematical object that represents the connections (called *edges*) between some set of entities (called *vertices*) [101]. In the case of a metapopulation model, the vertices represent regions, and the edges represent the rates of migration or other contact between the regions. A vertex is written simply as its index, e.g. $i$, and an edge is written as an ordered set of the vertices which it connects, e.g. $\{i, j\}$. In the network component of a metapopulation model, $\lambda_{ij}$ is the *weight* of the edge $\{i, j\}$. So, if 1% of the individuals in region 1 travel to region 2 per unit time, then the weight of the edge $\{1, 2\}$ is equal to $\lambda_{12} = 0.01$. In a *symmetric* network, $\lambda_{ij} = \lambda_{ji}$. A symmetric network can be *undirected*, meaning that its edges are unordered sets rather than ordered sets. A network is said to be *weighted* when $\lambda_{ij}$ can have values other than 0 and 1.

The parameters of a spatial SIR metapopulation model are: the transmission

rate $\beta_i$ in each region $i$, the recovery rate $\mu_i$ in each region, and the migration or contact rate $\lambda_{ij}$ between each pair of regions. The variables are: the time $t$, the number of susceptible individuals $S_i(t)$ in each region $i$ as a function of time $t$, the number of infective individuals $I_i(t)$ in each region, and the number of recovered individuals $R_i(t)$ in each region. In a deterministic model with migration, these variables are related by the equations:

$$\frac{\mathrm{d}S_i}{\mathrm{d}t} = -\frac{\beta_i}{N_i} S_i I_i + \sum_j (\lambda_{ji} S_j - \lambda_{ij} S_i) \tag{1.14}$$

$$\frac{\mathrm{d}I_i}{\mathrm{d}t} = \frac{\beta_i}{N_i} S_i I_i - \mu_i I_i + \sum_j (\lambda_{ji} I_j - \lambda_{ij} I_i) \tag{1.15}$$

$$\frac{\mathrm{d}R_i}{\mathrm{d}t} = \mu_i I_i + \sum_j (\lambda_{ji} R_j - \lambda_{ij} R_i), \tag{1.16}$$

with $N_i(t) = S_i(t) + I_i(t) + R_i(t)$. These are similar to equations 1.1–1.3, but with a region subscript added to each quantity, and a sum added to each equation to capture the migration of individuals into and out of each region.

In a stochastic SIR metapopulation model with migration the following events can occur in a very small time $\delta t$:

- transmission in region $i$, where $S_i \to S_i - 1$ and $I_i \to I_i + 1$, with probability $(\beta_i/N_i) S_i I_i \delta t$;
- recovery in region $i$, where $I_i \to I_i - 1$ and $R_i \to R_i + 1$, with probability $\mu_i I_i \delta t$;
- migration of an individual in state $X \in \{S, I, R\}$ from region $i$ to region $j$, with probability $\lambda_{ij} X_i \delta t$;
- no event, with probability $1 - \sum_i \left( (\beta_i/N_i) S_i I_i + \mu_i I_i + \sum_j \lambda_{ij} (S_i + I_i + R_i) \right) \delta t$.

In the case that individuals do not migrate between regions, but $\lambda_{ij}$ rather represents the rate of contact between regions, the total rate of infection events in a given region changes from $(\beta_i/N_i) S_i I_i$ to $(\beta_i/N_i) S_i \left( I_i + \sum_j \lambda_{ji} I_j \right)$, and the terms representing migration are removed. When $\lambda_{ij}$ is symmetric and small compared

to the other parameters, there is not a great difference in behaviour between these two types of models—the important thing is that $\lambda_{ij}$ represents the extent to which the pathogen dynamics in region $i$ and region $j$ are coupled, however this coupling occurs.

Metapopulation models introduce spatial structure into the dynamics of pathogen spread. This allows us to distinguish between local and global outbreaks, and therefore to model pandemics as a distinct class of outbreak [102, 103]. However, challenges remain in the use and interpretation of metapopulation models to understand the future risk of pandemics, and our ability to control them. In the following chapters, I present the results of my research into the properties of metapopulations models and their use as models of pandemics.

## Pandemics and policy 1.5

The potentially severe impact of pandemics on the health of populations worldwide has inspired various public health bodies to formulate specific policies regarding the global spread of an infectious disease, including the health departments of national governments and international organisations such as the World Health Organisation (WHO). For instance, many national governments have pandemic plans for a public health response in the case of the emergence of a new strain of pandemic influenza [104–106]. These plans incorporate guidance from the WHO, which helps to co-ordinate pandemic response between different national governments [107–109].

The WHO is empowered to make recommendations for responding to a PPP by the International Health Regulations [110]. These regulations make reference to the concept of a public health emergency of international concern (PHEIC), a concept which overlaps partially with that of a PPP. In order for an event involving the transmission of a pathogen to be investigated by the WHO as a potential PHEIC, it

must meet at least two of the following four conditions: the public health impact of the event is serious, the event is unusual or unexpected, there is a significant risk of international spread, and/or there is a significant risk of international trade or travel restrictions. The declaration of a PHEIC does not necessarily mean that there is a risk of a pandemic—a pathogen can be of international concern without having the potential to spread globally.

The WHO has at times had official policies for responding to pandemic risks in particular, depending on the year and on the pathogen concerned. For instance, the 2009 WHO pandemic influenza guidelines included criteria for the declaration of the pandemic phase of an influenza outbreak, "indicating that a global pandemic is under way" [108, 111]. As of late 2021, however, the WHO no longer uses the term 'pandemic' as an official status of any outbreak [111, 112]. With policy responses to the COVID-19 pandemic still in active consideration, it is possible that this will change again, with many public health workers and organisations calling for increased pandemic preparedness of the WHO and national governments [113, 114]. Any new policies aiming to increase pandemic preparedness will require a well-informed community of epidemiologists to help policy-makers understand which of the possible responses to a PPP are likely to be effective and proportionate. In the following chapters, I aim to contribute to the understanding of pandemics from a general perspective, and to generate recommendations for future work on pandemic preparedness.

## Thesis outline ⬡ 1.6

This thesis can be divided into two parts. The first, composed of chapters 2 and 3, explores the use of metapopulation models to understand the effects of various parameters on pandemic risk. The second, composed of chapters 4 and 5, explores the use of metapopulation models to evaluate strategies for the control

of a pandemic pathogen.

As noted above, the historical range of mortality caused by various pandemic pathogens is very wide. Understanding this range requires understanding how the various differences between pandemic pathogens contributes to the difference in the outcomes of the pandemics they cause. In chapter 2, I investigate one potential driver of these differences—the cross-immunity between a pandemic pathogen and previously circulating strains in the same pathogen family. I replicate results concerning cross-immunity and pandemic risk from a 2019 article by Thompson et al. [115], before extending these results using a novel Markov chain approach [116]. I show that cross-immunity can contribute to the variation in pandemic risk and size in a metapopulation model, but that its contribution is complex and can depend on other parameters such as the rate of long-distance travel. I demonstrate that, despite their relative simplicity, metapopulation models can capture complex non-linear effects of key epidemiological parameters.

The definition of 'pandemic' that I have used up to this point is that of a global outbreak. In chapter 3, I discuss the insufficiency of this definition for theoretical studies of pandemic risk, and explore potential quantitative definitions that resolve its ambiguity. I show for the first time that a quantitative change in the pandemic definition can result in a qualitative change in the results of a metapopulation model of pandemic risk—an effect that has not been taken into account in existing studies of pandemic risk [116]. I use this finding as a basis on which to make recommendations regarding the use of pandemic definitions in mathematical models.

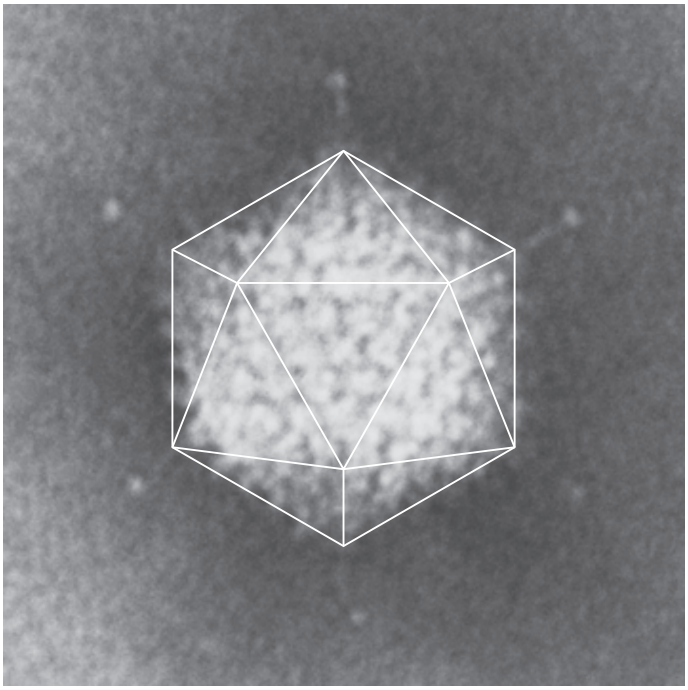
Metapopulation models are used not only to estimate the risk of pandemics, but in order to generate recommendations for their control. There exist many studies using metapopulation models to predict the regions that are at most risk of outbreaks of a pathogen, which recommend that control measures be preferentially targeted to these regions [68, 72, 86, 117]. In chapter 4, I examine whether this

control strategy is more successful at preventing pathogen spread in a metapopulation model than alternatives that target control based on the *centrality* of a region in the travel network. I perform the first direct comparison between predictive risk-targeting and centrality-targeting, and find that various model parameters can affect which control strategy is most effective at reducing infections.

As discussed above, the rarity of large pandemics makes them difficult to prepare for. One way to reduce the risk from viral pandemics is to create stockpiles of antiviral drugs [86, 95, 104–106]. In chapter 5, I examine whether a stockpile of a prophylactic drug could be used to prevent or slow the invasion of a pandemic pathogen in England, making use of data collected on pathogen spread during the COVID-19 pandemic. I use a metapopulation model parameterised using travel data collected by Facebook Data for Good to simulate outbreaks under a range of control strategies [118, 119]. I find that an increased stockpile of prophylactic drugs can have a dramatic effect on the probability of successful containment of an outbreak, and on the timing of an epidemic.

In chapter 6, I reflect on the implications of the work presented in this thesis for the use of metapopulation models to understand pandemic risk and control. I discuss the advantages and disadvantages of using metapopulation models for these purposes, and make recommendations for future research in this area.





An electron micrograph of an adenovirus capsid, reproduced with permission from Springer, with white lines highlighting the symmetrical shape of the protein shell [120]. This shape is the regular icosahedron, a 20-sided polyhedron. It is the Platonic solid with the largest number of faces. Many viruses have icosahedral capsids, and this structure can be derived from energy minimisation of the capsid's protein subunits [121].

# 2 Cross-immunity and pandemic risk

The outcomes of pandemics over the course of history have been highly heterogeneous. In particular, the mortality due to various pandemics has differed greatly. The deaths due to influenza pandemics alone have varied over two or three orders of magnitude—from roughly 200,000 during the 2009 pandemic, to 1 million during the 1957 pandemic, to over 50 million during the 1918 pandemic [4, 122, 123]. The global mortality due to COVID-19 as of late 2021 is estimated to lie on the medium-high end of this spectrum, at 4 million, while HIV/AIDS joins the 1918 strain of influenza A with a death toll in the tens of millions, albeit spread out over a much longer period [5, 124]. The variation is even greater if we consider those outbreaks whose pandemic potential fails to be realised, perhaps due to successful containment, or random chance. If we include such outbreaks which experience stochastic extinction, with only a handful of individuals being infected, then the range of pandemic mortality covers six to seven orders of magnitude.

The global factors that drive the dynamics of pandemics are changing quickly in the 20$^{th}$ and 21$^{st}$ centuries, among them the rate of long-distance travel [125]. Over the last few decades, there has been a consistent trend towards more individuals travelling more quickly over long distances [126]. One driver of this growth in mobility is the tourism industry, with roughly a fifth of the entire world population making at least one international journey in 2018 [127]. Long-distance travel is thought to be an important driver of pandemics, and while humans have been

travelling over long distances since at least the first migrations out of Africa, the current regime of international travel means that it is now possible to move any distance across the world within the incubation period (the time from infection to symptom onset) of almost any infectious disease [128–130]. Understanding the effects of these changes on pandemic risk will require understanding the existing drivers of the variability in pandemic outcomes.

There are many possible sources of this variation. Pandemic pathogens have innate variation in transmissibility and virulence, depending, among other things, on their mechanisms of infection [131, 132]. Climate can influence pathogen emergence, and anthropogenic climate change is therefore a factor in pandemic risk [133, 134]. Human activities can also impact the outcome of a pandemic more directly. Conflict can exacerbate the spread and severity of a disease, and differing approaches to public health policy and heterogeneous access to pharmaceuticals can also change the course of a pandemic [122, 135–138]. Finally, interactions with other pathogens can influence the progress of a pandemic, for better or worse. Co-infections with endemic bacteria have been known to intensify the symptoms of pandemic influenza, and exposure to the 1889 strain of pandemic influenza may have caused dysregulated immune responses to the 1918 strain amongst young adults [135, 139, 140]. Cross-immunity is the other side of this coin, where exposure of the population to previous strains of a pathogen can decrease the transmission rate and virulence of a widely circulating pandemic strain [31, 141, 142].

The effective functioning of the adaptive immune system requires the recognition of specific parts of a pathogen called *antigens* [143]. When the body encounters an antigen that it recognises as foreign, it mounts an immune response. Immune responses are highly complex, and can vary from pathogen to pathogen and host to host. In general, they have two main effects—to reduce the likelihood that a pathogen causes an infection in the host (*infection-reducing* immunity), and/or to reduce the severity of disease caused by that infection, possibly reducing the

likelihood of onward transmission (*disease-reducing* immunity). While infection-reducing immunity can sometimes be complete, effectively reducing the probability of re-infection to zero, it can also be partial, merely reducing the probability of re-infection.

One reason that immunity may be partial is if a pathogen is similar, but not identical, to a previously circulating pathogen to which the body has developed immunity [144, 145]. Such a pathogen may be a member of a different but related strain of microorganism to the original infection. In this case, the immune system may be able to recognise parts of the pathogen, but less effectively or reliably than it would recognise the original strain, leading to partial immunity. This kind of partial immunity between strains of a pathogen is termed *cross-immunity*.

Cross-immunity is well documented in certain pathogens. It is known to drive the dynamics of paramyxovirus infections in the United States and of cholera in Bangladesh, among many other examples [46, 146, 147]. Cross-immunity is likely also to play a role in the severity of pandemic influenza [148–150]. The 2009 pandemic of influenza A caused lower mortality than might reasonably have been expected based on previous influenza pandemics [122]. In particular, the elderly experienced less severe outcomes than typical for pandemic influenza. This may well have been due to cross-immunity acquired from previous strains of influenza A, most likely of swine origin [151, 152]. Cross-immunity is also thought to have played a role in the age-structure of mortality in the 1918 influenza A pandemic, where older adults had unusually low mortality levels relative to younger adults [153].

In 2019, Robin Thompson and his collaborators released a paper exploring the effects of cross-immunity and long-distance travel on pandemic risk, using a metapopulation model with two sub-populations [49]. In this paper, the authors show that cross-immunity between strains of a pathogen can greatly reduce the size and probability of pandemics of that pathogen. In particular, they show

that, when cross-immunity is sufficiently high, increased long-range travel rates can lead to decreases in the size and probability of pandemics of late-emerging pathogen strains, due to the increased circulation of precursor strains that grant cross-immunity. Therefore, as the rate of long-range travel increases around the globe, there is the potential for differences in cross-immunity to explain some of the variance in pandemic outcomes.

In this chapter I will explain in detail the mathematics behind Thompson et al.'s results, replicating said results with a particular focus on explaining the wide range of pandemic outcomes. Figure 2.1 will treat a similar scenario to figure 2 in Thompson et al. 2019, but with a slightly different model to improve continuity with later sections. Figure 2.2 will correspond to figure 4 in Thompson et al. 2019, but with the pandemic rather than the epidemic probability shown, and the cross-immunity $\alpha$ chosen to illustrate the complexity of the model dynamics. In section 2.3 I will go on to extend Thompson et al.'s original model to treat metapopulations with more than two sub-populations, using a novel Markov chain approach. This will allow me to explore the effects of cross-immunity with variable structures of the underlying travel network. I will conclude with a discussion of this Markov chain framework and its potential further applications.

## Cross-immunity and epidemic risk



2.1

I will start with the simple case of an outbreak in a single panmictic population. By definition, a panmictic population has no geographic structure, so there is no distinction between an epidemic and a pandemic (all outbreaks are equally "global"). In this section, therefore, I explore the effect of cross-immunity on epidemic dynamics, before extending the model to treat pandemics later in the chapter.

To investigate the effects of cross-immunity, we need to consider at least two strains of a pathogen. The population is assumed to be completely susceptible

to the first strain that emerges, which I call $X$. Infection with $X$ and subsequent recovery grants an individual infection-reducing cross-immunity against a second strain, which I call $Y$, reducing their susceptibility to $Y$ by the proportion $\alpha$, where $0 \leq \alpha \leq 1$. When $\alpha = 0$, there is no cross-immunity, and when $\alpha = 1$, infection and recovery from $X$ provides complete protection against infection with $Y$.

I assume that the outbreak of $X$ concludes before the first infection with $Y$, so that there are no individuals infected with $X$ by the time that the outbreak of $Y$ occurs. I assume also that cross-immunity does not wane over time, so that the level of cross-immunity $\alpha$ remains constant between outbreaks. As my focus is on the effects of cross-immunity, I take the outcome of the outbreak of $Y$ to be of greater interest than the outcome of the outbreak of $X$. In practical terms, this could be because the $X$ outbreak is historic, or because $Y$ has higher virulence than $X$.

In order to understand the effect of $X$ on the outbreak of $Y$, we need to know how many individuals become infected with $X$ and recover. When $X$ first arrives in the host population, the probability that there will be an epidemic of $X$ is given by:

$$\mathbb{P}^X = 1 - \left(\frac{1}{R_0^X}\right)^{I^X(0)}, \tag{2.1}$$

where $R_0^X > 1$ is the basic reproduction number for $X$ in this panmictic population, and $I^X(0)$ is the initial number of individuals infected with $X$ [66]. If $X$ does indeed cause an epidemic, its expected final size is given by the following equation:

$$R^X(\infty) = N - N \exp\left(\frac{-R_0^X R^X(\infty)}{N}\right), \tag{2.2}$$

where $R^X(\infty)$ is the final number of individuals who recover from an infection with $X$, and $N$ is the population size [38]. These standard results of the stochastic SIR model were introduced in sections 1.2 and 1.3.

When the $Y$ outbreak occurs, the presence of cross-immune individuals will

reduce the probability of an epidemic and reduce the size of any epidemics that do occur. In effect, if an epidemic of $X$ has occurred, then the initial population of those susceptible to $Y$, $S^Y(0)$, is effectively split into two compartments—one of initial size $S^Y(0)-R^X(\infty)$ comprised of those who have not been infected with $X$ and therefore become infected with $Y$ at rate $\beta^Y I^Y$, and one of initial size $R^X(\infty)$ comprised of those who have recovered from infection with $X$ and become infected with $Y$ at rate $(1-\alpha)\beta^Y I^Y$. Therefore, at the beginning of the outbreak, infections occur at a rate $r_\beta$:

$$r_\beta = \left(S^Y(0)-R^X(\infty)\right)\beta^Y I^Y(0) + R^X(\infty)(1-\alpha)\beta^Y I^Y(0). \tag{2.3}$$

Assuming $I^Y(0)$ is small, $N \approx S^Y(0)$, so this simplifies to:

$$r_\beta = \left(1-\alpha\frac{R^X(\infty)}{N}\right)\beta^Y I^Y(0)S^Y(0). \tag{2.4}$$

Note that we measure time for the quantities with superscript $Y$ differently from those with superscript $X$, treating the zero of each time scale as the time of the first infection with the respective strain.

If there has been no outbreak of $X$, the initial rate at which $Y$ infections occur is simply $\beta^Y I^Y(0)S^Y(0)$. The ratio of the reproduction number in the two scenarios is given by the ratio of the rates of infection, since the recovery rate $\mu^Y$ is unaffected by infection with $X$. Thus we obtain the expression for $R_e^Y$, the effective reproduction number of $Y$ when an outbreak of $X$ has previously occurred:

$$R_e^Y = \left(1-\alpha\frac{R^X(\infty)}{N}\right)R_0^Y, \tag{2.5}$$

where $R_0^Y$ is the basic reproduction number of $Y$. Note that $R_e^Y$ is *not* the time-dependent reproduction number—it gives the reproduction number only for the *start* of the $Y$ outbreak, in the case that an epidemic of $X$ has occurred. The pro-

bability $\mathbb{P}^Y$ of an epidemic of $Y$ occurring is:

$$\begin{aligned}\mathbb{P}^Y &= \left(R_0^X\right)^{-I^X(0)}\left(1-\left(R_0^Y\right)^{-I^Y(0)}\right)\\ &\quad+\left(1-\left(R_0^X\right)^{-I^X(0)}\right)\left(1-\left(R_e^Y\right)^{-I^Y(0)}\right),\end{aligned} \tag{2.6}$$

where the fist term gives the probability of an epidemic of $Y$ without an epidemic of $X$, and the second term gives the probability of an epidemic of both $X$ and $Y$.

The expected size of the epidemic of $Y$ depends on whether an epidemic of $X$ has occurred. If an epidemic of $X$ has not occurred, then the expected size of the $Y$ epidemic is $R^Y_{\text{no}X}(\infty)$, given by:

$$R^Y_{\text{no}X}(\infty) = N - N\exp\left(\frac{-R_0^Y R^Y_{\text{no}X}(\infty)}{N}\right), \tag{2.7}$$

in analogy to equation 2.2. In the case where an epidemic of $X$ *has* occurred the final size $R^Y_X$ is given by:

$$R^Y_X(\infty) = N - N\exp\left(\frac{-R_e^Y R^Y_X(\infty)}{N}\right). \tag{2.8}$$

To find the overall expected number infected by strain $Y$ we multiply $R^Y_{\text{no}X}(\infty)$ and $R^Y_X(\infty)$ by the probability of their respective scenarios, giving:

$$\begin{aligned}\mathbb{E}[R^Y(\infty)] &= \left(R_0^X\right)^{-I^X(0)}\left(1-\left(R_0^Y\right)^{-I^Y(0)}\right)R^Y_{\text{no}X}(\infty)\\ &\quad+\left(1-\left(R_0^X\right)^{-I^X(0)}\right)\left(1-\left(R_e^Y\right)^{-I^Y(0)}\right)R^Y_X(\infty).\end{aligned} \tag{2.9}$$

Using these equations, we can see how cross-immunity affects the risk of an epidemic in a simple example. In figure 2.1, I illustrate the behaviour of equations 2.6–2.9 when $I^X(0) = I^Y(0) = 1$, i.e. each epidemic is seeded by a single infected individual, and $R_0^Y = \frac{3}{2}R_0^X$. The reproduction numbers scale together in this way to emulate the effect of extrinsic factors that affect the spread of both strains, such as the mixing rate $\kappa$ of the population. $R_0^Y$ is assumed to be greater than $R_0^X$ to reflect

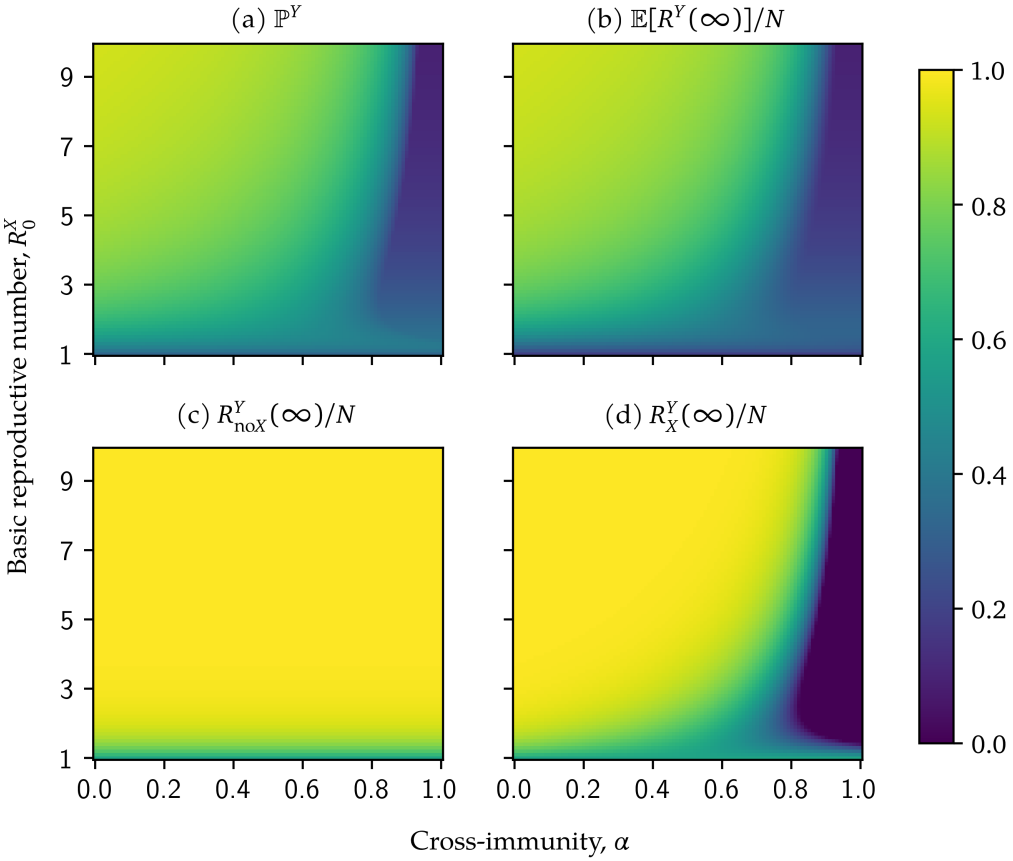



**Figure 2.1** Plots showing how the behaviour of strain $Y$ in the single-population model varies with $R_0^X$ and $\alpha$, when $R_0^Y = \frac{3}{2}R_0^X$. (a) The probability of a pandemic of strain $Y$ generally becomes more likely with increased $R_0^X$, and always decreases with increased $\alpha$, falling sharply at very high $\alpha$ where $R_e^Y$ goes below 1. (b) The expected number of individuals infected with strain $Y$ follows a similar pattern. This number is a linear combination of (c), which shows the scenario in which there is no epidemic of strain $X$ (in which case there is no dependence on $\alpha$); and (d), which shows the scenario in which there *is* such an epidemic.

the greater pandemic potential of strain $Y$.

In general, increasing cross-immunity decreases the probability and expected size of an epidemic of strain $Y$, while increasing the reproduction number of both $X$ and $Y$ increases the probability and expected size. There are, however, some non-monotonic effects for high $\alpha$, where increasing $R_0^X$ can cause the probability or size of an epidemic to decrease, and then increase. This is due to the fact that, at high levels of cross-immunity, a sufficiently large outbreak of strain $X$ will effectively completely block the invasion of strain $Y$, notwithstanding its higher $R_0$.

We can gain additional insight into these non-monotonic effects by considering equation 2.6 when $R_0^Y = \xi R_0^X$ and $I^Y(0) = I^X(0) = 1$, so that:

$$\begin{aligned}\mathbb{P}^Y &= 1 - \frac{1}{\xi\left(R_0^X\right)^2} - \frac{1-\frac{1}{R_0^X}}{\left(1-\alpha\frac{R^X(\infty)}{N}\right)\xi R_0^X} \\ &= 1 - \frac{1}{\xi R_0^X}\left(\frac{1}{R_0^X} + \frac{1-\frac{1}{R_0^X}}{1-\alpha\frac{R^X(\infty)}{N}}\right).\end{aligned} \tag{2.10}$$

Differentiating this with respect to $R_0^X$ we obtain:

$$\begin{aligned}\frac{\mathrm{d}\mathbb{P}^Y}{\mathrm{d}R_0^X} &= \frac{1}{\xi\left(R_0^X\right)^2}\left(\frac{1}{R_0^X} + \frac{1-\frac{1}{R_0^X}}{1-\alpha\frac{R^X(\infty)}{N}}\right) \\ &\quad - \frac{1}{\xi R_0^X}\left[-\frac{1}{\left(R_0^X\right)^2} + \frac{1}{\left(1-\alpha\frac{R^X(\infty)}{N}\right)\left(R_0^X\right)^2} + \frac{\alpha}{N}\frac{1-\frac{1}{R_0^X}}{\left(1-\alpha\frac{R^X(\infty)}{N}\right)^2}\frac{\mathrm{d}R^X(\infty)}{\mathrm{d}R_0^X}\right].\end{aligned} \tag{2.11}$$

The derivative of $R^X(\infty)$ with respect to $R_0^X$ has no analytic solution. Setting equation 2.11 to zero gives the turning point (or points) of the probability, i.e. the point(s) at which $\mathbb{P}^Y$ changes from decreasing with increasing $R_0^X$ to increasing with increasing $R_0^X$, or vice versa. Since the derivation of $\mathbb{P}^Y$ assumed that $R_0^X > 1$, we therefore have the condition that non-monotonic effects of $R_0^X$ on $\mathbb{P}^Y$ exist if and only if there exists some critical value of $R_{0,c}^X$ greater than 1 for which

$\frac{d\mathbb{P}^Y}{dR_0^X}\left(R_0^X = R_{0,c}^X\right) = 0$. Substituting this condition into equation 2.11, we therefore require that there is at least one value of $R_{0,c}^X > 1$ that satisfies the expression:

$$\begin{aligned}\frac{2}{R_{0,c}^X} = &-\frac{1}{1-\alpha\frac{R_{0,c}^X(\infty)}{N}} + \frac{2}{R_{0,c}^X\left(1-\alpha\frac{R^X(\infty)}{N}\right)} \\ &+ R_{0,c}^X\frac{\alpha}{N}\frac{R_{0,c}^X - 1}{\left(1-\alpha\frac{R^X(\infty)}{N}\right)^2}\frac{dR^X(\infty)}{dR_0^X}\left(R_0^X = R_{0,c}^X\right).\end{aligned} \tag{2.12}$$

Since $\xi$ does not appear in this equation, we conclude that the existence or otherwise of non-monotonic effects does not depend on the ratio of the basic reproduction numbers of $X$ and $Y$. It does, however, depend on the degree of cross-immunity $\alpha$, as we saw in figure 2.1. In fact, when $\alpha = 0$, equation 2.12 is a self-contradiction, meaning that non-zero cross-immunity is a necessary condition for a turning point to exist, and therefore for non-monotonicity to occur. It is not a sufficient condition—pandemic risk can be monotonic in $R_0^X$ with non-zero cross-immunity in the case that $R_{0,c}^X \leq 1$.

$\mathbb{P}^Y$ is monotonic with respect to $\alpha$, since

$$\frac{d\mathbb{P}^Y}{d\alpha} = -\frac{R^X(\infty)}{\xi\left(R_0^X\right)^2 N}\frac{R_0^X - 1}{\left(1-\alpha\frac{R^X(\infty)}{N}\right)^2} = 0 \tag{2.13}$$

implies that $R_0^X = 1$, which violates the assumption that $R_0^X > 1$. Therefore there is no turning point in $\mathbb{P}^Y$ with respect to $\alpha$.

In this section I have shown that cross-immunity can have a large effect on epidemic risk, and that it can introduce interesting non-monotonic effects to the infection dynamics. However, to understand how cross-immunity affects pandemic risk, we need to expand this model to include geographic structure.

# Pandemics on two regions ⬡ 2.2

An extremely simple kind of geographic structure can be introduced to an epidemiological model by dividing the population into two regions, numbered 1 and 2. Individuals in region $i$ are panmictic among themselves, but may travel to region $j$ at some rate given by $\lambda_{ij}$. Parameters such as $N$ and $R_e^Y$ could vary from region to region, so we add a subscript to distinguish the various values of these parameters, yielding $N_i$ and $R_{e,i}^Y$. For $N_i$ to be approximately constant over time, travel must either be low over the duration of the outbreak, or symmetric between regions such that $\lambda_{ij} = \lambda_{ji}$. The basic reproduction number $R_{0,i}^\Xi$ for strain $\Xi$ in region $i$ must now be understood as the *within-region* basic reproduction number, since it ignores the possibility for an individual to travel to a region with different epidemiological parameters and infect others there.

Now that the model has a geographic structure, we can define a pandemic on this population as an outbreak that goes on to become epidemic in both region 1 and region 2. Let us assume that an outbreak of $X$ is seeded in region 1. If an epidemic occurs in region 1, then there will be a total of $R_1^X(\infty)$ individuals who become infected over time, where the value of $R_1^X(\infty)$ can be obtained from equation 2.2 by adding the subscript 1 to all quantities. Note that equation 2.2 applies only approximately to the case where individuals can travel, under the assumption that the travel rate is small compared to the rate of recovery from infection. If travel rate is large, the flux of infective individuals due to travel can significantly impact the course of the epidemic, resulting in a different final size.

Infective individuals in region 1 recover from infection at a rate $\mu_1^X$, and travel to region 2 at a rate $\lambda_{12}$. So, the probability for any infective individual in region 1 to travel to region 2 before they recover is given by $\lambda_{12}/\left(\lambda_{12} + \mu_1^X\right)$. Under the assumption that each individual acts independently, the probability for a particular number of infective individuals $0 \leq n \leq R_1^X(\infty)$ to travel from region 1 to region 2

while infected with $X$ is given by the binomial distribution:

$$\Pr(n) = \binom{R_1^X(\infty)}{n}\left(\frac{\lambda_{12}}{\lambda_{12}+\mu_1^X}\right)^n\left(\frac{\mu_1^X}{\lambda_{12}+\mu_1^X}\right)^{R_1^X(\infty)-n}, \tag{2.14}$$

assuming that $\lambda_{12} \ll \mu_1^X$, so that infective individuals do not migrate and return during their infectious period. The probability that any of these $n$ individuals, infected with $X$, arriving in region 2 causes an epidemic of $X$ in region 2 is $1-\left(1/R_{0,2}^X\right)^n$ (see section 1.2). In what follows it will be easier to work with the probability that they do *not* cause an epidemic, i.e. $\left(1/R_{0,2}^X\right)^n$. Using this quantity and equation 2.14, we can perform a probability-weighted sum over each possible number of infective migrants from region 1 to 2, multiplied by the probability of that number of infective individuals failing to cause an epidemic. This gives us the probability $q_{12}^X$ that an epidemic of $X$ does *not* occur in region 2, given that an epidemic of $X$ *does* occur in region 1:

$$q_{12}^X \approx \sum_{n=0}^{R_1^X(\infty)}\left(\frac{1}{R_{0,2}^X}\right)^n\binom{R_1^X(\infty)}{n}\left(\frac{\lambda_{12}}{\lambda_{12}+\mu_1^X}\right)^n\left(\frac{\mu_1^X}{\lambda_{12}+\mu_1^X}\right)^{R_1^X(\infty)-n}, \tag{2.15}$$

Where the approximation is valid when $\lambda_{12} \ll \mu_1^X$ and $\lambda_{21} \ll \mu_2^X$, i.e. outward travel does not significantly affect the course of the epidemic in region 1, and infective individuals do not migrate and return within their infectious period. This sum has the closed form of a binomial expansion, so it can be rewritten as:

$$q_{12}^X \approx \left(\frac{\lambda_{12}}{R_{0,2}^X(\lambda_{12}+\mu_1^X)}+\frac{\mu_1^X}{\lambda_{12}+\mu_1^X}\right)^{R_1^X(\infty)}. \tag{2.16}$$

I assume that no epidemic of $X$ can occur in region 2 if none occurs in region 1.

It is important to understand that $q_{12}^X$ is a dimensionless quantity—a probability, rather than a rate. The final size equation 1.8 for $R_1^X(\infty)$ in section 1.2 is obtained by (implicitly) integrating over time, $R_{0,2}^X$ is a dimensionless ratio, and the rates $\mu_1^X$ and $\lambda_{12}$ with dimensions of inverse time appear in equation 2.16 only

in ratios with each other. This means that $q_{12}^X$ cannot give any information about *when* an epidemic of $X$ might occur, nor the duration of such an epidemic—it is the static all-time probability that no epidemic occurs in region 2 at *any* point in time, given that an epidemic occurs in region 1 at *some* point in time.

The total expected size of the outbreak of $X$ can be calculated by taking the sum of the expected size $R_i^X(\infty)$ of an epidemic of $X$ in each region $i$, multiplied by the probability of such an epidemic. The expected size $\mathbb{E}\left[R^X(\infty)\right]$ of an outbreak of $X$ that is seeded in region 1 is therefore:

$$\mathbb{E}\left[R^X(\infty)\right] = \left(1 - \left(R_{0,1}^X\right)^{-I_1^X(0)}\right)\left(R_1^X(\infty) + \left(1 - q_{12}^X\right)R_2^X(\infty)\right), \tag{2.17}$$

where the term in the first set of brackets gives the probability of an epidemic of $X$ in region 1, and $1-q_{12}^X$ gives the probability of an epidemic of $X$ in region 2, given that an epidemic of $X$ occurs in region 1.

Strain $Y$ is introduced following the outbreak of strain $X$. It has basic reproduction numbers $R_{0,i}^Y$ and recovery rates $\mu_i^Y$. If it is seeded in a region $i$ where no epidemic of $X$ has occurred, it will have a probability $Q_{i;\text{no}X}^Y = \min\left((1/R_{0,i}^Y)^{I_i^Y(0)}, 1\right)$ of not causing an epidemic in that region. If $R_{0,i}^Y < 1$, then no epidemic can occur and $Q_{i;\text{no}X}^Y = 1$. If $Y$ is seeded in a region where an epidemic of $X$ has occurred, it will have a probability $Q_{i;X}^Y = \min\left((1/R_{\text{e},i}^Y)^{I_i^Y(0)}, 1\right)$, where $R_{\text{e},i}^Y$ is defined analogously to $R_{\text{e}}^Y$ in equation 2.5. The probability $q_{ij}^Y$ of region $j$ not experiencing an epidemic of $Y$ given that region $i$ does experience an epidemic of $Y$ is given by:

$$q_{ij;\text{no}X,\text{no}X}^Y \approx \left(\frac{\lambda_{ij}}{R_{0,j}^Y(\lambda_{ij}+\mu_i^Y)} + \frac{\mu_i^Y}{\lambda_{ij}+\mu_i^Y}\right)^{R_{i;\text{no}X}^Y(\infty)}, \tag{2.18}$$

if no epidemic of $X$ has occurred in either region $i$ or region $j$;

$$q_{ij;\text{no}X,X}^Y \approx \left(\frac{\lambda_{ij}}{R_{e,j}^Y(\lambda_{ij}+\mu_i^Y)} + \frac{\mu_i^Y}{\lambda_{ij}+\mu_i^Y}\right)^{R_{i;\text{no}X}^Y(\infty)}, \tag{2.19}$$

if no epidemic of $X$ has occurred in region $i$, but one *has* occurred in region $j$;

$$q^Y_{ij;X,\text{no}X} \approx \left( \frac{\lambda_{ij}}{R^Y_{0,j}(\lambda_{ij}+\mu^Y_i)} + \frac{\mu^Y_i}{\lambda_{ij}+\mu^Y_i} \right)^{R^Y_{i;X}(\infty)}, \tag{2.20}$$

if an epidemic of $X$ has occurred in region $i$, but none has occurred in region $j$, or:

$$q^Y_{ij;X,X} \approx \left( \frac{\lambda_{ij}}{R^Y_{\text{e},j}(\lambda_{ij}+\mu^Y_i)} + \frac{\mu^Y_i}{\lambda_{ij}+\mu^Y_i} \right)^{R^Y_{i;X}(\infty)} \tag{2.21}$$

if an epidemic of $X$ has occurred in both region $i$ and region $j$, where the expected size $R^Y_i(\infty)$ of an epidemic of $Y$ in region $i$ is given by:

$$R^Y_{i;\text{no}X}(\infty) = N_i - N_i \exp\left( \frac{-R^Y_{0,i} R^Y_{i;\text{no}X}(\infty)}{N_i} \right), \tag{2.22}$$

or

$$R^Y_{i;X}(\infty) = N_i - N_i \exp\left( \frac{-R^Y_{\text{e},i} R^Y_{i;X}(\infty)}{N_i} \right). \tag{2.23}$$

Now, in order to focus on the effects of cross-immunity, I ignore the case where there is no epidemic of $X$, and therefore no cross-immune interactions. The probability $\mathbb{P}^Y_X$ that a pandemic of $Y$ occurs (i.e. both regions experience an epidemic of $Y$), given that at least one epidemic of $X$ occurs, is given by:

$$\begin{aligned}
\mathbb{P}^Y_X = {} & q^X_{12} Q^Y_{2;\text{no}X} \left(1 - Q^Y_{1;X}\right)\left(1 - q^Y_{12;X,\text{no}X}\right) + \left(1 - q^X_{12}\right) Q^Y_{2;X} \left(1 - Q^Y_{1;X}\right)\left(1 - q^Y_{12;X,X}\right) \\
& + q^X_{12} Q^Y_{1;X} \left(1 - Q^Y_{2;\text{no}X}\right)\left(1 - q^Y_{21;\text{no}X,X}\right) + \left(1 - q^X_{12}\right) Q^Y_{1;X} \left(1 - Q^Y_{2;X}\right)\left(1 - q^Y_{21;X,X}\right) \\
& + \left(1 - Q^Y_{1;X}\right)\left[ q^X_{12}\left(1 - Q^Y_{2;\text{no}X}\right) + \left(1 - q^X_{12}\right)\left(1 - Q^Y_{2;X}\right) \right],
\end{aligned} \tag{2.24}$$

where the first two terms correspond to $Y$ causing its first outbreak in region 1, before transmitting to region 2, and the second two correspond to $Y$ causing its first outbreak in region 2. The first term in each of these pairs treats the case where there is an epidemic of $X$ in region 1 only, and the second term treats the

case in which both regions experience an epidemic of $X$ (i.e. there is a pandemic of $X$). The final term corresponds to the initial $Y$-infected population seeding independent epidemics in both patches.

The expected final size $R_X^Y(\infty)$ of the outbreak of $Y$ is given by:

$$\begin{aligned}
R_X^Y(\infty) &= \mathbb{E}\left[\sum_i R_i^Y(\infty) \,\middle|\, \#\text{ epidemics of } X \geq 1\right] \\
&= \left(1-Q_{1;X}^Y\right)\Big[R_{1;X}^Y(\infty) + q_{12}^X\left(\left(1-Q_{2;\text{no}X}^Y\right) + Q_{2;\text{no}X}^Y\left(1-q_{12;X,\text{no}X}^Y\right)\right)R_{2;\text{no}X}^Y(\infty) \\
&\quad + \left(1-q_{12}^X\right)\left(\left(1-Q_{2;X}^Y\right) + Q_{2;X}^Y\left(1-q_{12;X,X}^Y\right)\right)R_{2;X}^Y(\infty)\Big] \\
&\quad + Q_{1;X}^Y\Big[q_{12}^X\left(1-Q_{2;\text{no}X}^Y\right)\left(R_{2;\text{no}X}^Y(\infty) + \left(1-q_{21;\text{no}X,X}^Y\right)R_{1;X}^Y(\infty)\right) \\
&\quad + \left(1-q_{12}^X\right)\left(1-Q_{2;X}^Y\right)\left(R_{2;X}^Y(\infty) + \left(1-q_{21;X,X}^Y\right)R_{1;X}^Y(\infty)\right)\Big].
\end{aligned} \tag{2.25}$$

We can now calculate the results of these equations for a particular case. In figure 2.2, I present a range of scenarios with the cross-immunity $\alpha$ varying from 0.5 to 0.8. I assume that $R_{0,i}^Y = \frac{3}{2}R_{0,i}^X$ for all $i \in \{1,2\}$, that $R_{0,1}^X = R_{0,2}^X$, and that the outbreak of $Y$ is seeded by a single infected individual appearing in either region 1 or region 2. The between-region travel is symmetric so that $\lambda_{12} = \lambda_{21} = \lambda$. The population size of each region is set to $N_1 = N_2 = 1000$ and the recovery rate is set to $\mu_i^\Xi = 0.14/\text{day} \approx 1/\text{week}$ for all $i \in \{1,2\}$ and $\Xi \in \{X, Y\}$.

From figure 2.2, we see that, for the relatively small value of $\alpha = 0.5$, increasing between-region travel or the basic reproductive number causes both the probability of a pandemic of $Y$ and the expected size of the outbreak to increase. However, with an increase of $\alpha$ to 0.7, the effect of travel on the expected size of an outbreak is reversed. So, although increased travel rates still increase the probability of a pandemic of $Y$, the expected severity of the outbreak is nevertheless diminished. This is due to the increased probability that $X$ causes an epidemic in region 2 as well as region 1, thereby granting those in region 2 partial immunity

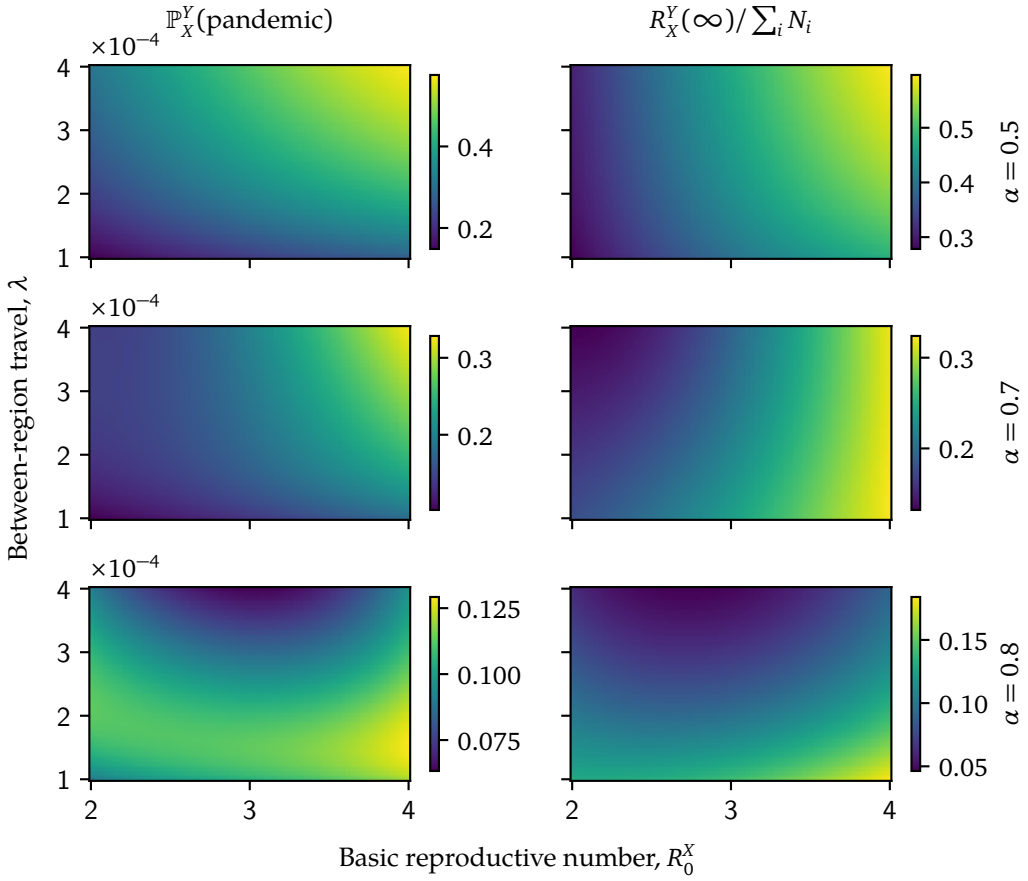



**Figure 2.2** Plots showing how the behaviour of strain $Y$ in the two-region model varies with $R_0^X$, $\lambda$, and $\alpha$, when $R_0^Y = \frac{3}{2}R_0^X$, assuming that there has been at least one epidemic of $X$. The first column shows the probability of a pandemic of $Y$ as given by equation 2.24, and the second shows the total proportion of the population who are expected to become infected as given by equation 2.25. Each row gives the results for a different value of the cross-immunity, $\alpha$. For $\alpha = 0.5$, increasing $R_0^X$ or $\lambda$ always increases pandemic probability and expected outbreak size, whereas for $\alpha = 0.7$ increasing $\lambda$ increases pandemic probability while reducing expected outbreak size. For $\alpha = 0.8$ there is non-monotonicity with both parameters. Each plot in this figure has a different colour scale, to illustrate the relative change in pandemic risk due to variations in $R_0^X$ and $\lambda$. For comparisons in risk between different values of $\alpha$, this figure and all following heat map figures are reproduced in appendix A with a single consistent colour scale.

to $Y$ and reducing the value of $R_2^Y(\infty)$.

At $\alpha = 0.8$, we see the return of the non-monotonicity in $R_0^X$ that we first encountered in figure 2.1, as well as some non-monotonicity in $\lambda$. In general, however, increasing $\lambda$ at this level of cross-immunity leads to a drop in both the probability of a pandemic of $Y$ and the expected size of the outbreak.

These results constitute a replication of Thompson et al. 2019 [49], although with an added focus on the pandemic outcome and on non-monotonic effects. We see that, even with simple geographic structure, cross-immunity can induce complex effects of long-distance travel on pandemic risk. To investigate further effects when the geographic structure is more complex, we must expand the metapopulation to more than two regions.

## Pandemics on networks ⬡ 2.3

Having described the dynamics of outbreaks occurring on two regions, in this section I will generalise this investigation to treat outbreaks spreading on populations with more complex geographic structure, as represented by travel networks (see section 1.4). This requires extending the two-region metapopulation model to allow arbitrary numbers of regions in the metapopulation, for which I use a Markov chain approach.

Suppose that the metapopulation contains $n$ regions, each with its own set of epidemiological parameters $N_i$, $R_{0,i}^X$, $\mu_i^X$, and $I_i^X(0)$. These yield values of $Q_i^X$ and $R_i^X(\infty)$ for each region. The rate of travel from a region $i$ to a region $j$ is given by $\lambda_{ij}$. Now suppose that a region $i$ (and only region $i$) of this network is currently experiencing an epidemic of $X$. For any given neighbour $j$ of $i$, the probability $q_{ij}^X$ that infective individuals travelling from region $i$ do not cause an epidemic in

region $j$ is given by:

$$q_{ij}^X \approx \left( \frac{\lambda_{ij}}{R_{0,j}^X \left( \mu_i^X + \sum_k \lambda_{ik} \right)} + \frac{\mu_i^X + \sum_{k \neq j} \lambda_{ik}}{\mu_i^X + \sum_k \lambda_{ik}} \right)^{R_i^X(\infty)}, \tag{2.26}$$

where $\sum_{k \neq j} \lambda_{ik}$ gives the total rate of travel from region $i$ to neighbouring regions other than region $j$. Equation 2.26 has the same structure as equation 2.16, since we are dealing with the dynamics of pathogen spread between pairs of regions. However, on a metapopulation with $n > 2$, $q_{ij}^X$ does not give us the overall probability that no epidemic occurs in region $j$, since an epidemic may occur in a neighbouring region other than region $i$, and be exported to region $j$. Instead, it gives us the conditional probability that no epidemic occurs in region $j$ due to the epidemic in region $i$, *given that* no region other than $i$ exports infective individuals to $j$ who go on to cause an epidemic in $j$.

To determine how the final probabilities of epidemics depend on these pairwise probabilities $q_{ij}^X$, I will use a Markov chain. This is a general mathematical tool for modelling the outcomes of a stochastic process by representing it as a series of steps, each of which assigns a probability to each possible state of the system being modelled [154]. The probabilities assigned at each step of the Markov chain depend only on the probabilities assigned at the immediately preceding step. The steps can represent discrete steps in time (e.g. days), or they can be more abstract steps used to explore the state space of a system, with no direct correspondence to calendar time. My Markov chain approach is of the latter type, with each state of the system not indicating the configuration of epidemics at a particular time, but rather acting as a bookkeeping device to keep track of which regions might export or import the pathogen.

I will now show how to obtain the final probability for each possible combination of regions experiencing an epidemic, using the conditional probabilities $q_{ij}^X$. We know the initial probability $1 - Q_i^X$ that the initial population of infective

individuals in a region $i$ causes an epidemic there. Let us call this the probability of region $i$ starting in state $E$, for epidemic. Then, let us assign the leftover probability $Q_i^X$ to a state called $N$, for neutral, meaning that we have not yet determined whether an epidemic will occur in this region.

For a region $i$ that is in state $E$, and therefore experiences an epidemic, we know that the probability that it fails to export this epidemic to a given neighbouring region $j$ is $q_{ij}^X$, as defined above. We can therefore assign a probability $\Pr(x_j = N \to y_j = E)$ for each region $j$ in state $N$ to transition to state $E$, given by:

$$\Pr(x_j = N \to y_j = E) = 1 - \prod_i \left(q_{ij}^X\right)^{\delta(x_i,E)} \tag{2.27}$$

where $x_i$ is the initial state of region $i$ , $y_i$ is the next state of region $i$, and

$$\delta(a,A) = \begin{cases} 1 & \text{if } a = A \\ 0 & \text{otherwise.} \end{cases} \tag{2.28}$$

This gives the probability that region $j$ experiences an epidemic due to infective individuals travelling there from a region which experiences an epidemic due to its initial infected population.

Each region which experiences an epidemic in this way also has a chance to export its epidemic to those of its neighbours which have not yet experienced epidemics, so we must calculate this transition probability a second time—this is the second step of the Markov chain. However, $q_{ij}^X$ is a probability calculated over all time, and therefore if the regions that experience initial epidemics remain in state $E$ we will end up double-counting their influence on their neighbours. We therefore set all regions that started the step in state $E$ to state $T$, for terminal, meaning that they no longer contribute to the probability for neighbouring regions to transition from state $N$ to state $E$—we have already decided that these regions will experience epidemics, and have determined to which further regions

they will export these epidemics.

This allows us then to define a transition matrix $\mathbf{T}^X$ for our Markov chain, whose elements $t^X_{x_1x_2\ldots x_n\to y_1y_2\ldots y_n}$ give the probability of transitioning from a system state $x_1x_2\ldots x_n \in \{N,E,T\}^n$ to a system state $y_1y_2\ldots y_n \in \{N,E,T\}^n$, at each step in the chain. We define these elements thus:

$$\begin{aligned} t^X_{x_1x_2\ldots x_n\to y_1y_2\ldots y_n} = \prod_j \Bigg[ & \Big(\delta(x_j,N)\delta(y_j,N) + \delta(x_j,N)\delta(y_j,E) \\ & + \delta(x_j,E)\delta(y_j,T) + \delta(x_j,T)\delta(y_j,T)\Big) \\ & \times \left(1 - \prod_i \left(q^X_{ij}\right)^{\delta(x_i,E)}\right)^{\delta(x_j,N)\delta(y_j,E)} \\ & \times \prod_i \left(q^X_{ij}\right)^{\delta(x_i,E)\delta(x_j,N)\delta(y_j,N)} \Bigg]. \end{aligned} \tag{2.29}$$

The first expression inside the square brackets ensures that the only acceptable transitions for any given region are $N \to E$ and $E \to T$, and requires that all epidemic regions in the initial state must be terminated in the transition. The second expression gives the probability of each $N \to E$ transition, and the third expression gives the probability of each $N \to N$ (non-)transition, given the regions that are in state $E$ before the transition. Since each region can be in one of three states, the total number of possible system states is $3^n$.

If all regions in the initial state are in state $N$ or $T$, then equation 2.29 implies that $t^X_{x_1x_2\ldots x_n\to y_1y_2\ldots y_n} = \prod_i \delta(x_i,y_i)$, i.e. the system state can no longer change. This means that the potential for each region to export the epidemic of $X$ has been calculated, and the final state assigns $T$ to those regions which experience epidemics, and $N$ to those which do not. Because each region can only be in state $E$ for a maximum of one step, we are guaranteed to reach the final state within $n$ steps or fewer.

The final element we require for our Markov chain is an expression for the

initial probability $P^X_{\text{initial},z_1z_2\ldots z_n}$ of each system state $z_1z_2\ldots z_n \in \{N,E,T\}^n$, given by:

$$P^X_{\text{initial},z_1z_2\ldots z_n} = \prod_i \Big[ (1-\delta(z_i,T)) \times (1-Q_i^X)^{\delta(z_i,E)} (Q_i^X)^{\delta(z_i,N)} \Big], \tag{2.30}$$

where $Q_i^X = \min((1/R^X_{0,i})^{I_i^X(0)}, 1)$. The first term inside the square brackets ensures that no region begins in state $T$, while the second term gives the initial probability of an epidemic due to the initial infected population. The vector of these probabilities is written $P^X_{\text{initial}}$.

The vector $P_{\text{final}}$ of the final probability assigned to each system state is therefore given by:

$$P^X_{\text{final}} = \left(\mathbf{T}^X\right)^n P^X_{\text{initial}}. \tag{2.31}$$

To find the expected final size $R^X(\infty)$ of the outbreak of $X$ we take the sum over the total number of infective individuals of each system state, multiplied by the final probability of that state:

$$\begin{aligned} R^X(\infty) &= \mathbb{E}\left[\sum_i R_i^X(\infty)\right] \\ &= \sum_{(z_1z_2\ldots z_n)\in\{N,T\}^n} P^X_{\text{final},z_1z_2\ldots z_n} \sum_i \delta(z_i,T) R_i^X(\infty). \end{aligned} \tag{2.32}$$

An example of the calculations involved in this Markov chain modelling framework is given in figure 2.3. This figure shows the case of a 4-region metapopulation, where travel is highly restricted and occurs only from region 1 to 2, region 2 to 4, region 1 to 3, and region 3 to 4. The probability that only region 1 begins in state $E$ is given by $(1-Q_1^X)Q_2^XQ_3^XQ_4^X$. I ignore the other possible initial states for the purpose of this demonstration. Reading from right to left, the possible transitions from this state are then shown, with their associated probabilities as given by equation 2.29. The probability for each final state can then be determined by multiplying the probabilities on each arrow that leads to it—for instance, the state $TNNN$ has final probability $p_{TNNN} = (1-Q_1^X)Q_2^XQ_3^XQ_4^Xq_{12}^Xq_{13}^X$.

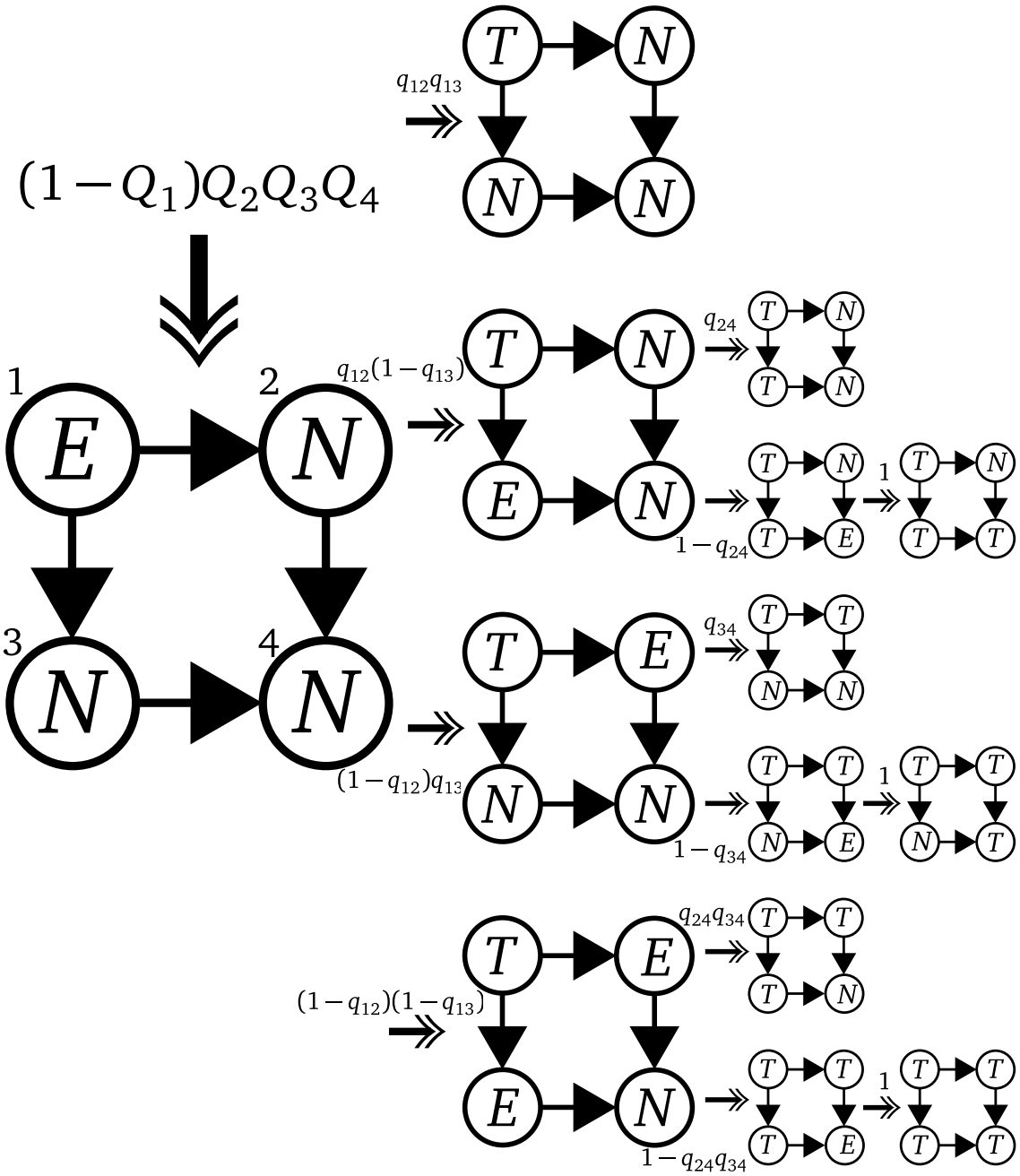



**Figure 2.3** An illustration of the Markov chain modelling framework as summarized in equation 2.31, with the superscript $X$ omitted on all quantities for clarity. A network of regions, represented by circles, is shown at several different points on a tree-like Markov chain, with capital letters $N$, $E$, and $T$ representing the states of each region at each step. Single-headed arrows represent travel routes between regions. Double-headed arrows represent transitions between states of the model, and are labelled with the corresponding transition probability given by equation 2.29. The leftmost, downward-pointing double-headed arrow is labelled with the probability of the illustrated initial state as given by equation 2.30.

Note that the order of transitions does not necessarily correspond to the order in which the pathogen spreads—for instance, on the path to the final state in which all regions are in state $T$, populations 2 and 3 become terminated before population 4. However, it is entirely possible for a pathogen in this metapopulation to spread to population 4 via population 2 before causing an epidemic in population 3—the order in which populations become terminated is simply an artefact of the model, and doesn't reflect underlying epidemiological dynamics.

To apply this model to an outbreak of strain $Y$ following the outbreak of $X$, we define $Q_i^Y$ and $q_{ij}^Y$ as functions of the final state $S = s_1 s_2 \dots s_n$ of the outbreak of $X$, so that:

$$Q_i^Y(S) = \begin{cases} \min((1/R_{0,i}^Y)^{I_i^Y(0)}, 1) & \text{if } s_i = N \\ \min((1/R_{\mathrm{e},i}^Y)^{I_i^Y(0)}, 1) & \text{if } s_i = T, \end{cases} \tag{2.33}$$

and

$$q_{ij}^Y(S) \approx \begin{cases} \left(\tilde{\lambda}_{ij}/R_{0,j}^Y + (1-\tilde{\lambda}_{ij})\right)^{R_{i;\mathrm{no}X}^Y(\infty)} & \text{if } s_i = N \text{ and } s_j = N \\ \left(\tilde{\lambda}_{ij}/R_{\mathrm{e},j}^Y + (1-\tilde{\lambda}_{ij})\right)^{R_{i;\mathrm{no}X}^Y(\infty)} & \text{if } s_i = N \text{ and } s_j = T \\ \left(\tilde{\lambda}_{ij}/R_{0,j}^Y + (1-\tilde{\lambda}_{ij})\right)^{R_{i;X}^Y(\infty)} & \text{if } s_i = T \text{ and } s_j = N \\ \left(\tilde{\lambda}_{ij}/R_{\mathrm{e},j}^Y + (1-\tilde{\lambda}_{ij})\right)^{R_{i;X}^Y(\infty)} & \text{if } s_i = T \text{ and } s_j = T, \end{cases} \tag{2.34}$$

where

$$\tilde{\lambda}_{ij} = \frac{\lambda_{ij}}{\mu_i^X + \sum_k \lambda_{ik}}. \tag{2.35}$$

We can then write the initial probability vector for $Y$ as:

$$P_{\mathrm{initial},z_1 z_2 \dots z_n}^Y(S) = \prod_i \Big[ (1-\delta(z_i,T)) \times (1-Q_i^Y(S))^{\delta(z_i,E)} (Q_i^Y(S))^{\delta(z_i,N)} \Big], \tag{2.36}$$

in analogy to equation 2.30, and the elements of the transition matrix for $Y$, $\mathbf{T}^Y$ as:

$$\begin{aligned} t^Y_{x_1x_2\ldots x_n\to y_1y_2\ldots y_n}(S) = \prod_j \Big[ &\Big(\delta(x_j,N)\delta(y_j,N)+\delta(x_j,N)\delta(y_j,E) \\ &+\delta(x_j,E)\delta(y_j,T)+\delta(x_j,T)\delta(y_j,T)\Big) \\ &\times\left(1-\prod_i\left(q^Y_{ij}(S)\right)^{\delta(x_i,E)}\right)^{\delta(x_j,N)\delta(y_j,E)} \\ &\times\prod_i\left(q^Y_{ij}(S)\right)^{\delta(x_i,E)\delta(x_j,N)\delta(y_j,N)}\Big], \end{aligned} \tag{2.37}$$

in analogy to equation 2.29. Finally, we write the probability vector $P^Y_{\text{final}}$ for the final states of the $Y$ outbreak as:

$$P^Y_{\text{final}}(S) = \sum_{(S)\in\{N,T\}^n} P^X_{\text{final},S}\left(\mathbf{T}^Y(S)\right)^n P^Y_{\text{initial}}(S). \tag{2.38}$$

To look exclusively at the case where there is at least one outbreak of $X$, we can instead use

$$P^Y_{\text{final};X} = \sum_{S\in\{N,T\}^n} P^X_{\text{final}\geq 1,S}\left(\mathbf{T}^Y(S)\right)^n P^Y_{\text{initial}}(S), \tag{2.39}$$

where

$$P^X_{\text{final}\geq 1,S} = \frac{\left(1-\prod_i\delta(s_i,N)\right)P^X_{\text{final},S}}{\sum_S P^X_{\text{final}\geq 1,S}}. \tag{2.40}$$

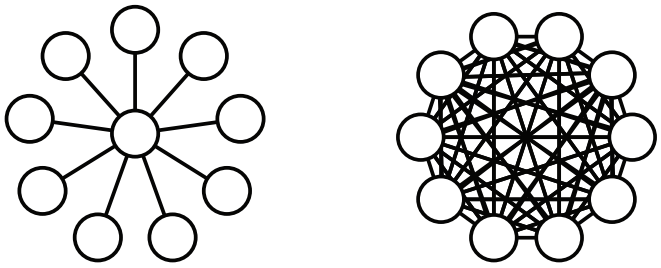
The final size of this outbreak is given by:

$$\begin{aligned} R^Y_X(\infty) = \sum_{S\in\{N,T\}^n}\sum_{(z_1z_2\ldots z_n)\in\{N,T\}^n} \Big[ &P^X_{\text{final}\geq 1,S}\left[\left(\mathbf{T}^Y(S)\right)^n P^Y_{\text{initial}}(S)\right]_{z_1z_2\ldots z_n} \\ &\times\sum_i \delta(z_i,T)R^Y_{i;S}(\infty)\Big], \end{aligned} \tag{2.41}$$

where

$$R^Y_{i;S}(\infty) = \begin{cases} R^Y_{i;\text{no}X}(\infty) & \text{if } s_i = N, \\ R^Y_{i;X}(\infty) & \text{if } s_i = T. \end{cases} \tag{2.42}$$

With these quantities defined, we can explore the effects of long-distance travel



(a) (b)

**Figure 2.4** Illustrations of (a) a star network and (b) a full network, each with ten regions. Circles represent regions, and straight lines represent travel routes between regions. These illustrations make clear the low connectivity of the star network, and the high connectivity of the full network.

and cross-immunity on pandemic risk in metapopulations with more than two regions. First, however, we must choose a definition of 'pandemic'. In the two-region model, it is clear that an outbreak that causes an epidemic in both regions is a pandemic, but the extension to the multiple region case is not obvious—is a pandemic any outbreak that spreads to more than one region, or must it cause epidemics in all regions? I will return to this question in chapter 3, but for our current purposes I will choose the most restrictive definition of a pandemic, requiring an epidemic to occur in every region of the metapopulation. This means that the probability of a pandemic is equal to the $TT\ldots T$ element of $P^Y_{\text{final};X}$, which I will denote $\mathbb{P}^Y_X$ in agreement with earlier sections.

In a metapopulation with more than two regions, the geographic structure is defined by the network of travel between the regions. Here I will explore the dynamics of cross-immunity in two alternative networks, illustrated in figure 2.4. The star network, defined by $\lambda_{ij} = \lambda(\delta(i,1) + \delta(j,1))$, has lower overall motility, and comprises a set of peripheral regions with no direct travel routes between them, connected by a hub region with equally weighted routes connecting it to each peripheral region. The fully connected network, defined by $\lambda_{ij} = \lambda$, has higher overall motility and is fully symmetric, with each region identically connected by equally weighted travel routes to each other region. Even with such a simple structure, however, cross-immunity can produce interesting dynamics.

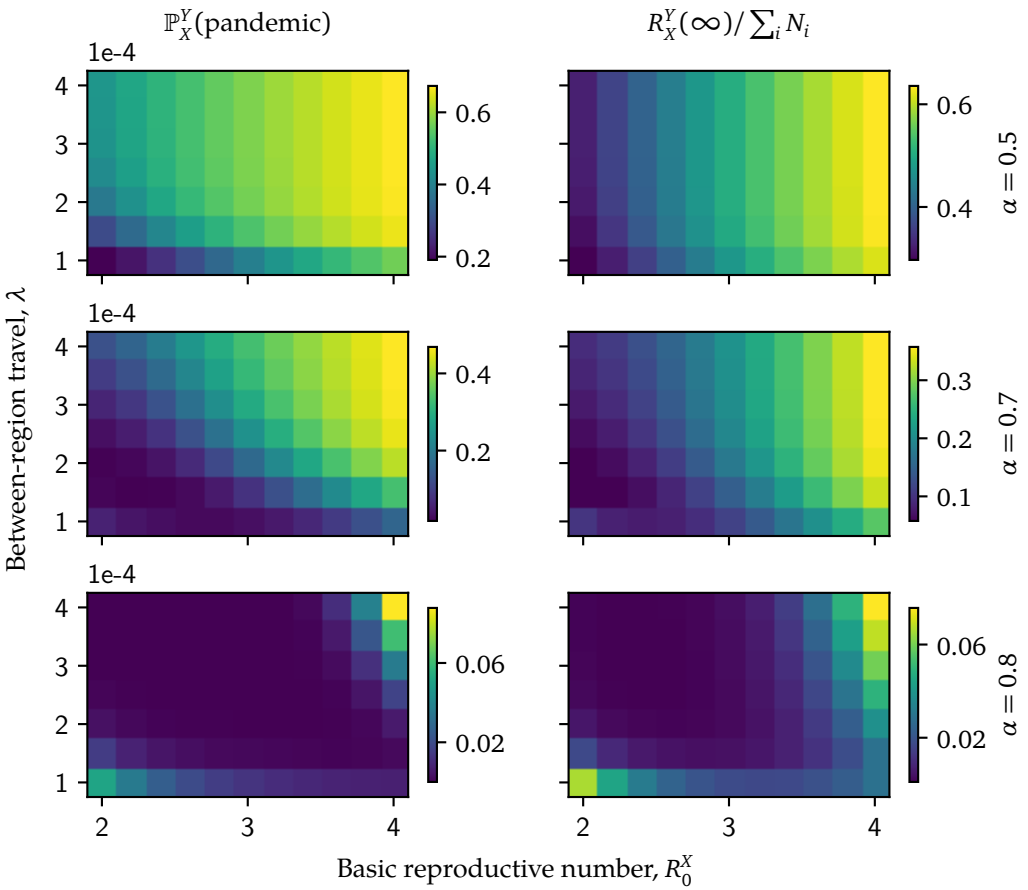



**Figure 2.5** Plots showing how the behaviour of strain $Y$ in the ten-region fully connected network varies with $R^X_0$, $\lambda$, and $\alpha$, when $R^Y_0 = \frac{3}{2}R^X_0$, assuming that there has been at least one epidemic of $X$. The first column shows the probability of a pandemic of $Y$, given by the $TT\dots T$ element of $P^Y_{\text{final};X}$ in equation 2.39. The second column shows the total proportion of the population who are expected to become infected, as given by equation 2.41. Each row gives the results for a different value of the cross-immunity, $\alpha$. Slight non-monotonicity with both $\lambda$ and $R^X_0$ is visible at $\alpha = 0.7$, and becomes much stronger when $\alpha = 0.8$. A version of this figure with a single consistent colour scale is included as figure A.2 in appendix A.

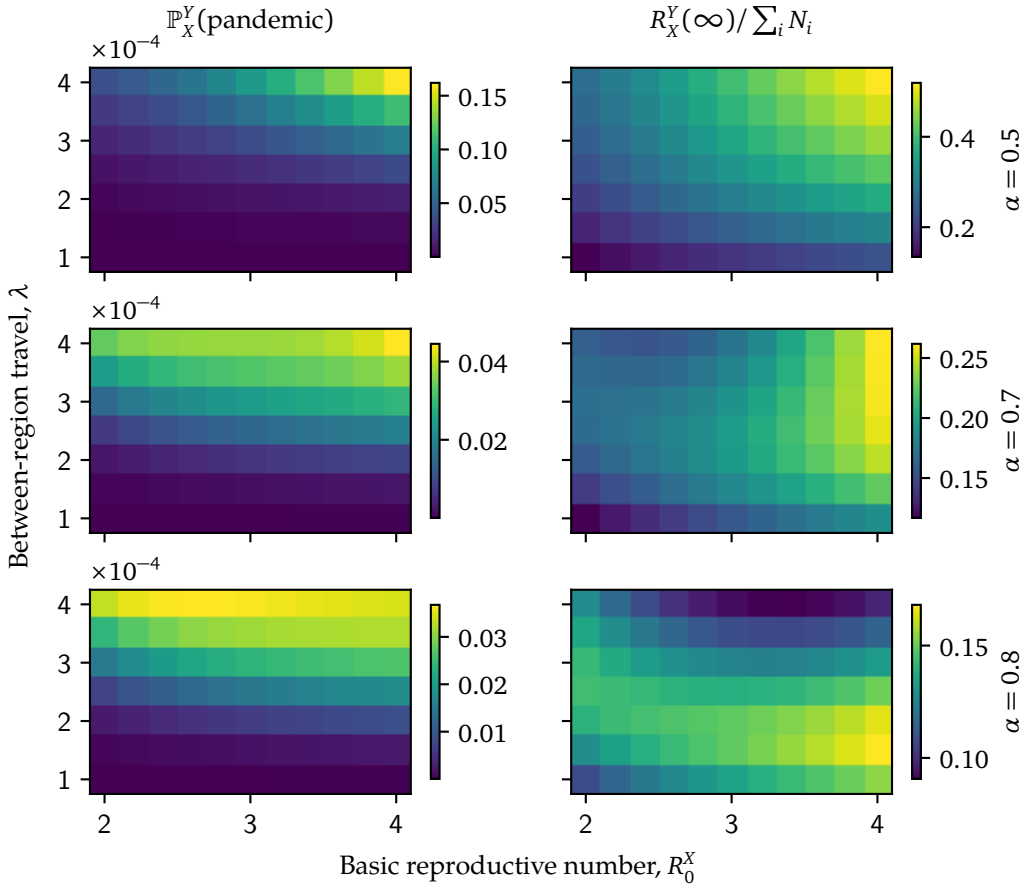



**Figure 2.6** Plots showing how the behaviour of strain $Y$ in the ten-region star network varies with $R_0^X$, $\lambda$, and $\alpha$, when $R_0^Y = \frac{3}{2}R_0^X$, assuming that there has been at least one epidemic of $X$. The first column shows the probability of a pandemic of $Y$, given by the $TT\ldots T$ element of $P^Y_{\text{final};X}$ in equation 2.39. The second column shows the total proportion of the population who are expected to become infected, as given by equation 2.41. Each row gives the results for a different value of the cross-immunity, $\alpha$. Pandemic probability always increases with increasing $\lambda$, but the same is not true for the expected infected proportion, which is non-monotonic with $\lambda$ for $\alpha = 0.7$ and $\alpha = 0.8$. A version of this figure with a single consistent colour scale is included as figure A.3 in appendix A.

In figure 2.5, I present the pandemic probability and expected outbreak size in a range of scenarios in the 10-region fully connected network, with the cross-immunity $\alpha$ varying from 0.5 to 0.8. The outbreak of $X$ is seeded by a single individual, with the region in which that individual appears being arbitrary due to the symmetry of the network. I choose to always seed the outbreak of $X$ in region 1, but this is fully equivalent to seeding it randomly. As in the sections above, I assume that $R^Y_{0,i} = \frac{3}{2}R^X_{0,i}$ for all $i \in \{1,\dots,10\}$, that $R^X_{0,1} = R^X_{0,2} = \cdots = R^X_{0,10}$, and that the outbreak of $Y$ is seeded by a single infected individual with a 1 in 10 probability of appearing in each region. The population size of each region is set to $N_i = 1000$ for all $i \in \{1,\dots,10\}$, and the recovery rate is set to $\mu^{\Xi}_i = 0.14/\text{day}$ for all $i \in \{1,\dots,10\}$ and $\Xi \in \{X, Y\}$. These plots are less smooth in $\lambda$ and $R^X_0$ due to the greater computational intensity of the multi-region model compared to the two-region model (see section 2.4).

In general, increasing cross-immunity decreases both the probability of a pandemic of $Y$ and the expected size of an outbreak of $Y$, as can be seen by the changing scale on the plots in figure 2.5. For $\alpha = 0.5$, increasing $R^X_0$ and $\lambda$ both always result in increased pandemic probability and outbreak size. However, at $\alpha = 0.7$, a small non-monotonic effect can be seen at low values of $R^X_0$ or $\lambda$, and for $\alpha = 0.8$ there is large variation in the effect of $\lambda$. For lower values of $R^X_0$, the circulation of $X$ as driven by $\lambda$ must be very low in order for $Y$ to have any chance of causing a pandemic. This is because the effective reproduction number $R^Y_e$ is very low, so $Y$ has a very low probability of causing epidemics in regions that have experienced outbreaks of $X$. However, as $R^X_0$ increases, $R^Y_e$ increases, and increased between-region travel begins to favour circulation of $Y$, leading to an increase in pandemic probability with increasing $\lambda$.

In figure 2.6 I present analogous results for pandemic probability and expected outbreak size in the 10-region star network. Due to the low connectivity of the network, the probability of a pandemic is in general very low compared to

the fully-connected network. However, because the low connectivity impedes the transmission of $X$ around the metapopulation, increasing the cross-immunity has a less dramatic effect on pandemic probability, because fewer regions are likely to have partially immune populations. Another consequence of this low connectivity is that, for the range of parameters considered, increasing travel rates always increase pandemic probability, no matter the value of $\alpha$. However, the same is not true of the expected size of an outbreak. For high $\alpha$ this can show non-monotonic effects in $\lambda$, even as $\mathbb{P}_X^Y(\text{pandemic})$ increases monotonically. In certain parameter ranges, the pandemic risk can increase while the expected size decreases, as in the $\alpha = 0.7$ row of figure 2.2. This difference exists in the star network for high $\alpha$ because a large portion of the expected size comes from scenarios in which $Y$ does not become a pandemic. At high levels of $\lambda$ an "all-or-nothing" dynamic emerges. The probability that $Y$ becomes a pandemic is enhanced by the greater travel rates, but the greater circulation of $X$ reduces the likelihood and size of non-pandemic outbreaks, leading to an overall reduction in the expected final size.

In metapopulations with multiple regions we see the complex non-linear effects of cross-immunity in the two-region model preserved and extended. Model behaviours such as pandemic probability increasing but expected outbreak size decreasing, or non-monotonicity of pandemic probability in $R_0^X$ and $\lambda$, can both be replicated in larger metapopulations. The network structure partially determines which of these effects we see, with non-monotonicity taking a markedly different form in the star network than in the full network. This means that not only the scale of travel, but also the connectivity of certain regions on the travel network, can be important to understanding the effects of cross-immunity on pandemic risk.

## Efficiency of the Markov chain

The Markov chain modelling framework defined in section 2.3 is a novel method I devised in order to extend the results of Thompson et al. 2019 [49] to metapopulations with more than two regions. This framework can be used to calculate pandemic probabilities exactly (up to the approximations included in the definitions of $q_{ij}^X$ and $q_{ij}^Y$), instead of approximating it with a large number of individual-based simulations. This approach is flexible, and in chapter 3 I will use it to investigate different aspects of the dynamics of infectious disease on travel networks.

In some simulation-based epidemiological models, the proportional error for the estimated probabilities of unlikely events can be very large, since these events occur in very few simulations. This can be an issue for agent-based models, for example. An advantage of the Markov chain framework is that small probabilities require no more computation to estimate than large ones. Since pandemics are rare events and can have very low probabilities, this is a distinct advantage in theoretical pandemic preparedness research.

Here and in chapter 3 I focus on metapopulations with 10 regions or fewer. This relatively small number of regions is sufficient to explore a wide range of theoretical questions about pandemic dynamics, and for this number of regions the model described above can run quickly on a laptop computer. The time taken to compute the $n^{\text{th}}$ power of the transition matrix in equation 2.38 is of the order $O(n3^{\omega n})$, where $2 \leq \omega \leq 3$ varies depending on the matrix multiplication algorithm used, meaning that the model runs slowly for large $n$. Networks based on empirical data can have large numbers of regions (see chapters 4 and 5). In this case, the Markov chain calculations can take a long time, and a less exhaustive simulation-based approach, such as a Monte Carlo method, may be more appropriate.

To calculate pandemic probabilities and final sizes in a metapopulation where

each region has the same population size, we only need to know the probability for a given number of regions to experience epidemics, and not the particular regions in which these epidemics occur. This means that, in highly symmetric travel networks, such as the star and full networks, the efficiency of the computation of $P^Y_{\text{final}}$ can be increased by exploiting the fact that many final states of the $X$ outbreak are equivalent up to a relabelling of the regions. Without this simplification, the sum in equation 2.38 has $2^n$ terms—i.e. 1024 terms for a 10-region network.

Suppose that some set $S_e$ of possible final states of the $X$ outbreak are identical up to a renumbering of the regions. We can sum the probabilities of these states to obtain $P_S = \sum_{s_1 s_2 \ldots s_n \in S_e} P^X_{\text{final},s_1 s_2 \ldots s_n}$. We can then define a new $\hat{P}^X_{\text{final}}$, in which the probability for some state $s_1 s_2 \ldots s_n \in S_e$ is set to $P_S$, and the probability for all other states in $S_e$ is set to zero. Replacing $P^X_{\text{final}}$ in equation 2.38 with $\hat{P}^X_{\text{final}}$ yields the expression:

$$\hat{P}^Y_{\text{final}}(S) = \sum_{(S)\in\{N,T\}^n} \hat{P}^X_{\text{final}}(S) \times \left(\mathbf{T}^Y(S)\right)^n P^Y_{\text{initial}}(S). \tag{2.43}$$

$\hat{P}^Y_{\text{final}}(S)$ is in general different from $P^Y_{\text{final}}(S)$, but gives the same probability of a pandemic of $Y$ and expected final size of an outbreak of $Y$.

Performing this procedure for each set of equivalent final states can greatly reduce the number of non-zero terms in 2.38, allowing much faster computation. For the fully connected network, where all regions are equivalent, the only salient feature of the final state of the $X$ outbreak is the number of terminated regions from 0 to $n$—yielding 11 non-zero terms in the simplified sum for a 10-region fully connected network, a near 100× efficiency gain. For the star network, the central region differs from the peripheral regions, so there are $2n$ distinct scenarios, one for each combination of peripheral regions terminated from 0 to $n-1$ with either a neutral or terminated central region. Therefore there are 20 non-zero terms in the simplified version of equation 2.43 in the case of the 10-region star network.

## Conclusions ⌬ 2.5

In this chapter, I have replicated the results of Thompson et al. regarding the effects of long-distance travel and cross-immunity on pandemic risk in a two-region metapopulation [49]. I have also extended this work to treat the case of metapopulations with an arbitrary number of regions, computing results for two 10-region metapopulations with star-shaped and fully connected travel networks. I have found that pathogens that encounter high levels of cross-immunity due to previously circulating strains pose a smaller pandemic risk, both by being less likely to cause pandemics and by causing smaller outbreaks on average (including non-pandemic outbreaks). Cross-immunity interacts with long-distance travel in interesting ways—the presence of cross-immunity can lead to deviations from the assumption that increased long-distance travel always increases pandemic risk, and can even cause it to have non-monotonic effects. The nature of these deviations depends on the scale and structure of travel, and on the level of cross-immunity. Taken together these results suggest that cross-immunity could have large and manifold effects on pandemic outcomes as international travel continues to grow and change.

Some pathogens with pandemic potential (PPPs) belong to families of pathogens that have circulated widely among the global population. Such pathogens include pandemic strains of influenza, such as the 2009 strain of influenza A [122, 151, 152]. These pathogens are likely to encounter populations with cross-immunity reducing either the probability of infection or the probability of disease, and we should expect that the higher this cross-immunity is, the smaller the risk these pathogens pose. The modelling I have presented in this chapter shows that, for pathogen families that induce particularly high levels of cross-immunity, increasing rates of long-distance travel may serve to increase the risk-reducing effects of cross-immunity.

These effects rely on previously circulating relatives of a pathogen, which do not exist for every PPP. HIV is one of only two kinds of retrovirus which cause disease in humans, and it has highly successful strategies for evading the immune system [155, 156]. This has helped make the pandemic of HIV/AIDS one of the most severe in human history, with the global population having no immunity or cross-immunity to impede its spread. For a pathogen like this, we do not expect to see risk-reducing effects of long-distance travel—instead, pandemic risk from novel viruses is likely to increase monotonically with increasing global mobility.

SARS-CoV-2 falls between these two categories, since it had a precursor strain in the form of SARS-CoV, but this strain did not cause large outbreaks worldwide [157]. In effect, SARS-CoV-2 played the role of pathogen $X$, with negligible amounts of pre-existing immunity in the human population. However, if SARS-CoV had spread more widely, it is possible that the spread of SARS-CoV-2 would have been impeded (although the details of cross-immunity between the two strains are not well known). Note that this would not have been a desirable outcome—SARS-CoV had a much higher fatality rate than SARS-CoV-2. The work presented here does not in general support a strategy of allowing pandemic pathogens to spread unimpeded in the hope that they will induce immunity against future outbreaks.

Cross-immunity cannot give the whole story of how pandemic outcomes vary, and how pandemic risk will change in the future. At the very least, different kinds of immunological interactions are possible, with some kinds of immune responses actually increasing mortality from infection [140, 158]. This analysis also does not take into account any of the epidemiological effects of climate change, or changes in the agricultural sector that may affect the rate of zoonosis [134, 159, 160]. Perhaps the most important thing to learn from this investigation, in which we have encountered non-linearity and non-monotonicity in several regimes, is that the consequences of changing epidemiological parameters

are not likely to be straightforward. In a changing world, those working towards pandemic preparedness must be ready to face unknown unknowns.





A photograph of an icosahedral die with its faces inscribed with Greek letters, from the Metropolitan Museum of Art in New York [161]. The regular icosahedron is a popular shape for dice, being a highly symmetric polyhedron with many faces, and icosahedral dice have gained a contemporary association with risk and reward from their use in the game Dungeons & Dragons and in Magic 8-Balls. This die was produced in Ptolemaic Egypt, and may have been used for divination, "throwing lots to obtain a number that would, through certain algorithms, lead to ready-prepared oracle questions and responses."

# 3 Pandemic definitions

In chapter 2, I made the choice to define a pandemic in a metapopulation model as an outbreak in which epidemics occur in all regions of the metapopulation, using the definition of an epidemic in terms of the bimodal distribution of outbreak sizes in the stochastic panmictic SIR model (see section 1.2). This choice is not the only one I could have made, and I will show in this chapter that the choice is consequential—a quantitative change in the pandemic definition used can result in a qualitative change to conclusions regarding pandemic risk.

Determining pandemic risk in a metapopulation model requires a specific quantitative definition of a pandemic. However, such definitions are not widely agreed upon in the epidemiology community. The International Epidemiology Association's Dictionary of Epidemiology defines a pandemic as "an epidemic occurring worldwide, or over a very wide area, crossing international boundaries and usually affecting a large number of people" [1]. Notably this definition makes an explicit reference to national borders. Contrastingly, a World Health Organization (WHO) source makes reference to a pandemic as "the worldwide spread of a new disease" [3]. The use of the word 'new' here is ambiguous in the context of infectious diseases. HIV/AIDS is often referred to as a global pandemic, but is certainly not new on the timescale of, say, the emergence of influenza strains [162, 163]. A study by Morens et al. in 2009 finds that there is little in common between all disease outbreaks that have been referred to as pandemics, except that they have a wide geographical extension [2].

Both definitions above use qualitative terms such as "large number" and "worldwide", without providing numerical criteria for their application to pandemics. Perhaps as a result of this, many quantitative studies on pandemics do not make use of a quantitative definition of a pandemic, but instead focus on causally related concepts, such as sustained transmission [164], or emergence of novel pathogens [165]. Others treat the global spread of a pathogen as a context in which to study transmission dynamics, without paying special attention to how those dynamics might lead to a pandemic as distinct from an epidemic or a more limited outbreak [148, 166, 167]. As such, there is little precedent in the epidemiological literature for pandemic definitions suitable for use in a metapopulation model.

One possible explanation for this lack of a consensus quantitative pandemic definition is that using different definitions of 'pandemic' could make little difference to the results of quantitative analyses of pandemic risk. In this scenario, a variety of implicit definitions could be used without affecting the communication of key scientific conclusions. We can test whether this is true by creating a simple epidemiological model, and interpreting its results using a variety of pandemic definitions.

I will explore two different kinds of pandemic definition, following Morens et al. 2009 [2], specifically:

- the family of *transregional* definitions, where a pandemic is defined as an outbreak in which the number of regions experiencing epidemics meets or exceeds some threshold number $n$. I refer to specific transregional definitions as *n-region* transregional definitions, e.g. a 3-region transregional definition.

- the *interregional* definition, where a pandemic is defined as an outbreak in which two or more non-adjacent regions experience epidemics.

Note that these definitions require a specific sense of 'region'. These regions could

be countries, or they could be smaller or larger than individual countries—from counties to health zones to WHO regions. My Markov chain modelling framework can be used to model regions of any size. I have chosen not to include definitions with criteria relating to the number of people infected or killed, instead of, or in addition to, geographical extension. Instead, I focus on extension, this being the only universal factor in pandemic definitions [2].

Three questions that help form public health policy at the beginning of an outbreak are:

- Would interventions restricting travel reduce the risk of a pandemic?
- Does a portion of the population have pre-existing immunity, and does this affect the risk of a pandemic?
- How is the risk of a pandemic affected by regional differences in transmission?

In this chapter, I will use the Markov chain modelling framework defined in section 2.3 to explore how changing the pandemic definition does or does not affect our answers to these questions. I will show that the precise definition of a pandemic used in modelling studies can (but does not always) affect the inferred risk. The predicted effects of travel restrictions, the influence of pre-existing immunity, and the impact of regional differences in transmission can all vary when alternative definitions of 'pandemic' are used. These differences can be qualitative in nature, such as an increase in pandemic risk changing to a decrease under a different definition. This demonstrates clearly the need to consider carefully the pandemic definition used to assess the risk from an invading pathogen.

# Between-region travel 3.1

In chapter 2, we saw that between-region travel can have large effects on the probability of a pandemic [49, 95, 129, 168]. Here, I will discuss how these effects are interpreted under different definitions of 'pandemic'.

Using the Markov chain modelling framework introduced in section 2.3, and the star and fully-connected networks shown in figure 2.4, we can investigate pandemic risk under a range of different transregional definitions. The pandemic probability under a range of $n$-region transregional definitions for a 10-region network with a variety of travel rates is shown in figure 3.1. I set $\mu = 0.14$, and $R_0 = 2$, dropping the superscript $X$ since we are treating the case of a single pathogen. An $R_0$ of two is within the plausible range for both seasonal and pandemic influenza, so it is an appropriate choice for a simulated PPP [167]. As in chapter 2, each region is given a population size of 1000, and the outbreak is seeded by a single infected individual in one region. For the full network I choose a single arbitrary region in which to seed the outbreak, since each region is connected identically. In the star network, I take the average pandemic probability over outbreaks seeded in each region.

An $n$-region transregional definition effectively provides a threshold number $n$—if more than $n$ regions experience epidemics, the outbreak is counted as a pandemic, and otherwise it is not. Thus I indicate the different possible $n$-region definitions through their threshold numbers on the y-axis. The 1-region transregional definition merges the definitions of 'pandemic' and 'epidemic' in an implausible way, but it is included in these figures for comparison. The comparison between the pandemic probability according to the 2-region definition and according to the 10-region definition shows the difference between pandemic definitions that are satisfied by any transregional transmission, and definitions that are satisfied only by truly global spread. For the star network, or for the fully connected network

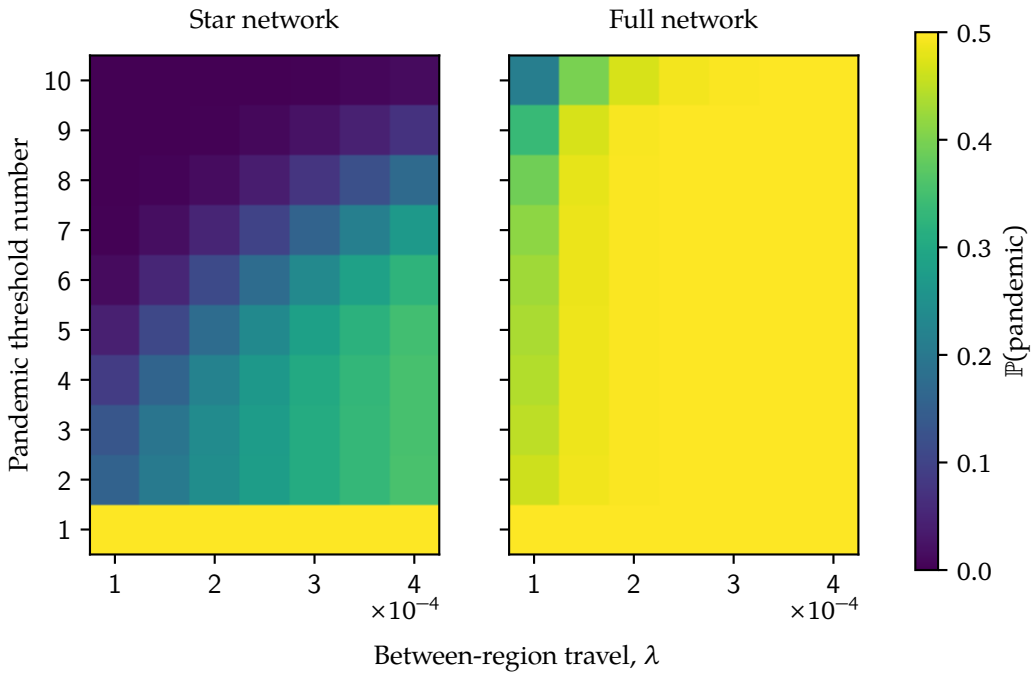



**Figure 3.1** Plots showing how pandemic probability varies for a range of between-region travel rates under a range of transregional pandemic definitions, in the 10-region star and fully connected networks. The "pandemic threshold number" refers to the minimum number of regions that must experience epidemics before a pandemic is judged to have occurred. The pandemic probability is, in general, sensitive to the pandemic definition used, but the degree of sensitivity depends on network structure and travel rates. In the star network, the pandemic probability is, in general, highly sensitive to the pandemic definition used, while in the full network this sensitivity is significantly reduced at high travel rates.

with low travel rates, there is a marked difference between the probability of either of these definitions being satisfied. However, for the fully connected network at medium or high rates of travel, if the pathogen causes at least one epidemic, then it has a very high probability of spreading globally. As such, the probability of a pandemic nears the maximum of 0.5 (i.e. $1 - 1/R_0$) at all thresholds. For any definition, the probability of a pandemic increases with the connectivity of the network, and with travel rates across the network.

We can also explore the difference in pandemic probability between the transregional and interregional definitions, which make use of a distinction between adjacent and non-adjacent regions. This is shown for a 10-region star network in figure 3.2, which illustrates difference in pandemic risk under the 2-region transregional and 2-region interregional definitions, compared to that under the 10-region transregional definition. There is a difference in pandemic probability between the 2-region interregional and transregional definitions, but the difference is much smaller than that between the 2-region interregional and 10-region (global) definition, and reduces as travel rates increase. In the case of a fully connected network, all regions are adjacent to each other, so I compare only the 2-region transregional and global definitions. The definitions are clearly distinct for low travel rates, but as the travel rate increases the difference between the likelihood of a pathogen causing an epidemic in two regions and the likelihood of it causing epidemics in all regions approaches zero. This is due to the fact that the pathogen can be introduced into any population from any other.

When a pandemic is defined in terms of the regions that experience epidemics of a disease, we have seen that different definitions can produce very different estimates of the pandemic probability at low travel rates or in the star network, but have a much smaller effect at high travel rates or in the full network. The 10-region star network has much lower connectivity than the 10-region fully connected network, having 9 edges as opposed to 45. Therefore, if $\lambda$ is constant, the total volume

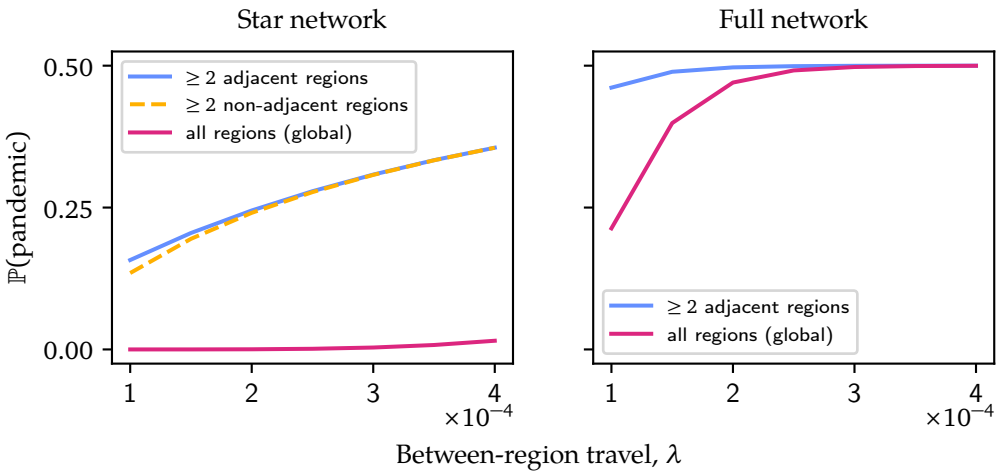



**Figure 3.2** Plots of pandemic probability against between-region travel rate under a range of pandemic definitions, in the 10-region star and fully connected networks. The difference in probability under different pandemic definitions changes as travel rates increase. In a fully connected network all regions are adjacent, so no line is shown for the interregional definition, which requires non-adjacent regions to experience epidemics.

of travel per unit time in the star network is one fifth of that in the full network. To understand how much of the different behaviour in the two networks is due to their different connectivity, and how much is due to the different overall volume of travel, we can multiply the travel rates in the star network by five, and divide those in the full network by five. The results are shown in figures 3.3 and 3.4.

Referencing these figures, we can isolate the effects of the connectivity of these different networks from the effect of the change in overall travel volume. The high connectivity of the full network amplifies the effects of increasing travel in general, as circulation around the network is high. Contrastingly, the lower connectivity of the star network means that risk increases more gradually with travel, since there is only ever one possible route along which the pathogen can be transmitted from one region in the network to another.

We are now in a position to return to the first of the questions introduced on page 63, i.e. "Would interventions restricting travel reduce the risk of a pandemic?" We have seen that reductions in travel rates always reduce risk in the star network, where travel occurs mainly through a central hub. However, in a network

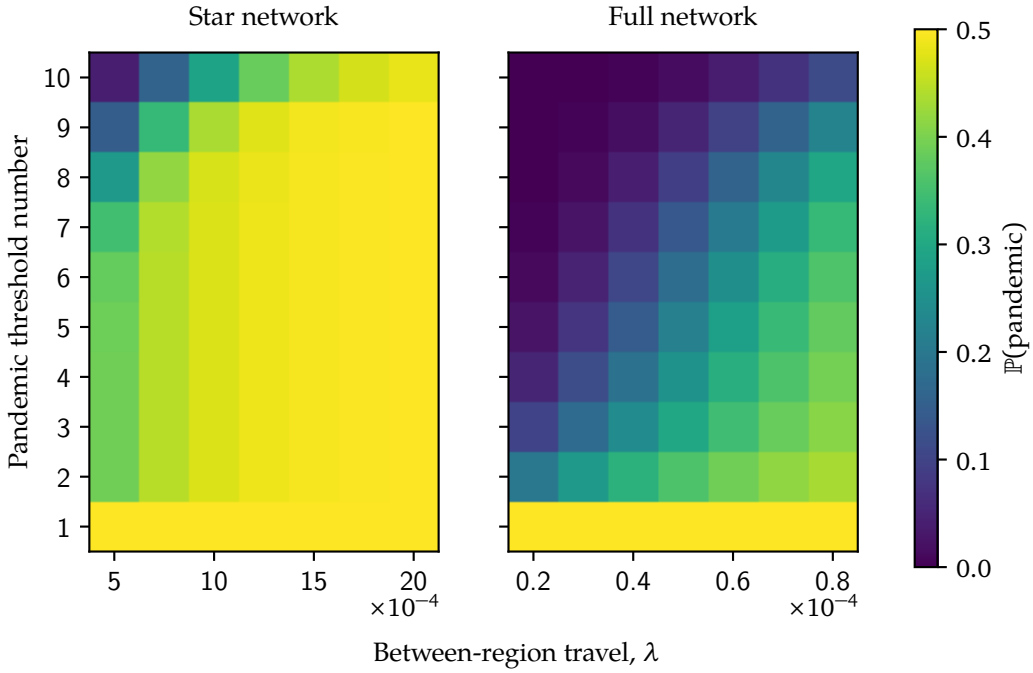



**Figure 3.3** Analogous plots to figure 3.1, but with the travel rate for the star network increased by a factor of five, and that for the full network decreased by the same, so as to swap the total volume of travel in the two networks. We see that the star network with high travel rates shows similar behaviour to the full network at medium travel rates, but with more gradual variation as travel increases. The full network at low travel rates shows similar behaviour to the star network at medium travel rates, but with greater changes in pandemic probability as the pandemic threshold number changes.

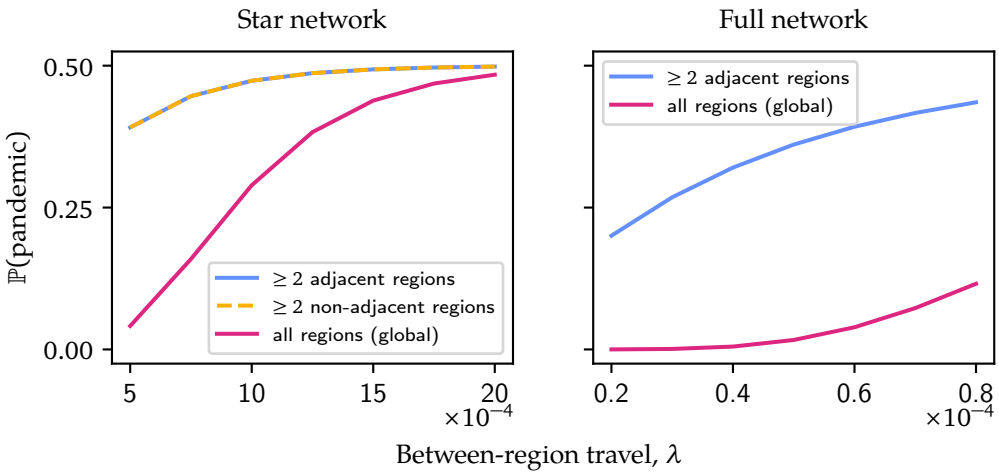



**Figure 3.4** Analogous plots to figure 3.2, but with the travel rate for the star network increased by a factor of five, and that for the full network decreased by the same, so as to swap the total volume of travel in the two networks. The 2-region transregional and interregional definitions are equivalent at very high travel rates, but in the star network the global definition remains distinguishable from the 2-region definition even with high travel. The full network at low travel rates resembles the star network at medium travel rates, but the probability of a global pandemic increases much more quickly with increasing travel.

with high connectivity and high travel rates, travel would have to be extremely highly suppressed to result in a substantial change in the probability of a pandemic, under most definitions. This accords with previous findings regarding the effectiveness of restricting travel [169]. Additionally, when connectivity is high, changing the definition of a pandemic makes little difference to the pandemic probability, for high enough values of the travel rate.

The line plots in figures 3.2 and 3.4 further illustrate the effects of different definitions. Changing the pandemic definition can sometimes greatly alter the estimated probability of a pandemic, illustrated by the gap between the blue and red lines, which give the pandemic probability under the 2-region transregional definition, and the 10-region transregional definition, respectively. The effect of changing the travel rate also differs substantially between these two definitions. There are also situations where changing the definition does not significantly change the pandemic probability, as seen in the left hand plot in both figures. Comparing the blue line and the dashed yellow line, which represents the 2-region

interregional definition, both the estimated risk and the effect of reducing travel are very similar in either case. So, while some changes in definition do not cause a large change in the estimated risk of a pandemic, others may significantly alter both the absolute estimated risk and the predicted effects of key parameters. In the full network, this can depend on the values of those key parameters themselves, as seen in figure 3.2. For low travel rates, the pandemic probability is very different for the two illustrated definitions, but at high travel rates the pandemic probabilities for the two definitions converge.

## Cross-immunity 3.2

In chapter 2 we saw that cross-immunity can alter the effect of between-region travel on pandemic risk. To model cross-immunity, I assumed that a pathogen strain $X$ circulated initially, followed by a strain $Y$. Those who were infected with $X$ and subsequently recovered had infection-reducing cross-immunity $\alpha$ to $Y$. Here I make use of the same assumptions, and use the Markov chain modelling framework as introduced in section 2.3, but focus on how the results can be interpreted under different pandemic definitions. As in chapter 2, I assume that at least one epidemic of $X$ occurs, and that we are primarily interested in the outcome of the outbreak of $Y$.

The effects of varying between-region travel rates under two different values of the cross-immunity $\alpha$ are shown for a range of transregional pandemic definitions in figure 3.5, for $R_0^X = 2$ and $R_0^Y = 3$. The topmost rows of these plots are equivalent to the leftmost columns of the pandemic probability plots in figures 2.5 and 2.6, in which the 10-region transregional definition was used.

Comparing figure 3.5 to figure 3.1, we see that, at moderate cross-immunity $\alpha = 0.5$, the effect of pandemic definition on pandemic probability is similar to the case without cross-immunity, the exception being that the epidemic probab-

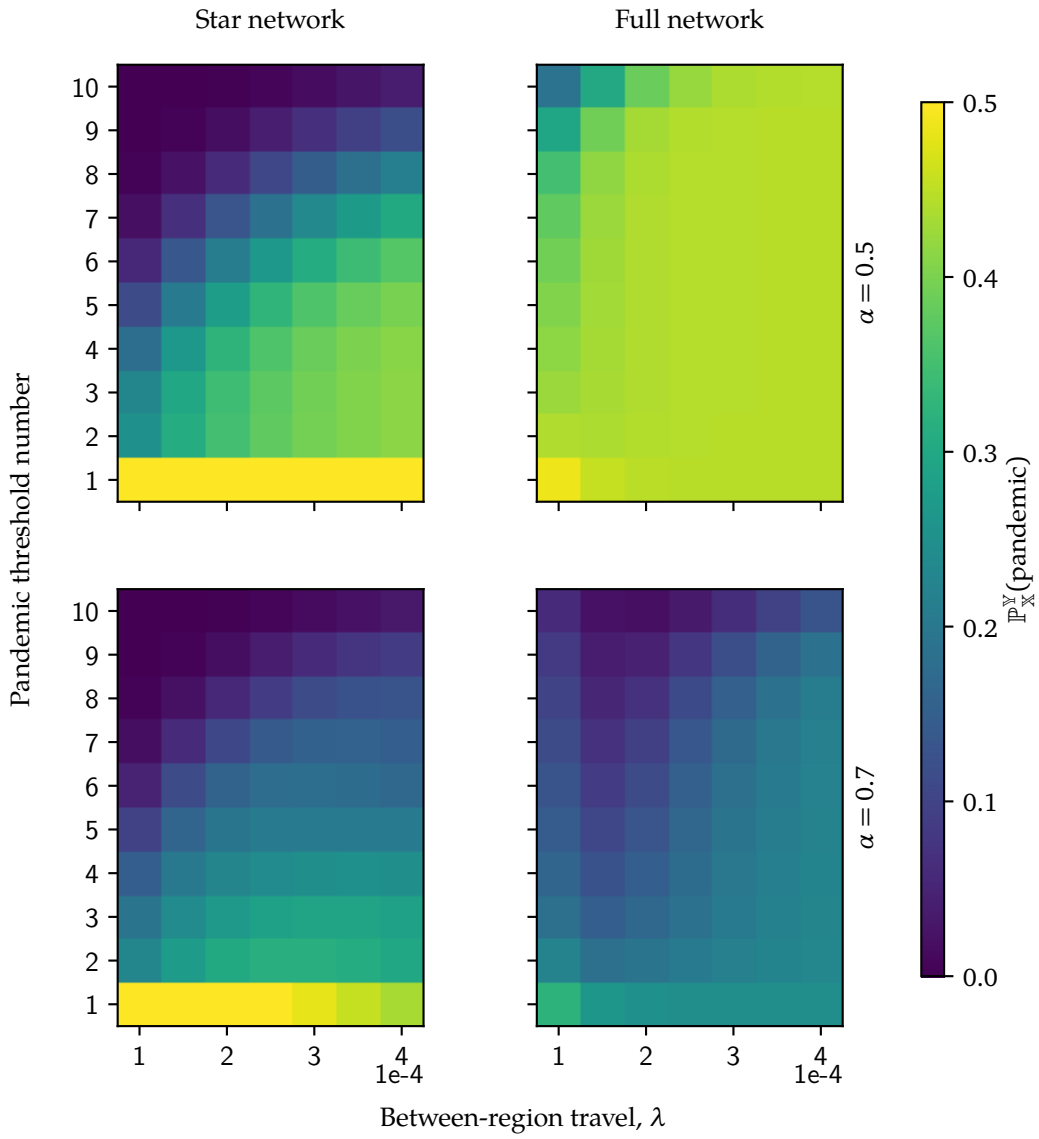



**Figure 3.5** Plots showing how the probability $\mathbb{P}_X^Y(pandemic)$ of a pandemic of strain $Y$, given at least one epidemic of strain $X$, varies with travel rate and cross-immunity, under a range of transregional pandemic definitions, in the 10-region star and full networks. At moderate cross-immunity, the plots resemble those in figure 3.1, though the pandemic risk is increased for the star network due to the higher value of $R_0^Y$, and decreased in the full network due to wide circulation of $X$. At high cross-immunity, pandemic probability is in general highly suppressed, and there are non-monotonic effects in $\lambda$.

ility (threshold $n$=1) decreases with increasing between-region travel in the full network. Overall probability is increased in the star network, since $R_0^Y = 3$ is higher than the value of $R_0 = 2$ used in figure 3.1, and decreased in the full network, due to the wide circulation of $X$ and subsequent immunity to $Y$. At high cross-immunity $\alpha = 0.7$ we see the return of the non-monotonic effects observed in chapter 2, though now we can see the dependence on the pandemic definition. In figure 2.6, which uses the 10-region transregional definition, pandemic probability appears monotonic in $\lambda$ for the star network when $\alpha = 0.7$. However, when we include lower threshold definitions, as in the lower left panel of figure 3.5, we see non-monotonic effects of the travel rate on pandemic probability in the star network when $\alpha = 0.7$. By contrast, non-monotonic effects exist at all thresholds greater than 1 in the full network (see the lower right panel of figure 3.5), although they are accentuated at high thresholds.

In figure 3.6 I show the effect of cross-immunity on pandemic risk under a range of transregional definitions when the travel rate $\lambda$ is fixed at $2 \times 10^{-4}$/day. In the full network there is a distinct transition from higher risk to lower risk, as cross-immunity approaches one. However, in the star network there is, on average, less circulation of $X$, so the effect of cross-immunity is less dramatic. Increased cross-immunity can also increase the difference in risk under different pandemic definitions—for the fully connected network, when cross-immunity exceeds $\alpha = 0.5$, differences in probability between different thresholds are visible that are much smaller at the lower values (including zero, as in figure 3.1). This suggests that the probability that an outbreak will develop into a pandemic may be more sensitive to the exact pandemic definition for outbreaks of pathogens that encounter pre-existing immunity than for pathogens which encounter only fully susceptible populations. However, this effect is not seen for the star network, in which the low connectivity of the network results in larger differences in probability between different thresholds even at low levels of cross-immunity.

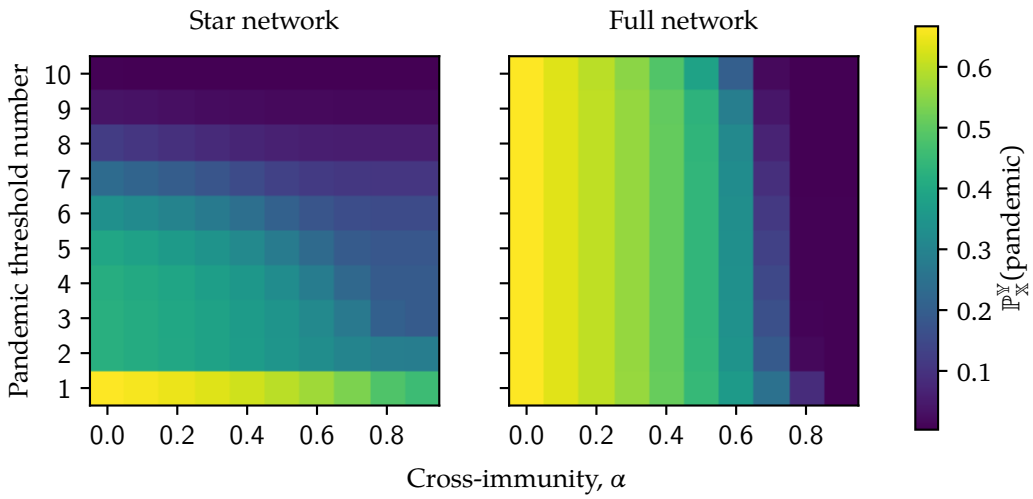



**Figure 3.6** Plots showing how the probability $\mathbb{P}_X^Y(pandemic)$ of a pandemic of strain $Y$, given at least one epidemic of strain $X$, varies with cross-immunity under various transregional definitions. In the star network, the pandemic probability is sensitive to changes in definition at all levels of cross-immunity. In the full network the sensitivity of the pandemic probability to the pandemic definition used increases with cross-immunity, until the probability of any epidemic becomes very low.

We can now return to the second of the questions introduced on page 63, "Does a portion of the population have pre-existing immunity, and does this affect the risk of a pandemic?" The presence of cross-immunity can significantly alter the effects of travel rate found in section 3.1. There we concluded that high travel rates in the full network greatly reduce the differences in risk under different pandemic definitions—but in figure 3.6, we have seen that these differences can re-emerge when cross-immunity is high. We have also seen the emergence of definition-dependent non-monotonic effects in pandemic risk at high cross-immunity in figure 3.5. In general, the presence of cross-immunity in the population greatly complicates the effects of both travel and pandemic definitions on estimated pandemic risk.

## Heterogeneous transmission 3.3

A topic of great concern during a pandemic is the heterogeneity in risk between different countries or regions [57, 170]. Cross-immunity can create one kind of heterogeneity, since it is common for previous exposure to a pathogen to differ between regions [142]. Another kind of heterogeneity is that due to different public health interventions. In this section I examine heterogeneous 10-region networks, in which five regions have a higher rate of pathogen transmission than the other five. This can be thought of as an approximation to the difference between poor regions with a relative lack of public health interventions, and wealthy regions with well-funded public health organisations and increased access to healthcare.

I set the basic reproduction number in the higher-transmission regions to $R_0^+ = 2\rho/(1+1/h)$, and that in the lower-transmission regions to $R_0^- = 2\rho/(1+h)$, where $h$ is the heterogeneity parameter, equal to the ratio between $R_0^+$ and $R_0^-$. If half of the regions have $R_0 = R_0^+$ and half have $R_0 = R_0^-$, then the mean basic reproduction number is $\rho$, where the mean is calculated over regions with equal population sizes. The larger the value of $h$, the greater the difference in $R_0$ between the two types of regions. When $\mu$ is fixed, this reflects a difference in the transmission rate. In the case of the 10-region star network, there are two ways to split the region into two equal groups—the central region can either have $R_{0,\text{centre}} = R_0^+$, or $R_{0,\text{centre}} = R_0^-$. In both networks, if an outbreak is seeded by a single infected individual in a single region, then that region can either have $R_{0,\text{seed}} = R_0^+$, or $R_{0,\text{seed}} = R_0^-$.

The variation of the pandemic probability with the heterogeneity $h$, in each of these cases, under a range of transregional pandemic definitions, is shown in figure 3.7, with $\rho = 2$ and $\lambda = 2 \times 10^{-4}$/day. In the case of the star network, when the chosen value of $R_{0,\text{centre}}$ is different from $R_{0,\text{seed}}$, this means that the pathogen must emerge in a peripheral region, which significantly impacts the probability of a pandemic. When $R_{0,\text{centre}} = R_{0,\text{seed}}$, the pathogen could emerge in the central re-

gion, but for consistency, I require that the pathogen always emerge in a peripheral region for the star network—both when $R_{0,\text{centre}} = R_{0,\text{seed}}$ and when $R_{0,\text{centre}} \neq R_{0,\text{seed}}$.

The row for the 1-region definition shows how the risk of an epidemic varies with the changing basic reproduction number of the pathogen in the region in which it emerges. If $R_{0,\text{seed}} = R_0^-$, then increasing heterogeneity reduces the chance of an epidemic. In this case, as the threshold number increases, the pandemic probability decreases, but the relationship to heterogeneity remains similar—there is a monotonic decrease in probability with increasing $h$.

More complex effects of heterogeneity occur when $R_{0,\text{seed}} = R_0^+$. For a threshold number of 1 the epidemic risk increases monotonically with $h$. However, as the threshold increases, then more and more lower-transmission regions are required to experience epidemics in order for a pandemic to be deemed to have occurred. This can change the relationship with heterogeneity, as increasing $h$ leads to a greater initial chance of an epidemic, but a reduced chance of epidemics in lower-transmission regions. In the case of the star network, any threshold greater than 1 requires the central region to experience an epidemic. Either the outbreak is seeded in the central region, or the central region is the second region to experience an epidemic. When $R_{0,\text{centre}} = R_0^-$, this leads to a dramatic decrease in pandemic probability with increasing threshold number at high heterogeneity. This, in turn, leads to the non-monotonic effects that we see for thresholds 3 and 4, in the top right panel of figure 3.7—greater heterogeneity at first increases pandemic probability by facilitating the initial emergence of the pathogen, but then, as it gets higher, blocks the transmission of the pathogen around the network.

A similar effect is seen in the full network when the threshold number is 10. Greater heterogeneity has a dual effect on the chance of a pathogen that emerges in a higher-transmission region causing an epidemic in a lower-transmission region. Increased $R_0$ in the higher-transmission regions leads to a greater number of epidemics with large final sizes in those regions, increasing the risk of importation

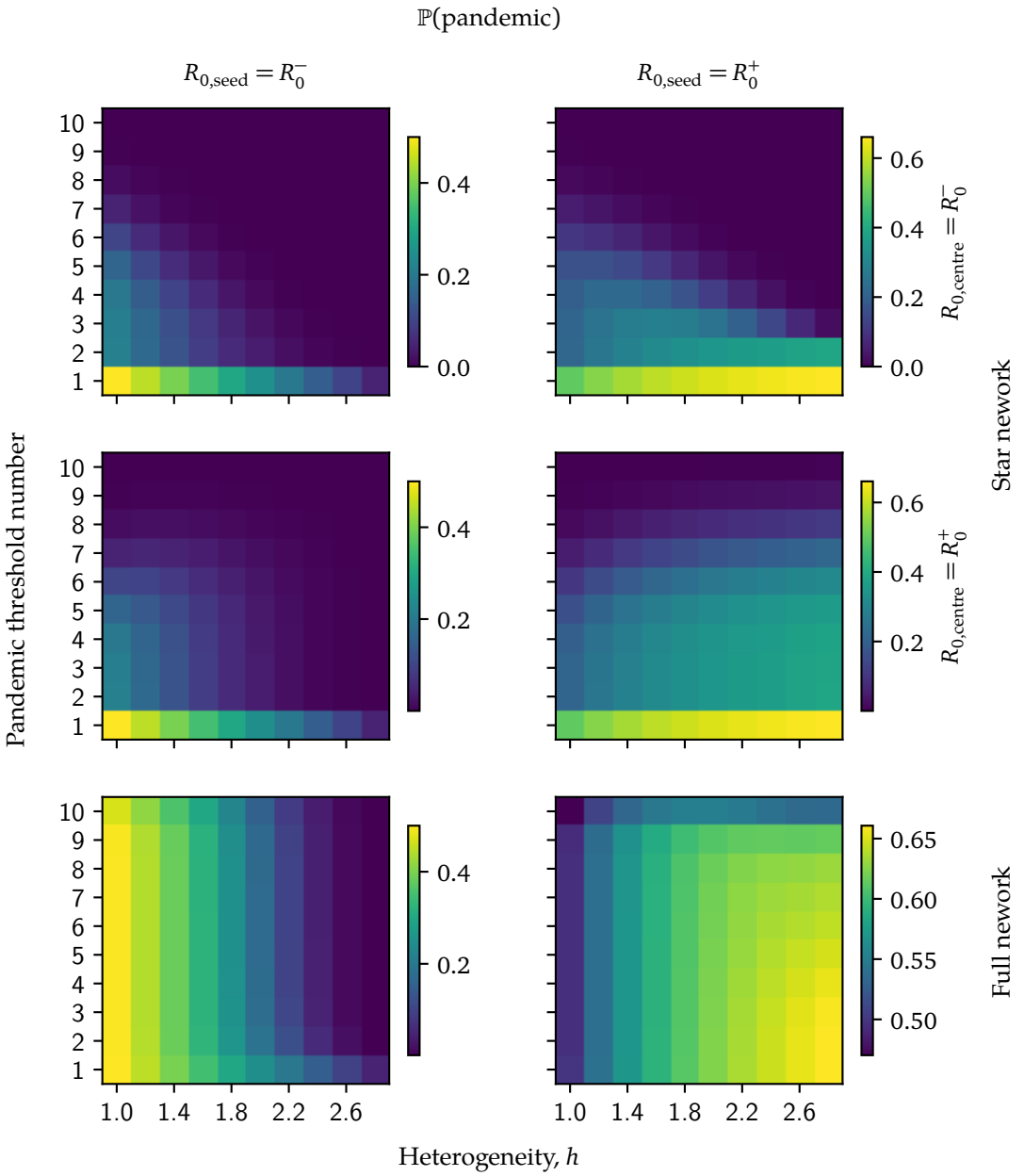



**Figure 3.7** Plots showing how pandemic probability varies with heterogeneity $h$ in transmission rates, for different networks and initial states, under a range of different transregional pandemic definitions. When the outbreak is seeded in a lower-transmission region, increased heterogeneity always results in reduced pandemic probability. When the outbreak is seeded in a higher-transmission region, the effect of heterogeneity on pandemic probability depends on the network and on the pandemic definition, with non-monotonic effects occurring in the star network with $R_{0,centre} = R_0^-$, and in the full network when the pandemic threshold number is high. Each plot in this figure has a different colour scale, to illustrate the relative change in pandemic risk due to variations in $h$ and the pandemic threshold. For comparisons in risk between different networks, this figure and the preceding heat map figures are reproduced in appendix A with a single consistent colour scale.

of the pathogen to the lower-transmission regions. However, the reduced $R_0$ in the lower-transmission regions decreases the probability that an imported pathogen will cause an epidemic. For a threshold number of 10, when all lower-transmission regions are required to experience epidemics, the risk-increasing effects of increasing heterogeneity win out at first, but when $h$ exceeds 2, the risk-decreasing effects dominate.

For the star network with $R_{0,\text{centre}} = R_0^+$, pandemic probability is monotonic with $h$ for all pandemic definitions. This is because increasing $h$ always increases the ability of the central region to export the pathogen to the periphery.

The third question introduced on page 63 was, "How is the risk of a pandemic affected by differences between regions?" Many pathogens have higher transmission rates in lower income countries, and novel pathogens are more likely to emerge in low income countries [57, 134, 170]. This means that pathogens are most likely to emerge in countries in which they have higher transmission rates. In figure 3.7, we have seen that pandemic risk due to pathogens emerging in high-transmission regions can have a complex relationship with heterogeneity, depending on the pandemic definition. In particular, when the pandemic definition requires many countries to experience epidemics to qualify an outbreak as a pandemic, including countries with lower transmission rates, we can see non-monotonicity in the relationship between heterogeneity and pandemic probability. These results imply that, when the mean value of the transmission rate is fixed, a small gap in the effectiveness of public health infrastructure between wealthy and poor regions puts all regions at greater risk; but that a larger gap protects wealthier regions while the risk for poor regions continues to increase.

To illustrate this concept, consider the contrasting examples of Ebola and COVID-19. The 2014 outbreak of Ebola virus followed a pattern of high incidence in low-income countries but low incidence in high-income countries. The virus spread through several low-income African countries but was effectively

contained when introduced to high-income countries [69, 171, 172]. In this case, high-income countries had the capacity to prevent a pandemic from taking hold, being able to quickly isolate and treat symptomatic individuals. This generated high heterogeneity in transmission, corresponding to the $R_{0,\text{seed}} = R_0^+$, high-$h$ regime in figure 3.7, with low-income countries at high risk and high-income countries at low risk. In contrast, high-income countries have not been able to escape the pandemic of COVID-19, in part due to asymptomatic and presymptomatic transmission of SARS-CoV-2 allowing it to evade surveillance and public health measures [173, 174]. This has led to more similar transmission rates between different countries, corresponding to the low-$h$ regime in 3.7. For the fully connected network, this means that the probability of epidemics is more uniform between regions, and therefore that the probability of a pandemic is more uniform between pandemic definitions.

Since the initial stages of the COVID-19 pandemic, the conditions under which SARS-CoV-2 spreads have changed significantly. Wealthy regions now have much higher levels of vaccination than poorer regions—while 56% of the global population has received at least one dose of a SARS-CoV-2 vaccine as of December 2021, only 7% of people in low-income countries have received a dose [175, 176]. In addition, different regions have experienced outbreaks of different sizes, leading to variations in immunity against re-infection. Taking into account the multiple strains of SARS-CoV-2 furthers this heterogeneity, with different regions of the world having had different levels of exposure to different strains [177]. The spread of new variants of SARS-CoV-2 now occurs in a highly heterogeneous landscape. This could partially explain why some variants spread to only a few regions, while others spread globally.

# Conclusions ⬡ 3.4

In the absence of a widely accepted quantitative definition of a pandemic, a variety of different sorts of outbreaks could be classified as pandemics. As we have seen, various epidemiological parameters can have different effects on the risk of different kinds of outbreaks as characterised by the interregional and various transregional pandemic definitions. This means that, even when we can determine the effects of factors like travel rates and immunity on the risk of certain outbreak scenarios, the effect of these parameters on pandemic risk remains ambiguous due to the ambiguity of the definition of 'pandemic'.

In sections 3.2 and 3.3, we saw that quantitative differences in the pandemic definition, such as using transregional definitions with different thresholds, can result in qualitative differences in the predicted effects of cross-immunity and heterogeneity in the susceptible population. In figure 3.5 there is a marked qualitative shift from monotonic to non-monotonic dynamics at high cross-immunity in the star network, and in figure 3.7 we saw that pandemic risk can change from increasing with heterogeneity to decreasing with heterogeneity depending on the pandemic definition. This shows that two studies of pandemic risk using exactly the same model could have qualitatively different conclusions, if the pandemic definitions used were different. Because research on the effects of parameters like travel rate, cross-immunity, and heterogeneity help determine public health policy, these different definitions could lead to contradictory policy advice.

The epidemiological parameters explored in this chapter are only a subset of the factors that could be included when modelling pandemic risk. Additional parameters could incorporate the effects of different types of travel (e.g. within-country transport or international flights), nosocomial infections, or age structure [65, 178–181]. Metapopulation models are widely used in pandemic modelling, and the overall conclusions of this chapter should hold irrespective of the under-

lying modelling framework [83, 91, 102, 103, 182]. However, the specific interpretation of alternative models under various pandemic definitions might differ in interesting ways, and is a promising avenue for future work. Of particular interest are metapopulation models with additional epidemiological complexity [65, 103, 183, 184] and the widely used global epidemic and mobility (GLEaM) model [185–187]. In a model with explicit representation of time, the effect of pandemic definition on questions to do with the timescale of pandemics could be resolved, such as the time taken for a local epidemic to develop into a pandemic.

Contradictions due to vague or inconsistent pandemic definitions have caused controversy in the past. Although differences between pandemic definitions are rarely addressed within the scientific community, they can cause confusion between different stakeholders (e.g. between scientists and governments, or governments and the public), who may not have a shared background understanding of the term. In 2009, the WHO declared a pandemic of H1N1 influenza, using criteria related to the incidence and spread of the virus in different WHO regions [111]. The criteria did not include reference to morbidity or mortality [108]. This fact led to some controversy over whether the declaration of a pandemic was appropriate, as the declaration prompted some governments to mount an intensive response to an outbreak that resulted in fewer yearly deaths than a typical strain of seasonal flu [122, 188–190].

This incident did not discourage the epidemiology community from using the term 'pandemic', despite the existence of alternative concepts such as that of a public health emergency of international concern (PHEIC) [110]. It continues to appear in phrases such as 'pandemic preparedness' [191–193], 'pandemic influenza' [150, 194, 195], and 'pandemic potential' [49, 164, 196]. A Google Scholar search returns 25,800 results using the term 'pandemic' for 2019 alone. Although the WHO no longer uses the term 'pandemic' as an official status of any outbreak [111, 112], the WHO Director-General drew attention to their use of the term as

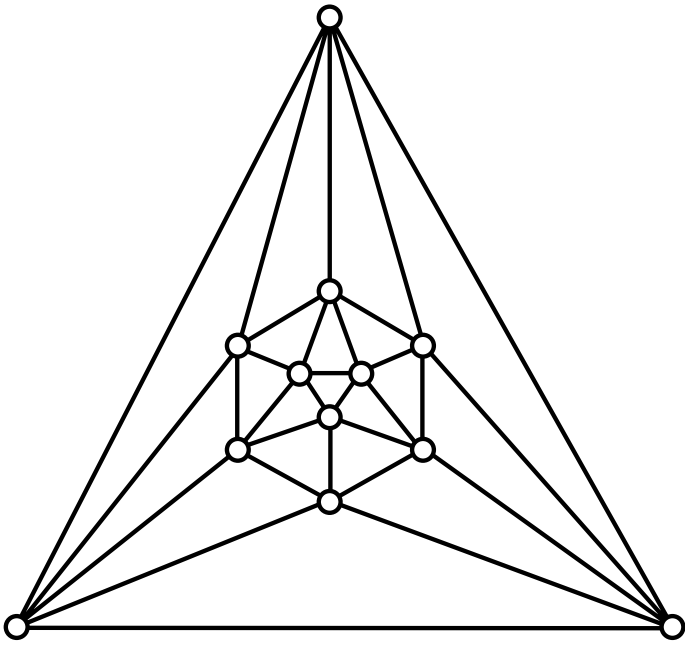
recently as March 2020, to describe the status of the COVID-19 outbreak [197]. The Director-General cited "alarming levels of inaction" as one of the reasons to use the term, along with the caveat that "describing the situation as a pandemic does not change WHO's assessment of the threat posed by this virus". Although the spread of COVID-19 had already been designated a PHEIC, the WHO's use of the term 'pandemic' was of interest to the public, receiving extensive press coverage [198–200], and to governments, being quoted in legal responses to the pandemic [201]. The term 'pandemic' clearly continues to be important in communications about infectious disease risk, especially in attempts to encourage a large-scale public health response.

If giving up the use of the term 'pandemic' is not an option, then it is important to resolve its ambiguity in some way. A single quantitative definition applied in all contexts is unlikely to be successful. There can be significant uncertainty about infectious disease dynamics, especially in early stages of outbreaks, meaning that the existence of a specific numerical criterion is unlikely to resolve all ambiguity in pandemic declarations [202]. Disease outbreaks are also dynamic, meaning that outbreaks that fail to meet a criterion now could do so in the future, and vice versa.

The work presented in this chapter provides a foundation for an alternative approach, in which pandemic preparedness researchers using quantitative methods carefully consider the appropriate pandemic definition in a given context and state their assumptions clearly. Under this approach, the different kinds of outbreaks that fall under the category of 'pandemic' needn't be absorbed into a single quantitative definition. The ambiguity is instead resolved on a case-by-case basis for the purpose of communicating the conclusions of a given modelling study of pandemic risk. I have demonstrated that these conclusions are more sensitive to the pandemic definition in certain situations, especially those in which different regions experience differences in transmission due to immunity or public health interventions. This provides an initial set of guidance as to how precise a defi-

nition is needed in different kinds of modelling studies. A more comprehensive set of guidelines could be developed by exploring the additional parameters and model types discussed above.





A graph whose vertices represent the vertices of the regular icosahedron, and whose edges represent the connections between those vertices by edges of the icosahedron. A graph is a mathematical object which represents entities (vertices) between some of which there exist pairwise relationships (edges). This graph is highly symmetric, due to the symmetry of the icosahedron.

# 4 Spatial control of pathogens

We saw in chapter 3 that pandemics can be characterised as outbreaks with wide geographical extension. This means that strategies to reduce the spread of pathogens over geographic space could help prevent pandemics. In this chapter I will compare a variety of strategies for spatial control of an invading pathogen, based on the allocation of prophylactic measures in a metapopulation. I use a general model of prophylaxis, meaning that my results are applicable to many different kinds of disease prevention, including vaccines and prophylactic drugs.

Administrators in charge of prophylaxis campaigns often find that there are fewer doses of the prophylactic drug or vaccine available (or capacity to administer these doses) than there are susceptible individuals willing to receive them. In this situation, some form of prioritisation is necessary, in which certain individuals are chosen from the susceptible population to receive prophylaxis. Often, vaccines and drugs are offered to individuals in certain demographics who are particularly at risk of catching, spreading, or experiencing severe outcomes of a disease. This kind of demographic targeting is used by the United Kingdom's National Health Service to distribute influenza vaccines—the NHS offers vaccines free of charge to individuals over 65, or with conditions such as asthma and diabetes, who are at risk of severe influenza; and a nasal vaccine to children, who play an important role in the spread of influenza [203]. Pre-exposure prophylaxis for HIV (PrEP) is often offered preferentially to high-risk groups such as men who have sex with men [204–206]. Targeting demographics at high risk of severe

disease has also been a key strategy in COVID-19 vaccination programmes, with elderly people often being prioritised [64, 207, 208].

An alternative way of selecting which individuals receive prophylaxis is by virtue of their location in geographic space. Spatial position can be used as an alternative to demographic membership as a predictive variable for risk of infection [68], but often spatial strategies are used when cases of a disease occur only within a certain region, and public health decision-makers seek to contain the pathogen locally and prevent it from spreading to other regions [72, 87, 209, 210]. It is this kind of control that is particularly relevant for pandemic prevention when seeking to contain a PPP [86].

Within spatial strategies, there are many ways in which certain regions can be prioritised for prophylaxis. Among alternatives, priority could be given to regions close to an existing epidemic, with high population density, or with many travel connections. In practice, many modelling studies that consider spatial prophylaxis allocation use a risk-targeting approach, in which those regions predicted to be most at risk of importation of a pathogen are prioritised for distribution of the drug or vaccine [68, 72, 86, 117]. There is, however, no proof that this leads to the maximum reduction in infections, and authors often mention possible further work examining "downstream effects" [117] or "spreading capacity" [99] of certain regions, or the "impact of vaccination on ... spatial expansion" [68].

Here I define a prophylaxis strategy as a method for sorting all individuals susceptible to an infection, from highest priority to receive the drug or vaccine to lowest priority. I define a prophylaxis allocation as a division of those individuals into two sets—those in one set receive the drug or vaccine, and those in the other do not. I define risk-targeting as any prophylaxis strategy in which a region is prioritised for receipt of drugs or vaccines based solely on the risk that a large outbreak of the relevant pathogen will occur in that region. The method for determining this risk can range from simple heuristics to sophisticated epidemiological

forecasts [72]. One particularly widely discussed form of risk-targeting is spatial ring vaccination (as opposed to ring vaccination of contacts), in which individuals are prioritised for vaccination based on their proximity to an ongoing outbreak. This kind of strategy has been discussed in settings from foot-and-mouth disease in the United Kingdom to Ebola in the Democratic Republic of the Congo [87, 209].

An alternative to risk-targeting in strategies for prophylaxis distribution is centrality-targeting [211–213]. Centrality is a term for any measure of the importance of a node in a network. Some of these measures can be used as proxies for the spreading capacity of a given region. I discuss specific measures of centrality in more detail below. In centrality-targeting strategies, drugs or vaccines are preferentially allocated to regions with high centrality in a travel network. Thus, they may act to impede the circulation of a pathogen between regions, rather than directly protecting those regions most likely to experience outbreaks themselves.

There is little existing work directly comparing risk-targeting of prophylaxis to centrality-targeting. A study by Piraveenan et al. (2013) compares targeting of prophylactic control in a disease transmission network according to three different measures: hop distance (a simple kind of risk-targeting), percolation centrality, and betweenness centrality [211]. Hop distance is the length of the shortest path through the network from a susceptible node to the closest infected node, and can be treated as a proxy for risk. Betweenness centrality is a very commonly used centrality measure based on a node's connectivity. Percolation centrality is a measure of node centrality of the authors' own devising, which takes into account a node's connectivity as well as the current state of infection in the network. The authors find that, in a scale-free and a random network, which of the three strategies best reduces infections depends on parameters reflecting how widespread the pathogen is, and how many nodes can be immunised. In general, when the pathogen is scarce in the network, or when many nodes can be immunised, risk-targeting (that is, prioritisation according to hop distance), is most effective at

reducing infection. In the opposite scenario, when the pathogen is widespread or when few nodes can be immunised, betweenness-targeting is most effective. In the intermediate regime, targeting according to percolation centrality leads to the largest mean reduction in infections.

Generalising from this result, I will start from the conjecture that risk-targeting may be less effective than centrality-targeting when resources available for control of a pathogen are very limited (e.g. few doses of a vaccine), or when the pathogen concerned has a high spreading potential (either by having a high $R_0$ or by travelling widely across space). Piraveenan et al.'s findings are intuitive—for instance, when many resources are available for control, a pathogen can be fully contained by ring vaccination (a form of risk-targeting), while the only chance to contain a pathogen when few resources are available is by targeting travel hubs (a form of centrality targeting). This intuitive understanding of the scenario leads to the following generalised hypothesis:

- when resources for control are widely available, and the potential for widespread pathogen transmission is limited, targeting regions at high risk is more effective than targeting regions with high centrality;
- when resources are limited and the pathogen has the potential to spread widely, targeting regions with high centrality is more effective than targeting regions at high risk.

The particular sense of 'effective' will depend on the control objectives. To investigate this hypothesis, I examine below whether the risk-targeting strategy is the most effective strategy to reduce the overall number of infections in a range of different epidemiological networks. I compare this strategy to two variations on a centrality-targeting strategy that prioritises regions for prophylaxis according to their centrality on the human travel network. The variations come about due to alternative centrality measures.

# Centrality 4.1

Certain nodes in a travel network can have a particularly important role in the dynamics of a pathogen's spread across a metapopulation. For instance, in the star network, we saw that the central node plays an outsized role in the disease dynamics compared to the peripheral nodes (see section 3.3). In network science, a measure of the importance of a node due to its connectivity in a network is called a *centrality* measure. In general, if a node is important to the disease dynamics on a network, then we might expect interventions in the corresponding region to be particularly effective.

There are many possible ways to define centrality [101, 214–217]. In the case of a pathogen spreading across a travel network, a common approach is to examine the paths that the pathogen can take when spreading between different individuals or populations. One kind of centrality measure defined in this way is *betweenness centrality*.

The betweenness centrality of a node $v$ is the proportion of the shortest paths from a source node $s$ to a target node $r$ that pass through $v$, averaged over all sources and targets $s$ and $r$ [218]. In a weighted travel network, the contribution to the length of a path of one of its constituent edges is equal to the reciprocal of the weight of that edge, since higher weights represent stronger connections. The betweenness $b_v$ of node $v$ in a network $G$ of size $n$ is expressed as

$$b_v = \frac{1}{(n-1)(n-2)} \sum_{s \neq v, r \neq v} \frac{g_v^{(s,r)}}{g^{(s,r)}}, \tag{4.1}$$

where $g^{(s,r)}$ is the number of geodesics (shortest paths) from node $s$ to node $r$, and $g_v^{(s,r)}$ is the number of such paths that pass through node $v$. This definition ensures that the betweenness of a given node is always between 0 and 1. There are several ways to calculate betweenness centrality in practice—I use Ulrik Brandes' algorithm as implemented in the *networkx* Python package [219, 220].

Betweenness centrality is widely used in studies of spreading processes on networks, and as such deserves consideration when treating the role of node centrality in infectious disease modelling [221–224]. However, every centrality measure based on paths through a network makes implicit assumptions about the spreading process being modelled [225]. In the case of betweenness centrality, the spread is assumed to travel on geodesic paths, with any node being a possible origin of the path. This means that betweenness centrality may not always be suitable for determining the epidemiological importance of a node. Pathogens transmitting between regions cannot be assumed to follow the shortest path from their source to their destination. In addition, when responding to an invading pathogen, we often know the possible sources of an infection and are seeking to disrupt transmission from these sources. Below, I will show that we can resolve the first issue with *random walk betweenness centrality*, and the second issue with *percolation centrality*. Combining these two solutions will result in the *random walk percolation centrality* (RWPC) measure.

Random walk betweenness centrality is defined similarly to betweenness centrality, but expands the paths counted, from only shortest paths from node $s$ to node $r$ that pass through a given node, to all random walks that pass through the node from $s$ to $r$ [226, 227]. Suppose a random walk in an $n$-node network is at node $v \neq r$. We write the travel rate from node $v$ to node $u$ as $\lambda_{vu}$, and the $n \times n$ matrix of these values as $\Lambda$. We assume that $\lambda_{vv} = 0$ for all nodes $v$. Then the probability $p_{vu}$ for the random walk to move to an adjacent node $u$ on its next step is:

$$p_{vu} = \frac{\lambda_{vu}}{\sum_u \lambda_{vu}}. \tag{4.2}$$

We denote the $n \times n$ matrix of these quantities $P$. Now, if $v = r$, then the random walk cannot escape because it has reached its final destination, so $p_{rk} = 0$ for all nodes $k$. We write $P_r$ to denote the $(n-1) \times (n-1)$ matrix which excludes the row and column corresponding to node $r$.

The probability that a random walk starting at node $s$ and ending at $r$ reaches node $v$ after a number of steps $m$ is given by $[(P_r)^m]_{sv}$. The probability that the walk then steps to an adjacent node $u$ is given by $p_{vu}[(P_r)^m]_{sv}$. So the expected number of times $N_{vu}$ for the random walk to step from node $v$ to node $u$ is given by:

$$\begin{aligned} N_{vu} &= \sum_{m=0}^{\infty} p_{vu}[(P_r)^m]_{sv} \\ &= p_{vu}\left[(\mathbb{1} - P_r)^{-1}\right]_{sv} \\ &= \lambda_{vu}\left[(W - \Lambda_r)^{-1}\right]_{sv}, \end{aligned} \tag{4.3}$$

where $\mathbb{1}$ is the $(n-1)\times(n-1)$ identity matrix, $W_{vu} = \delta_{vu}\Big/\sum_k \lambda_{vk}$, and $\Lambda_r$ is $\Lambda$ with the $r^{\text{th}}$ row and column set to zero.

Since a random walk can have an arbitrarily high number of steps, we allow steps passing along an edge in opposite directions to cancel out, so that we ignore random walks that go back and forth through a given node multiple times without progressing to another part of the network. We therefore define the net flow $F_e^{(s,r)}$ through an edge $e = \{v, u\}$ as the absolute difference between the flows either way along the edge,

$$F_e^{(s,r)} = |N_{vu} - N_{uv}|. \tag{4.4}$$

Writing this quantity in terms of the Laplacian of the network, $L = (W - \Lambda)$, will allow much faster computations of the random walk centrality. However, $L$ is singular, meaning that we cannot simply invert it and substitute it into equation 4.3. Instead we write:

$$C = \begin{pmatrix} 0 & 0_{n-1,1} \\ 0_{1,n-1} & \tilde{L}^{-1} \end{pmatrix}, \tag{4.5}$$

where $0_{k,m}$ is the $k \times m$ zero matrix, and the *reduced Laplacian* $\tilde{L}$ is an $n-1 \times n-1$ matrix obtained from $L$ by omitting the first row and column, yielding an invertible matrix. When $r = 1$, $F_e^{(s,r)}$ can easily be expressed in terms of $C$. In this case,

$N_{vu} = \lambda_{vu} C_{sv}$. When $r \neq 1$, we can still express $F_e^{(s,r)}$ in terms of $C$ if we continue to treat node 1 as a sink for the purpose of defining $P_1$, but take away any contribution to $N_{vu}$ from paths that emanate from node $r$. This cancels out any contributions to $N_{vu}$ from $p_{rk}$ for any node $k$, setting node $r$ as a sink in practice. That is,

$$F_e^{(s,r)} = |\lambda_{vu} C_{sv} - \lambda_{vu} C_{rv} - \lambda_{uv} C_{su} + \lambda_{uv} C_{ru}| . \tag{4.6}$$

Paths that pass first through node 1 and then along edge $e$ on the way from node $s$ to node $r$ are now included as a negative term in $N_{uv}$, rather than as a positive term in $N_{vu}$. The contribution to $F_e^{(s,r)}$ is equivalent.

The net flow $F_v^{(s,r)}$ through a node $v$ is equal to half the sum of flows along each of its connected edges, giving:

$$\begin{aligned} F_v^{(s,r)} &= \sum_{e:v\in e} \frac{1}{2} |\lambda_{vu} C_{sv} - \lambda_{vu} C_{rv} - \lambda_{uv} C_{su} + \lambda_{uv} C_{ru}| \\ &= \sum_{e:v\in e} \frac{1}{2} |M_{es} - M_{er}| \end{aligned} \tag{4.7}$$

where $M = BC^T$, and the oriented incidence matrix $B$ is defined:

$$B_{ev} = \begin{cases} \lambda_{vu} & \text{if } e = \{v,u\} \\ -\lambda_{vu} & \text{if } e = \{u,v\} \\ 0 & \text{otherwise.} \end{cases} \tag{4.8}$$

Finally, we define the random walk centrality $b_v^{rw}$ in terms of $F_v^{(s,r)}$ as:

$$b_v^{rw} = \frac{1}{(n-1)(n-2)} \sum_{s\neq v, r\neq v} F_v^{(s,r)}, \tag{4.9}$$

Expressing $F_v^{(s,r)}$ in terms of the oriented incidence matrix and the Laplacian allows this quantity to be calculated using standard matrix functions. Using random walk betweenness rather than betweenness based on geodesic paths means that

inefficient or winding routes through the network can contribute to the centrality measure. Since pathogens transmit randomly, this is more representative of the dynamics of pathogen spread on a travel network.

In order to incorporate knowledge about the origin of a pathogen into our centrality measure, we can use percolation centrality. Percolation centrality introduces a weighting to each element of the sums in equations 4.1 and 4.9 depending on the infection state of the source node $s$ [211]. The percolation centrality $p_v(t)$ of node $v$ at time $t$ is:

$$p_v(t) = \frac{1}{(n-2)} \sum_{s \neq v, r \neq v} \frac{g_v^{(s,r)}}{g^{(s,r)}} \frac{x_s(t)}{\left[\sum_i x_i(t)\right] - x_v(t)}, \tag{4.10}$$

where $x_i(t)$ is a scalar quantity representing the infection state of node $i$ at time $t$. If the infection state is binary (e.g. a node either belongs to state $I$ or a non-infected state such as $S$ or $R$) then equation 4.10 can be interpreted as equivalent to equation 4.1, with the source node $s$ restricted to the class of infected nodes, and the quantity normalised appropriately. This successfully captures the importance of paths starting from an infected node in a disease transmission network.

Combining the modifications to betweenness centrality introduced in equations 4.9 and 4.10 gives us random walk percolation centrality. The random walk percolation centrality $p_v^{rw}(t)$ of a node $v$ at time $t$ is:

$$p_v^{rw}(t) = \frac{1}{(n-2)} \sum_{s \neq v, r \neq v} F_v^{(s,r)} \frac{x_s(t)}{\left[\sum_i x_i(t)\right] - x_v(t)}. \tag{4.11}$$

Note that this equation is to equation 4.9 as equation 4.10 is to equation 4.1. If $F_v^{(s,r)}$ is defined according to equation 4.7 above, then this definition of random walk percolation centrality expands on that due to Piraveenan et al. by allowing the underlying network to have weighted, directed edges [211].

# A simple metapopulation model 4.2

In section 2.4 I note that the Markov chain modelling framework introduced in chapter 2 is well suited to theoretical investigations on small networks, but would be computationally unwieldy on larger empirical networks. Below, I evaluate infectious disease control on a range of artificial and empirical networks, requiring a pathogen transmission model that can be applied more easily to larger networks. To this end I make use of a simplified stochastic SIR metapopulation model similar to others used in spatial analyses of infectious disease epidemics and their control [68].

As in chapters 2 and 3, each node $i$ on the travel network is associated with a region, which has its own population size $N_i$ and within-region basic reproduction number $R_{0,i}$. The expectation value for the number of individuals travelling from region $i$ to region $j$ per unit time is again denoted $\lambda_{ij}$. Like all rates referred to in this section, this the rate of a stochastic process, where the number of individuals travelling over a time period $\Delta t$ is given by a Poisson distribution with rate parameter $\lambda_{ij}\Delta t$.

With this information we can calculate the expected number of infected individuals travelling from a region $i$ that experiences an epidemic, to any other region connected to it on the travel network. First we calculate the total number of infected individuals over the course of the epidemic in region $i$, $R_i(\infty)$, writing the final size equation (see equation 1.8) as:

$$R_i(\infty) = N_i - N_i \exp\left(\frac{-R_{0,i}R_i(\infty)}{N_i}\right). \tag{4.12}$$

Each infected individual can either recover while still in region $i$, or travel to a neighbouring region before recovering. The probability $P_{ij}$ that an individual

travels to a particular neighbouring region $j$ while infected is therefore

$$P_{ij} = \frac{\lambda_{ij}}{\sum_k \lambda_{ik} + \mu}, \tag{4.13}$$

where $\mu$ is the recovery rate, and $\sum_k \lambda_{ik}$ is the total rate of travel out of region $i$. This means that the expected total number $I_{ij}$ of infected individuals who arrive in region $j$ from region $i$ over the course of the epidemic in $i$ is

$$I_{ij} = \frac{\lambda_{ij} R_i(\infty)}{\sum_k \lambda_{ik} + \mu}. \tag{4.14}$$

When $\mu \gg \sum_k \lambda_{ik}$, meaning that individuals are much more likely to recover than travel while infectious, equation 4.14 can be approximated as:

$$I_{ij} \approx \frac{\lambda_{ij}}{\mu} R_i(\infty). \tag{4.15}$$

This approximation is motivated in part by the fact that, for the metapopulation model to be appropriate, the time-scale of the dynamics within each sub-population must be much faster than the time-scale of mixing between the populations. If the two time scales were similar, it would not be appropriate to represent the sub-populations as separate systems. I also make use of this approximation to assume that the population size $N_i$ of each region does not change over the course of the outbreak.

To avoid highly intensive computations, I use a simplified model with a square epidemic curve—meaning that, in this model, there are a constant number of infected individuals during an epidemic, and no infected individuals otherwise. I assume that the duration of such an epidemic is drawn randomly from an exponential distribution, but that the number of infected individuals is deterministic and independent of the epidemic duration. Since the size of the epidemic does not vary from its start to its end, this means that we can shift from treating the states

of individuals, to treating the states of whole regions. We can treat each region as belonging to a state $S$, meaning that no epidemic has occurred; $I$, meaning that an epidemic is ongoing; or $R$, meaning that an epidemic has occurred and the region is protected from further outbreaks due to herd immunity.

The expected total number of infected individuals who arrive in region $j$ from a region $i$, which experiences a square epidemic curve with height $A$ and mean duration $D$, is given by $\lambda_{ij}AD$. This quantity determines the probability for the pathogen to spread between regions, so its value in this simplified model should match the corresponding value in equation 4.15. To achieve this, I choose $A = R_i(\infty)/\mu$ and $D = 1$. This choice ensures that one unit of time is equal to the expected duration of a single epidemic, thus effectively rescaling time to match the simplified pathogen dynamics.

In a stochastic SIR model, the probability for a single infected individual to cause an epidemic when introduced into a large, entirely susceptible population is $1-1/R_0$ (see section 1.2). If the rate of arrival of infected individuals in region $j$ from region $i$ is $\lambda_{ij}R_i(\infty)/\mu$, and each arrival has a probability of causing an epidemic given by $1-1/R_0$, then the rate $E_{ij}$ at which the epidemic in region $i$ seeds an epidemic in region $j$ is:

$$E_{ij} = \left(1 - \frac{1}{R_0}\right)\lambda_{ij}R_i(\infty)/\mu. \tag{4.16}$$

This completes the structure of this simplified epidemic model: regions are assigned states $S$, $I$, and $R$; regions in state $I$ export the epidemic stochastically at rate $E_{ij}$ to neighbouring regions in state $S$; and regions in state $I$ recover stochastically into state $R$ at rate $1/D = 1$, after which they cannot be reinfected. I assume that the time between events is exponentially distributed, and use Gillespie's direct stochastic method to generate simulations of this model [77].

# Prophylaxis strategies ⬡ 4.3

I focus on the case of prophylactic measures that reduce the total number of individuals who are susceptible to infection—rather than, say, reducing the likelihood of severe outcomes of a disease, or reducing onward transmission without reducing susceptibility. The paradigmatic example of prophylaxis that works in this way is a single-dose vaccine with 100% efficacy and no waning immunity. Other susceptibility-reducing interventions include PrEP, which reduces the susceptibility to HIV infection by close to 100% for individuals with high adherence to an appropriate drug regimen [228, 229]. In the case of agricultural pathogens, culling of animals or inducing resistance in plants both reduce the number of susceptibles in the population. In what follows, a "dose" of prophylaxis will be considered to remove an individual from the susceptible class indefinitely—so each dose represents a single vaccination, an indefinite prescription for a prophylactic drug, an animal being culled, or any other measure that removes a potential host for the pathogen. Each of these measures is equivalent in the disease model outlined above, with each prophylaxis dose in region $i$ decrementing $S_i(0)$ by one. Therefore, to achieve herd immunity in region $i$, $N_i(1 - 1/R_{0,i})$ doses must be administered in that region.

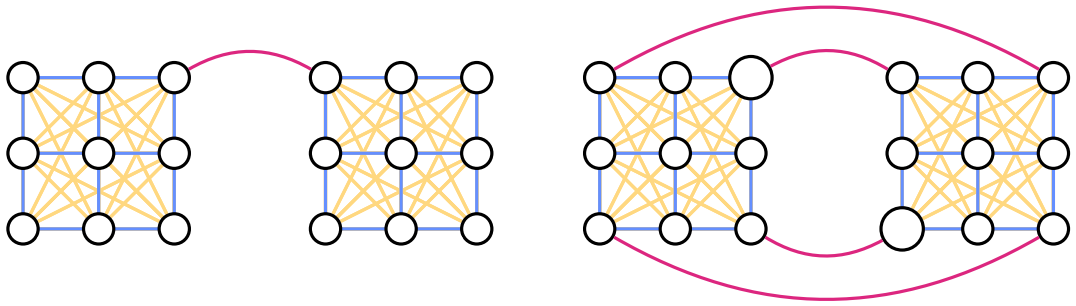
Below, I test three alternative measures for the prioritisation of regions for prophylaxis—risk, betweenness centrality, and random walk percolation centrality (RWPC). To simulate an invading pathogen, the initial state of the disease model has a single region in state $I$, with all other regions in state $S$. I assume that the emergence of the pathogen in this region triggers a prophylaxis campaign across the metapopulation, with all regions in state $S$ as possible targets for prophylaxis. Given a certain number of doses $n_d$, the campaign assigns enough doses to the highest priority region to achieve herd immunity in that region, then does the same for the second highest priority region, and so on, until it reaches a region

whose herd immunity threshold is higher than the number of doses remaining, in which case all of the doses remaining are assigned to that region and none to those with lower priority. I assume that the prophylaxis campaign in a targeted region is always completed before any epidemic occurs in that region. To obtain the mean number of infections I perform 10,000 simulations for each possible region of origination of the outbreak, using the simple metapopulation model described above, and take the mean final number of recovered individuals across all simulations.

To target regions based on risk, I perform 10,000 simulations of the outbreak from the initial state with no prophylaxis, and prioritise regions from most at risk to least at risk. We can treat the risk to region $i$ as equivalent to the probability that region $i$ experiences an epidemic. In principle, the risk is the expected number of infections, i.e the probability of the region becoming infected multiplied by $R(\infty)$. However, both $R(\infty)$ and the number of doses required to achieve herd immunity scale linearly with the size of the susceptible population. This means that the risk averted per dose is maximised by prioritising regions based on the probability of an outbreak in that region, from highest probability to lowest. For the RWPC strategy, I calculate the RWPC for each region according to the initial state of the network, and prioritise regions from highest RWPC to lowest. For the betweenness centrality strategy, I simply prioritise from highest betweenness centrality to lowest betweenness centrality. For both centrality measures, weighting the edges according to the rate of travel per individual results in more effective targeting than weighting them according to the total flux of travelling individuals.

## Artificial networks 4.4

In order to compare risk-targeting and centrality-targeting strategies, I created an artificial network that could represent pathogen spread on multiple scales. The structure of this network is illustrated in 4.1a. The network is based on a "joined



**(a)** The joined grids network **(b)** The "four corners" variation

**Figure 4.1** The "joined grids" arrangements of regions (circles), in which the movement of the population happens between adjacent regions in each grid (blue lines), on longer-range travel routes within a grid (yellow lines), and along inter-grid travel routes (red lines). All regions have a fixed population of 200,000, except for the regions corresponding to the two larger circles in (b), which have fixed populations of 400,000.

grids" structure, in which the 18 regions are split into two identical subsets of 9, each of which is arranged in a grid with unit distance between orthogonally adjacent regions. Transmission can occur at three different scales—within regions (represented by circles in figure 4.1), within grids, and across the whole network.

Travel occurs at the highest rate between regions (represented by nodes) which are adjacent on the grid, and at a rate $\lambda_{ij}$ determined by distance between more distant regions. The within-grid travel rate is given by

$$\lambda_{ij} = \frac{\lambda}{(r_{ij})^a}, \tag{4.17}$$

where $\lambda$ is the travel rate between adjacent regions; $r_{ij} \geq 1$ is the distance between region $i$ and region $j$, as a multiple of the distance between adjacent regions; and $a$ is a parameter determining the amount of long-distance travel—the higher the value of $a$, the less long-distance travel occurs. This is a kind of gravity model, which is one of the standard models of human movement [230–233]. In the basic version of this network illustrated in 4.1a, there is a single travel route between the two grids, which runs from a region at the corner of one grid to a region at the corner of the other. The travel rate along this route is set to $\lambda_{\text{join}} = \lambda$. I assume

that each region has a fixed population of 200,000.

The results of prophylaxis distribution on the joined grids network using risk-, betweenness-, and RWPC-targeting strategies is shown in figure 4.2. In each panel, the y-axis shows the mean number of infected individuals resulting from an epidemic that starts with a random region in state $I$ and the remaining regions in state $S$, over 10,000 stochastic simulations for each possible starting region, using the model outlined in section 4.2. The total number of simulations used to calculate each mean value is therefore 180,000. The lower the number of infected individuals, the more effective the control of the pathogen. Note that the scale varies from panel to panel, to allow the variation in effectiveness to be seen clearly in each case. The x-axis in each panel shows the number of doses administered, $n_d$, representing the resources available for prophylactic control. The three coloured lines show the mean number of infected individuals when doses are distributed according to the risk-targeting strategy (blue), the betweenness-targeting strategy (red), or the RWPC-targeting strategy (yellow). The panels are arranged left-to-right with increasing values of within-region basic reproduction number $R_0$, and top-to-bottom with increasing values of adjacent-region travel rate $\lambda$. The range of parameters for which simulation results are shown is chosen to illustrate a change in the relative effectiveness of different prophylaxis strategies.

We see that risk-targeting is the best strategy under some parameter combinations and not under others. When both $R_0$ and $\lambda$ are small, the risk-targeting strategy is always the most effective of the three considered, for any number of available doses. For some combinations of $R_0$ and $\lambda$, such as $R_0 = 8$ and $\lambda = 2 \times 10^{-6}$, the risk-targeting strategy is only the most effective when there are many doses available. Finally, when both $R_0$ and $\lambda$ are large, the centrality-targeting strategies are more effective than the risk-targeting strategy at all values of $n_d$. The RWPC strategy is always at least as effective as the betweenness centrality strategy, with the difference between the two being largest when $R_0$ and $\lambda$ are small and $n_d$ is

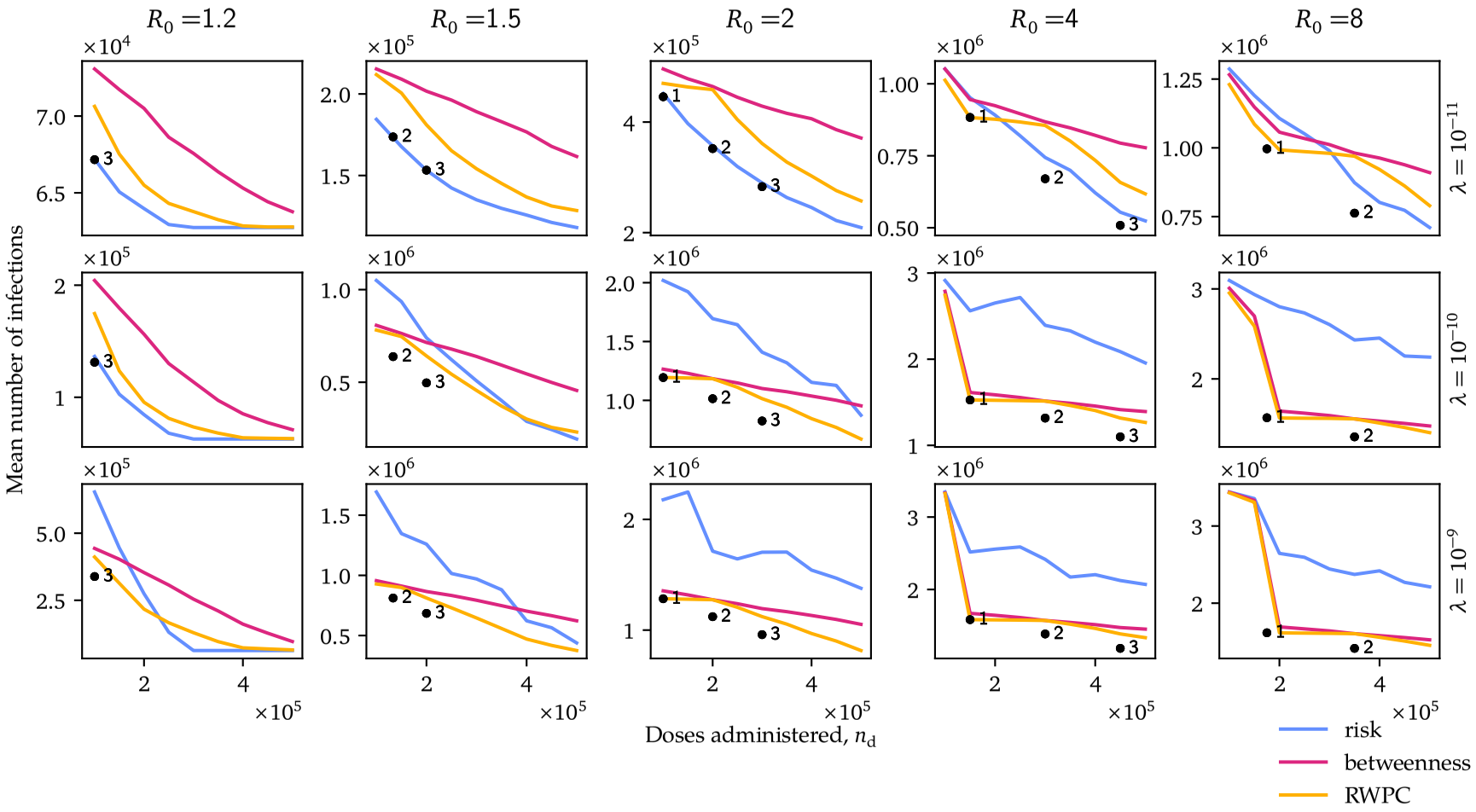



**Figure 4.2** Plots of the mean total number of infections due to a disease outbreak against the number of prophylaxis doses administered according to one of three strategies, for five different values of $R_0$ and three different overall travel rates, in the joined grids network. The long-distance travel parameter $a$ in equation 4.17 is set to 2 in all cases, and $\mu = 0.5$. It is possible to calculate the optimal prophylaxis allocation for some parameter values. The resultant mean numbers of infections are shown by the y-positions of the black dots, labelled with the number of regions which can be allocated enough doses for herd-immunity at that value of $n_d$.

large. The large dips in mean infections seen in the lower right panels of the figure correspond to the point at which sufficient doses become available to achieve herd immunity in one of the patches that join the two grids. This also happens when $R_0 \leq 2$, but outside of the range of doses shown, and it does not have as dramatic an effect when $\lambda = 10^{-11}$.

Since each region has the same population size, it is sometimes computationally feasible to determine the optimal allocation of prophylaxis using a brute force approach. That is, the approach in which we simulate pathogen spread under all possible prophylaxis allocation prioritisations, and select the prioritisation that leads to the smallest mean number of infections. To limit the number of possible allocations, I constrain the total number of doses to be equal to a small integer multiple $n_{\mathrm{H}}$ of the herd immunity threshold for a single population, and I assume that no less than the number of doses required for herd immunity are ever allocated to a given region, if any doses are allocated there at all. The number of possible allocations under these assumptions is $18!\,/(n_{\mathrm{H}}!\,(18-n_{\mathrm{H}})!\,)$, typically a large but manageable number for small $n_{\mathrm{H}}$. The resultant mean number of infections for each of the optimal allocations is indicated by the y-position of the black dots in figure 4.2, which have labels corresponding to the number of regions receiving prophylaxis. The spacing between these dots scales with the herd immunity threshold, as determined by $R_0$. In general, the best strategy out of those considered for any combination of parameters comes close to optimality in effectiveness, with risk-targeting at low travel rates being indistinguishable from the fully optimal strategy.

These results support the hypothesis outlined on page 88 concerning the allocation of prophylactic control in various regions—namely, that centrality-targeting would perform better than risk-targeting when $R_0$ and inter-region travel rates were high relative to the number of doses distributed. I have found that this general pattern remains constant when the joined grids network is changed in

small ways, such as varying the value of $a$, or adding extra connections between the two grids (see appendix B). However, it is possible to alter the joined grids network in such a way that the results of various control strategies change more substantially.

In figure 4.1b I illustrate a variation on the joined grids network with a set of alterations that result in significant changes to the dynamics of infectious disease control in the metapopulation model. I call this variation the "four corners" network. Its structure is similar to the joined grids network, but there are four routes between the two networks, joining each corner of each grid to a corner of the other grid. The inter-grid routes each have travel rate $\lambda_{\text{join}} = \lambda/4$. Each region has a fixed population of 200,000, except for one region on the corner of each grid, which has a population of 400,000, to represent high-population travel hubs. These two regions are not connected by any direct travel route. To account for the greater travel to more populous regions predicted by the gravity model of human movement, I now write a modified version of equation 4.17:

$$\lambda_{ij} = \frac{\lambda N_i}{(r_{ij})^a N}, \tag{4.18}$$

where $N = 200,000$ is the population size of every region except the two larger regions. In this case $\lambda$ characterises the overall volume of travel, but does not give the travel rate between adjacent regions in all cases—the travel rate from a region $k$ adjacent to a region $l$ with population size 400,000 is $\lambda_{kl} = 2\lambda$.

In figure 4.3, I illustrate the change in dynamics caused by changing the network structure. Because the regions do not all have the same population size, the number of possible distinct prioritisations becomes too large for brute force optimisation to be feasible. We see, however, that with multiple routes between the two grids, and an increase in the population size of two patches, the risk-targeting strategy becomes the most effective of the strategies considered for al-

most all parameter combinations. For very high spread rates the pathogen infects all those not protected by prophylaxis or by regional herd immunity, and so the mean number of infections becomes the same linear function of the number of doses for all strategies.

This adds an important caveat to the hypothesis outlined on page 88; the network structure must allow effective use of centrality-targeting. In the four corners network, the presence of many weak edges between the two grids means that the effectiveness of centrality-targeting never exceeds that of risk-targeting. This effect is however not seen without the two larger sized regions on the grid corners, or with only two connections. Given the specificity of the alterations required to elicit this behaviour, it is important to check whether we can reproduce it in a network based on empirical travel data.

## Empirical networks 4.5

To investigate prophylaxis strategies in empirical networks, I made use of of two freely available data sets concerning human movement patterns (see figure 4.4). The first is a subset of the US domestic air traffic network, sourced from the Bureau of Transport Statistics [236]. I used as nodes in the travel network only the 20 airports with the greatest sum of passengers arriving and passengers departing, and defined the weight of a directed edge between two of these nodes $i$ and $j$ in proportion to the total number of passengers travelling from airport $i$ to airport $j$ in 2019. To assign a population size to each node, I used the population size of the urban area in which the airport is situated [237]. There are two airports in the data set situated in Chicago. In this case, I split the population equally between the two airports, assigning half the total population of the Chicago urban area to each node.

The edge weight $\lambda_{ij}$ (i.e. the travel rate) from node $i$ to node $j$ in this network

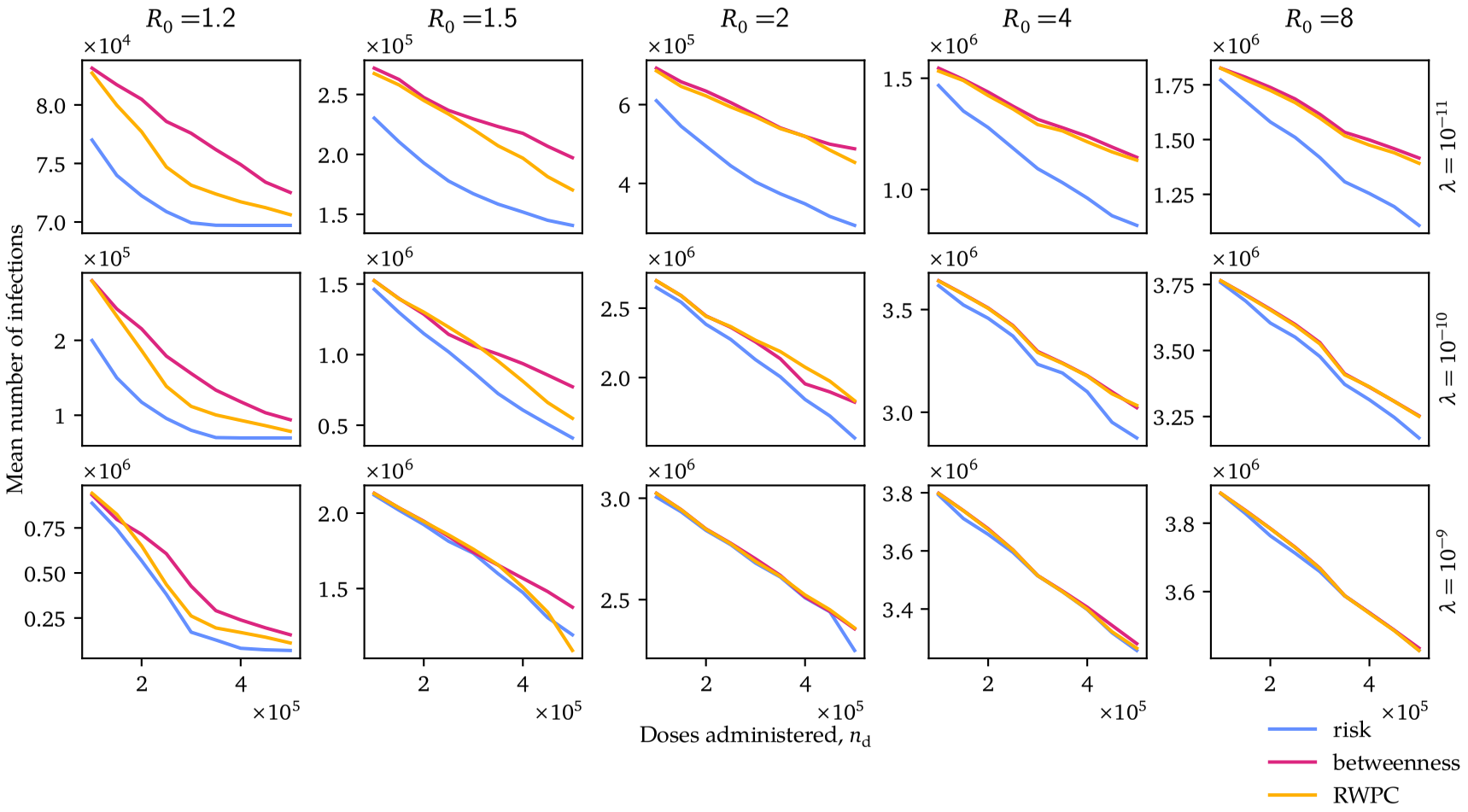
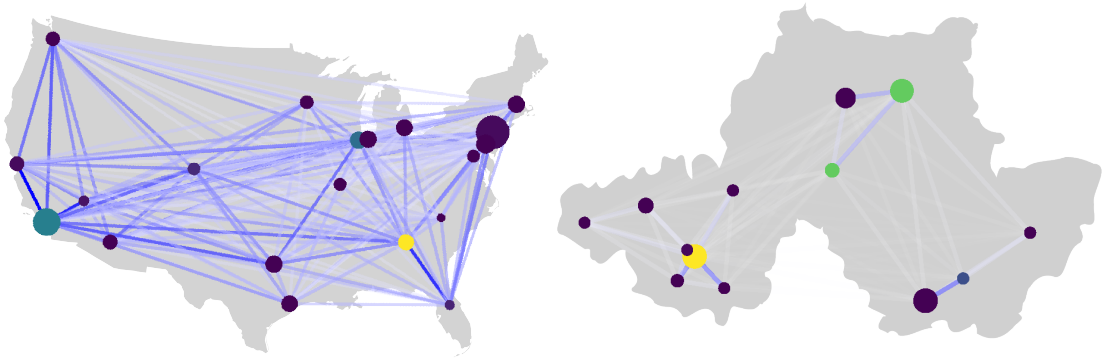



**Figure 4.3** Plots of the mean total number of infections due to a disease outbreak against the number of prophylaxis doses administered according to one of three strategies, for five different values of $R_0$ and three different overall travel rates, in the four corners variation of the joined grids network. The long-distance travel parameter $a$ in equation 4.18 is set to 2 in all cases, and $\mu = 0.5$.



**(a)** US air passenger traffic **(b)** Northwest England commuters

**Figure 4.4** Two travel networks based on empirical data. The colour and opacity of each edge scales with the sum of its weights in either direction, normalised separately in each network relative to the highest weighted edge. The area of each node scales with the population size it represents, and the colour indicates its betweenness centrality, with yellow regions having the highest centrality and purple regions the lowest. Chicago has two airports in the US air passenger traffic data set, so each of these nodes is slightly offset to make them both visible. [234, 235]

is therefore given by:

$$\lambda_{ij} = s\frac{t_{ij}}{N_i} \tag{4.19}$$

where $t_{ij}$ is the total number of passengers travelling from airport $i$ to airport $j$ in 2019, $N_i > 0$ is the population size of node $i$, and $s$ is an overall scaling factor that we can vary.

The second empirical network I used was sourced from the 2011 UK census [238]. Using this, I estimate the amount of travel between all cities and towns with populations over 100,000 in an area of Northwest England including Greater Manchester and West and South Yorkshire (see figure 4.4b). The census includes data for the location of the residence and place of the work of all respondents, which I use to estimate the number of people $c_{ij}$ who commute between any given pair of cities $i$ and $j$. I model the travel rate according to the equation:

$$\lambda_{ij} = s\frac{c_{ij} + c_{ji}}{2N_i}, \tag{4.20}$$

where $N_i > 0$ is the population size of city $i$, and $s$ is an overall scaling factor

that we can vary. The travel rate is forced to be symmetric because a commuter travelling between cities can transmit a pathogen either from the source city to the destination city (on the outbound journey) or from the destination city to the source city (on the inbound journey) [239]. The key assumption of these models of empirical travel data is that total travel is proportional either to air travel or to commuter travel, respectively.

Figures 4.5 and 4.6 illustrate the results of various prophylaxis strategies in these empirically derived networks. The structure of the figures is identical to figure 4.2, but with the scale of travel indicated by the parameter $s$ rather than $\lambda$. In figure 4.5, we see that the dynamics of prophylaxis targeting in the US air passenger traffic network is similar to that in the joined grids network, with the risk-targeting strategy performing best at low $R_0$ and low inter-regional travel, and the centrality-targeting strategies performing best at high $R_0$ and high inter-regional travel. The main differences include the overall high performance of the RWPC strategy, which is, or is very close to, the most effective strategy in all cases; the smoothness of the curves, with no sharp dips in mean infections; and the convergence of the effectiveness of the various strategies for $s = 10^{-6}$.

In figure 4.6, we see a case of an empirical network—the Northwest England commuter network—in which the risk-targeting strategy is consistently the most effective at reducing infections, as seen for the "four corners" artificial network in figure 4.3. The effectiveness of the different strategies converges as $s$ increases.

These results show that both types of behaviour illustrated in section 4.4 can be replicated in empirical networks. In general, in networks made up of distinct clusters of regions, with many weak links between these clusters, the hypothesis outlined on page 88 does not hold—the risk-targeting strategy is more effective than the centrality-targeting strategies for all parameter values considered. However, in the two other network structures I investigated (both the artificial joined grids network and the empirical US air passenger traffic network) the hypothesis

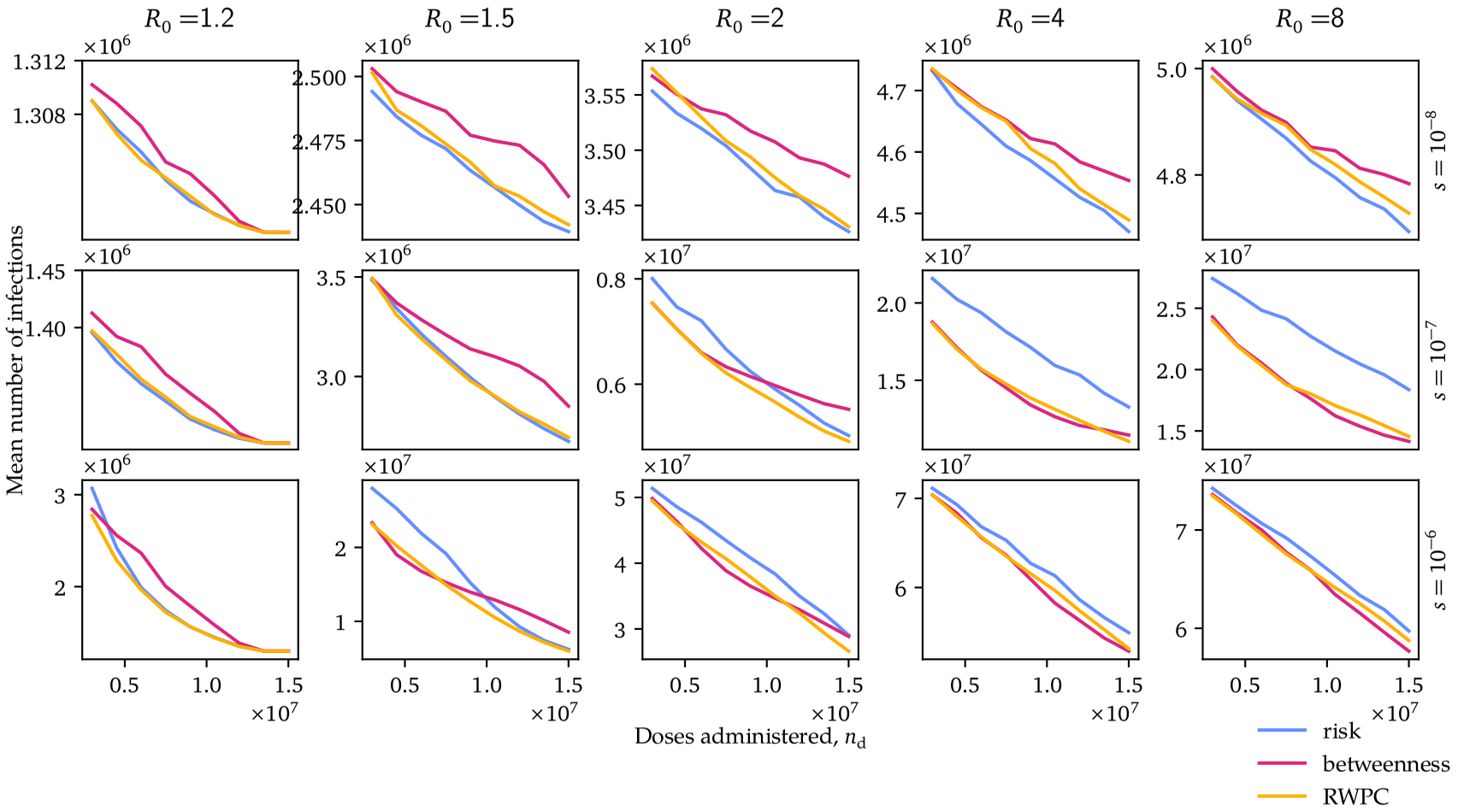



**Figure 4.5** Plots of the mean total number of infections due to a disease outbreak against the number of prophylaxis doses administered according to one of three region prioritisation policies, for five different values of $R_0$ and three different values of the travel scaling factor $s$, in the US air passenger traffic network.

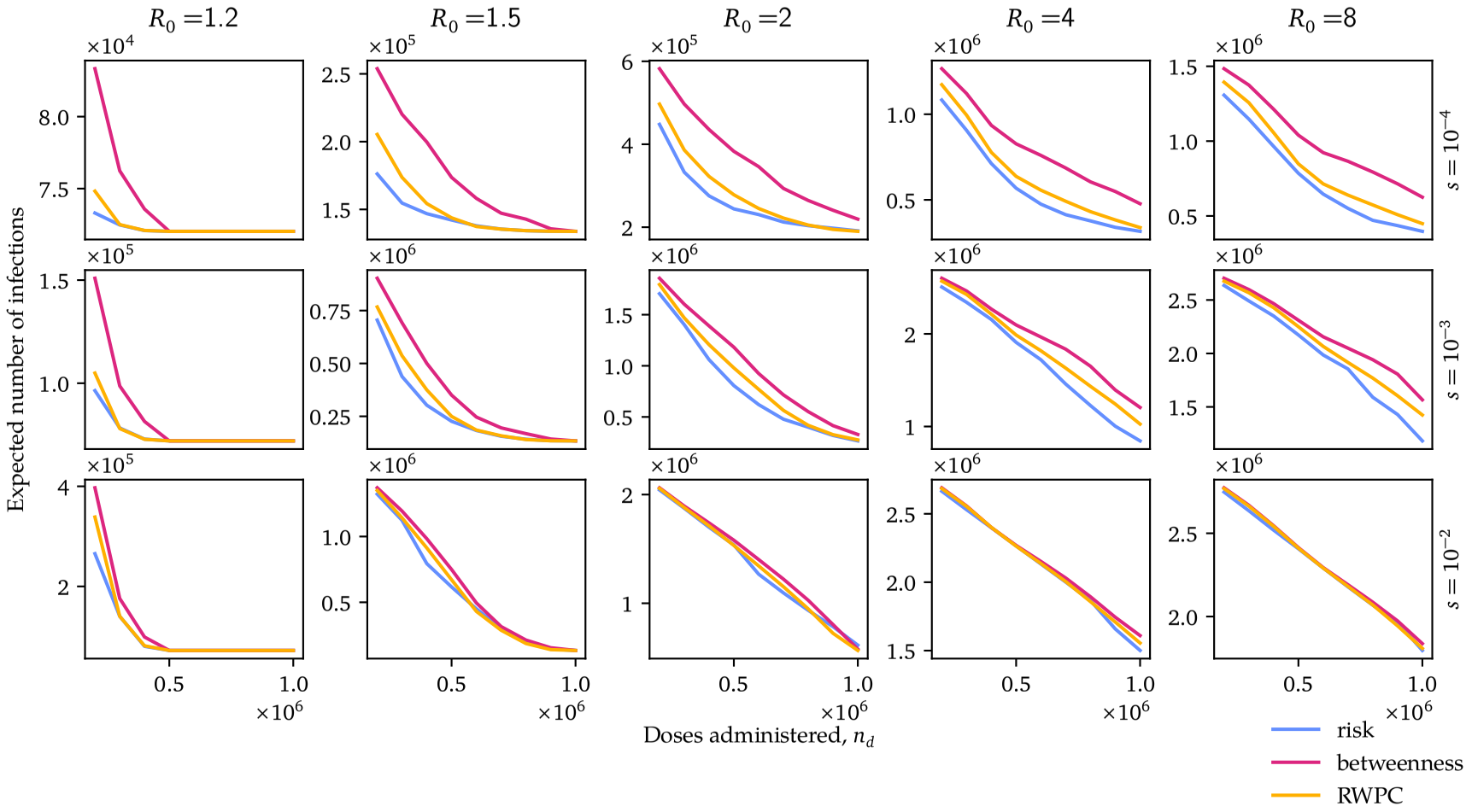



**Figure 4.6** Plots of the mean total number of infections due to a disease outbreak against the number of prophylaxis doses administered according to one of three region prioritisation policies, for five different values of $R_0$ and three different values of the travel scaling factor $s$, in the Northwest England commuter network.

holds, and centrality-targeting strategies can outperform the risk-targeting strategy at sufficiently high travel rates or $R_0$, or sufficiently low $n_d$.

## Non-pharmaceutical interventions ⬡ 4.6

The above results use a susceptibility-reducing intervention as an example of prophylactic control. However, other forms of pathogen control can also be allocated spatially [86, 240]. In this section I extend the framework introduced above to treat an intervention that reduces the transmission of the pathogen without reducing the susceptibility of the population. This can be thought of as a simple model of social distancing non-pharmaceutical interventions (NPIs) aimed at reducing community transmission of an infection. As seen during the COVID-19 pandemic, this kind of intervention can be used to reduce the spread of a pandemic pathogen. I ignore the possibility of placing only a subset of the population of a region under social distancing rules, assuming that prophylaxis allocations in this context target all the individuals in a region, or none of them.

I use the epidemiological model as defined in section 4.2. Instead of having a fixed number of prophylaxis doses $n_d$, I assume that some fixed maximum number of individuals $n_r$ can be placed under transmission-reducing restrictions. This allows comparison with the results above, and can be thought of as representing some limit on the amount of economic and social activity that a government is willing to suppress. Regions placed under these restrictions experience a 66% reduction in transmission rate, and no reduction in susceptibility. To create a list of regions under restrictions, I assume that the administrators of the NPIs place the entire population of the highest-priority region under restrictions, and continue down the priority ordering until placing the next region under restrictions would cause the total population under restrictions to exceed $n_r$. This last region is not placed under restrictions, meaning that the total population under restrictions

could be significantly below $n_r$.

The effects of various strategies for targeting NPIs in the US air travel network can be seen in figure 4.7. More plots illustrating NPI strategies in different networks are included in appendix B.3. In general, the results illustrated in these figures show that the relative performance of the various region prioritisation strategies is similar in the case of NPI allocation to the case of susceptibility-reducing prophylaxis allocation. In particular, the conclusion that risk-targeting performs better than centrality-targeting when $R_0$ and inter-region travel rates are high relative to the number of individuals targeted with prophylactic measures holds for the one-to-one joined grid network and the US air passenger traffic network, while risk-targeting almost always performs as well or better than centrality-targeting in the four corners joined grids network and the Northwest England commuter network. These plots are less smooth than the corresponding plots in figures 4.1–4.6, due to discontinuities when $n_r$ becomes large enough to place an additional region under restrictions.

## Conclusions ⬡ 4.7

In this chapter, we have seen that spatial targeting of outbreak control based on risk can be highly successful in many contexts, but that alternative strategies based on centrality measures can be more effective than risk-targeting in certain situations. In particular, in metapopulations with certain kinds of travel network, we see a validation of the hypothesis presented on page 88—that, when the resources available for control are limited and a pathogen has the potential to spread widely, targeting regions with high centrality can be more effective than targeting regions at high risk (see figures 4.2 and 4.5). We have seen that this is true whether control takes the form of a susceptibility-reducing pharmaceutical such as a vaccine, or transmission-reducing non-pharmaceutical interventions (see fig-

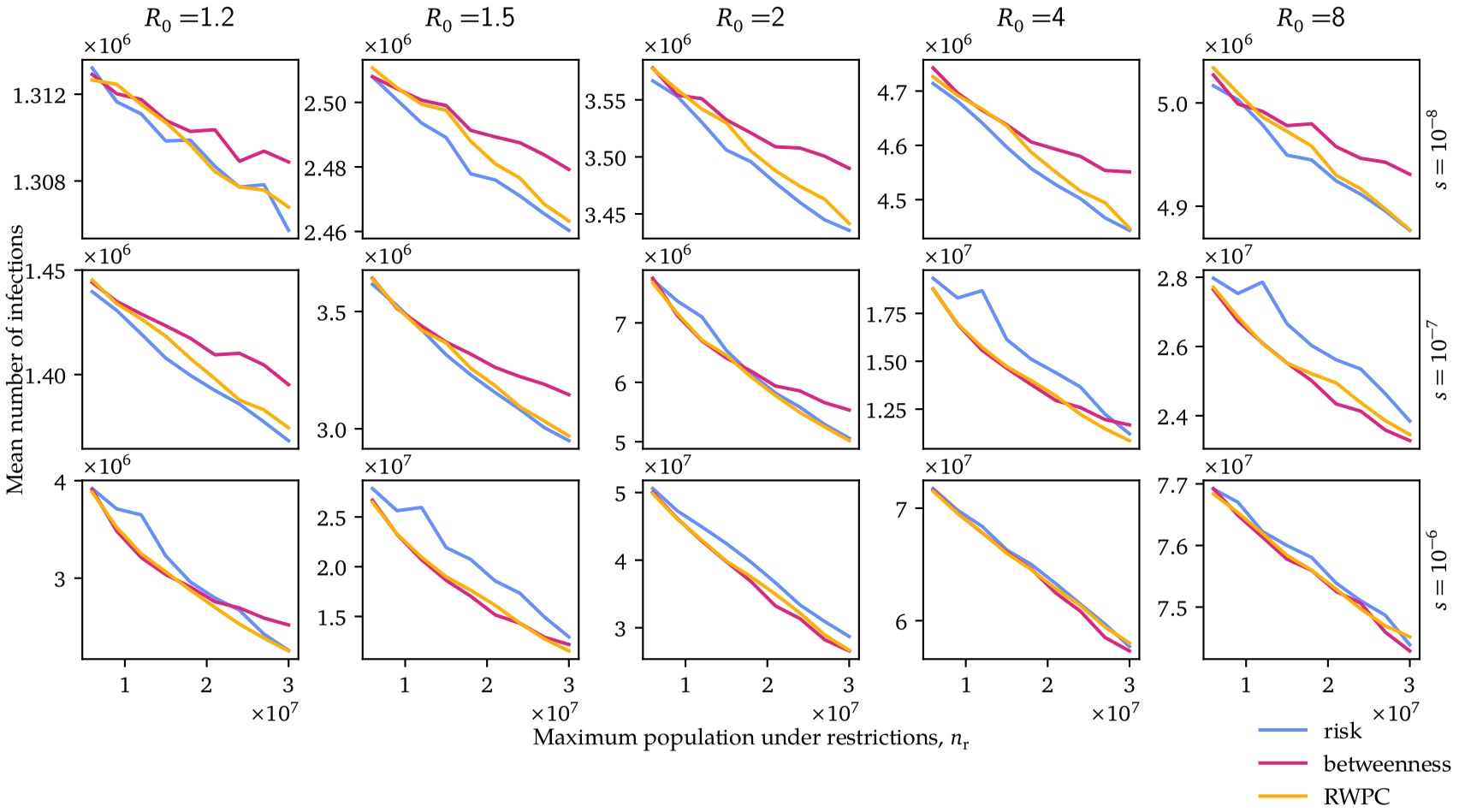



**Figure 4.7** Plots of the mean total number of infections due to a disease outbreak against the number of individuals placed under transmission-reducing restrictions, chosen according to one of three NPI strategies, for five different values of $R_0$ and three different values of the travel scaling factor $s$, in the US air passenger traffic network.

ure 4.7 and appendix B.3). In both cases, some travel networks are structured such that centrality-targeting is never more effective than risk-targeting (see figures 4.3 and 4.6).

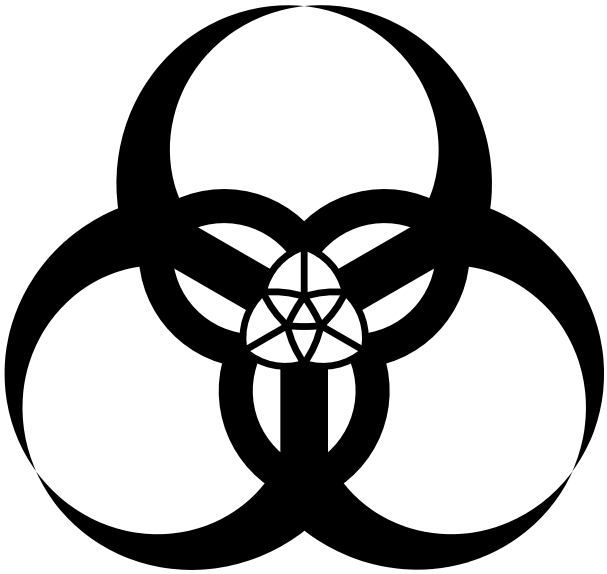
This provides an important perspective on the widespread use of risk-targeting in studies of spatial disease control. Risk-targeting can perform very well in certain networks, or in certain parameter regimes, but is far from universally superior to alternative strategies. In particular, when a pathogen is spreading quickly, or when resources available for control are limited, centrality-targeting strategies should be considered as an alternative. However, it is important to choose the correct centrality measure for the system at hand. We have seen that blunt application of a popular measure of node importance, such as betweenness, can result in a less effective allocation of vaccines compared to a measure tailored to the dynamics of a pathogen spreading on a network. In this chapter I have used a variation on the random walk percolation centrality measure due to Piraveenan et al. [211], but different scenarios may require the application of different measures. For instance, if the current distribution of infections is unknown, percolation centrality cannot be applied, and random walk betweenness centrality may be more effective.

The scenarios explored in this chapter assume a pathogen that is invading, and, in the case of vaccination, a distribution campaign that can be completed before the pathogen becomes widespread. The dynamics of control of a pathogen that is epidemic or endemic in a population differ significantly from those for an invading pathogen [96, 241, 242]. Further work is required to know whether risk-targeting is the most effective strategy for combating an endemic pathogen. With SARS-CoV-2 already prevalent in most regions of the world, this means that the results presented in this chapter are not applicable to ongoing COVID-19 vaccination efforts—however, they may provide insight into preventing the invasion of new strains of the virus, if these variants are detected early.

Whichever type of allocation strategy is used, actual implementation of pro-

phylaxis and NPIs requires understanding details not included in the simple model used here. Many prophylactic drugs and vaccines require more than a single dose to be effective, and timing and logistics of these doses should be included in a full model of pandemic pathogen control. Vaccines do not, in general, result in a 100% reduction in susceptibility, and their effects can be split into several components—infection-reducing, disease-reducing, and transmission-reducing [50, 243–246]. The relative efficacy of a vaccine in each of these components, and its interaction with NPIs, should be considered. Protection from vaccines can wane over time, increasing the importance of dynamical detail in the disease model [247–249]. The effects of NPIs, drugs, and vaccines can vary between individuals, and so a full model of pandemic control should consider demographic structure (e.g. age, occupation) in addition to spatial structure. Finally, implementation of a control strategy may require constraints on the spatial complexity of an allocation, since highly spatially resolved strategies may be logistically impractical.

Notwithstanding their potential complexity, spatial prophylaxis strategies can be useful as tools for pandemic prevention. When an outbreak of an emerging pathogen with pandemic potential is detected early, it may be possible to prevent global spread using spatial control. The results presented above suggest that the best methods for spatial prioritisation vary from context to context, but that a judiciously chosen method can lead to large reductions in infections.



A stereographic projection of the regular icosahedron, obtained by projecting the polyhedron onto a surrounding sphere, then projecting that sphere onto the 2D Cartesian plane. Here, the line weights are manipulated to emphasise the resemblance of the projection to the international standard biohazard symbol. This symbol indicates the presence of "infectious agents presenting a risk or potential risk to the well being of [humans], either directly through ... infection or indirectly through disruption of [our] environment." [250]

# 5 Prophylaxis against SARS-CoV-2 in England

In chapter 4, I modelled the distribution of a prophylactic drug or vaccine that grants long-lasting immunity to infection with a given pathogen. However, most prophylactic drugs do not provide long-term protection. Such drugs must be taken regularly in order to reduce the susceptibility to or transmissibility of an infection, and if the supply runs out, then the susceptibility and transmissibility revert to what they were before the distribution of the drug. The use of such prophylactic drugs to contain or mitigate an influenza pandemic has been discussed extensively in the influenza literature [251–255]. In this chapter, I extend this discussion to the case of SARS-CoV-2, considering a hypothetical scenario in which a stockpile of an effective antiviral against SARS-CoV-2 is available when the virus is first introduced into the UK. I assume that this antiviral can be used prophylactically to achieve a temporary reduction in susceptibility and transmission, and examine the results of spatially targeted distribution of such an antiviral. I go on to discuss the role that broad-spectrum antivirals could have in preventing future pandemics.

Modelling studies of influenza pandemics have shown that the prophylactic use of antivirals could form an important element of pandemic response, when effective antivirals are available [252, 253]. In particular, intensive use of a stockpile of antivirals early on in an outbreak can act to slow the progress of the epidemic in a given country, potentially allowing time for an effective vaccine to be

formulated and distributed [251, 253]. These results have been used to argue that large stockpiles of broad-spectrum influenza antivirals should be accrued in order to reduce the danger from pandemic influenza [254, 255].

General research on pandemic influenza that allows for uncertainty in the characteristics of the pandemic strain could generalise to other respiratory pathogens, such as SARS-CoV-2 [131, 255]. Over the course of the COVID-19 pandemic, many organisations and governments have compiled rich data sets on the spread of this virus [124, 256, 257]. These constitute an unprecedentedly large supply of data on a pandemic pathogen, and can inform our preparations for future pandemics. Below, I apply the results of modelling studies of influenza prophylaxis strategies to a model of SARS-CoV-2 spread in England, in order to evaluate whether an effective coronavirus antiviral could have helped mitigate the COVID-19 pandemic. If antiviral prophylaxis is an effective form of pathogen control in models of both influenza and SARS-CoV-2, this is a good reason to believe that prophylactic drug stockpiles could form an important part of pandemic preparedness strategy for a range of PPPs.

In practice, pharmaceutical prophylaxis is likely to form only part of the response to a pandemic pathogen. Non-pharmaceutical interventions, such as ventilation of indoor spaces, isolation of symptomatic cases, travel restrictions, closure of schools and universities, and general social distancing, have proven to be an important part of the response to COVID-19 [52, 136, 208, 240, 258]. Previous work on influenza prophylaxis has found that combining antivirals with NPIs may be necessary to control a pandemic pathogen [253, 254]. I therefore investigate how the distribution of a prophylactic drug might have helped to reduce the spread of SARS-CoV-2 under various levels of NPIs. I consider three possible levels of NPIs—low, similar to the interventions in place in England in mid-March 2020; medium, similar to the interventions in place in early October 2020; and high, similar to the interventions in place in mid-November 2020. Data from the Office

for National Statistics and the UK Health Security Agency on the spread of SARS-CoV-2 allow parameterisation of a model of pathogen spread in England at these different times (see section 5.2).

SARS-CoV-2 was introduced into the UK multiple times during the first phase of the epidemic [259]. This made containment of the outbreak particularly challenging. In section 5.6, I will consider a similar situation, in which a pathogen is introduced into multiple random regions simultaneously. In section 5.5 I consider a best-case scenario, in which the pathogen is detected, and a response launched, while the outbreak is still limited to a single region. In both cases, I assume that no new introductions of the pathogen occur after the pathogen is detected, either due to restrictions on international travel or a low base rate of introduction.

## Modelling travel in England ⬡ 5.1

A pandemic, and the government response to it, will affect the movement patterns of individuals as they adapt to the risk of disease and any restrictions put in place to reduce it. During the COVID-19 pandemic, researchers have collected a great deal of data on the movements of individuals at different levels of disease risk and under different stages of restrictions. This allows us to realistically model the movement of a human population during the COVID-19 pandemic. Here, I make use of travel data collected by Facebook Data for Good, and processed by Hamish Gibbs of the London School of Hygiene and Tropical Medicine, in order to parameterise a travel network for an epidemiological metapopulation model [118, 119].

The Facebook Data for Good dataset consists of the locations of Facebook users who have location services enabled. For each 8-hour period, a given such Facebook user is assigned to the Bing Maps cell in which they record the largest number of locations. The Bing Maps cells are derived from the Bing Maps Tile Sys-

tem [260]. This location data for each 8-hour period is aggregated into a matrix, with the entry in row $i$ and column $j$ giving the number of individuals assigned to cell $i$ in the previous period who are assigned to cell $j$ in the following period. Any elements in this matrix with values less than 10 are set to zero by Facebook before sharing with researchers, in order to preserve users' privacy. Travel to and from 12 Bing cells around Swindon was erroneously classed as international, and excluded from the data set. Hamish Gibbs and his collaborators found that the Facebooks users in a given cell was not strongly biased relative to the total population of that cell with regards to a range of demographic variables [119].

Gibbs and his collaborators used the Facebook data to arrange the UK population into movement communities—here called 'clusters' [119]. The clusters I use here are based on the InfoMap algorithm, which groups nodes in a weighted directed network based on the movement of a random walk through that network [35]. In short, an InfoMap cluster represents a group of nodes within which a random walk is likely to spend an unusual amount of time before moving to another cluster. The number of movements within a given cluster is typically much larger than the number of movements between that cluster and others. Gibbs aggregated the movement data over periods of one week, generated clusters from this aggregated network, and provided me with the cluster analysis and the total number of observed movements between clusters by week.

These clusters are illustrated in figure 5.1 for the week beginning 2020-08-24. Because COVID-19 case data is not easily available in terms of Bing Maps cells, Gibbs provided me with the clusters in terms of *middle layer super output areas* (MSOAs), with each cluster characterised by a set of constituent MSOAs. MSOAs are aggregates of Census output areas created by the UK's Office for National Statistics (ONS), each containing between 5,000 and 15,000 people [261]. Demographic information for MSOAs is available from the ONS website, including confirmed COVID-19 case numbers by week.



**Figure 5.1** Map illustrating the structure of the 160 InfoMap clusters on the Facebook Data for Good location network, in England, in the week beginning 2020-08-24. Each circle represents an MSOA, with its area proportional to the MSOA population, its position at the population-weighted centroid of the MSOA, and its colour determined by the cluster in which the MSOA falls. Each cluster is assigned a colour at random. A line of the same colour traces the convex hull of the centroids of the MSOAs in the cluster. These colours do not represent anything other than cluster membership, and clusters with similar colours are not necessarily similar in any other respect.

With clusters defined in terms of MSOAs, we can use census data to determine certain properties of the clusters. I assign to each cluster $i$ a population $N_i$, equal to the sum of the ONS mid-2019 estimates for the populations of each of the clusters' constituent MSOAs [262]. I assign each cluster a population weighted centroid, equal to the population weighted centroid of the population weighted centroids of the constituent MSOAs, as calculated in the 2011 Census [263]. I use this to define a measure of distance $d_{ij}$ between two clusters $i$ and $j$, equal to the Euclidean distance between their centroids.

Gibbs' network provides the number of observed movements from one cluster to another over the course of a week. Only the subset of the population that uses Facebook location services is observed, so a statistical model is needed to estimate the actual proportion of individuals moving between clusters. Here I use a Bayesian framework. I assume some prior distribution for the probability that any given proportion of individuals moves from one cluster to another in a given week. I then update this prior using the Facebook movement data, according to Bayes' rule [264–266]. This yields a posterior distribution, which I use to estimate the number of individuals moving between pairs of clusters.

My statistical model uses a beta prior and a binomial likelihood, giving a beta posterior distribution for the proportion $0 \leq p_{ij} \leq 1$ of individuals in cluster $i$ who travel to cluster $j$ during a given week. When zero movements are observed between a particular pair of clusters, the model should allocate some non-zero probability to unobserved movements having taken place. This probability should be larger when the clusters are close together. I therefore parameterise the prior so that its mean is equal to the expected travel $1-\exp(-\lambda_{ij})$ predicted by a gravity model (see section 4.4):

$$\lambda_{ij} \propto \frac{N_j}{d_{ij}^a}, \tag{5.1}$$

where $N_j$ is the population size of cluster $j$, $d_{ij}$ is the distance between the centroids of cluster $i$ and cluster $j$, and $a$ is a parameter determining how quickly

travel reduces with distance. For the UK, $a \approx 1.79$ [233].

The mean of the beta distribution $\text{Beta}(\alpha, \beta)$ is given by $\alpha/(\alpha+\beta)$, so setting the mean to $\mu_{ij} = 1 - \exp(-\lambda_{ij})$ leaves one parameter free. In order for a positive probability to be assigned to zero travel when zero movements are observed, $\alpha$ must equal 1. So I set the prior probability of a given value $x$ of $p_{ij}$ to:

$$\begin{aligned} \Pr(p_{ij} = x) &= \text{Beta}\left(x \middle| 1, \frac{1}{\mu_{ij}} - 1\right) \\ &= \frac{(1-x)^{\frac{1}{\mu_{ij}}-2}}{B\left(1, \frac{1}{\mu_{ij}} - 1\right)}, \end{aligned} \tag{5.2}$$

where the normalisation takes the form of a beta function:

$$B(\alpha, \beta) = \int_0^1 y^{\alpha-1}(1-y)^{\beta-1}\mathrm{d}y. \tag{5.3}$$

If the number of movements from cluster $i$ to cluster $j$ in a given week is equal to $p_{ij}N_i$, but we only have data on a certain number $n_{o,i}$ of the individuals in cluster $i$, then the likelihood $\mathscr{L}(m_{ij})$ to observe a number of movements $m_{ij}$ from $i$ to $j$ is binomially distributed:

$$\mathscr{L}(m_{ij}) = \binom{n_{o,i}}{m_{ij}} \left(p_{ij}\right)^{m_{ij}} \left(1 - p_{ij}\right)^{n_{o,i}-m_{ij}}. \tag{5.4}$$

Using this expression for the likelihood, we can use Bayes' rule to express the posterior distribution for $p_{ij}$ as:

$$\begin{aligned} \Pr(p_{ij} = x) &= \frac{\binom{n_{o,i}}{m_{ij}} x^{m_{ij}} (1-x)^{n_{o,i}-m_{ij}} (1-x)^{\frac{1}{\mu_{ij}}-2} / B\left(1, \frac{1}{\mu_{ij}} - 1\right)}{\int_0^1 \binom{n_{o,i}}{m_{ij}} y^{m_{ij}} (1-y)^{n_{o,i}-m_{ij}} (1-y)^{\frac{1}{\mu_{ij}}-2} / B\left(1, \frac{1}{\mu_{ij}} - 1\right) \mathrm{d}y} \\ &= \frac{x^{m_{ij}}(1-x)^{\frac{1}{\mu_{ij}}-2+n_{o,i}-m_{ij}}}{B(1+m_{ij}, \frac{1}{\mu_{ij}} - 1 + n_{o,i} - m_{ij})} \\ &= \text{Beta}\left(x \middle| 1 + m_{ij}, \frac{1}{1-\mathrm{e}^{-\lambda_{ij}}} - 1 + n_{o,i} - m_{ij}\right). \end{aligned} \tag{5.5}$$

Below, I parameterise this model using data on movement from different times, in order to simulate different levels of NPIs. To simulate travel during a period of low NPIs, I parameterise the movement model using data from the week beginning 2020-03-16, before the first period of intensive restrictions was introduced in the UK [208, 240]. For medium NPIs, I use the week beginning 2020-09-28, when long-distance travel was discouraged but no intensive restrictions were in place. For high NPIs, I use the week beginning 2020-11-16, when the public was generally encouraged to remain in their homes. In each case, the posterior tends to be tightly peaked around the observed number of travellers, due to the large number of observations in the Facebook data set.

## Pathogen spread

 5.2

The probabilistic travel network characterised by the values of $p_{ij}$ provides one component of a metapopulation model of the spread of an invading pathogen in England, using Gibbs' movement clusters as the regions of the metapopulation. I now describe a model of the spread of a pathogen through this metapopulation.

One way to model the pathogen spread would be to use a Bayesian framework as in section 5.1, fitting a statistical model to the case data for each cluster. A summary of this approach is given in appendix C. In practice, the case data are not sufficiently reliable on the cluster level for robust parameterisation of a statistical model of the spread of SARS-CoV-2 in the UK. This means either that a differently structured statistical model must be used, or that a mechanistic model should be used instead. In this chapter, I use a mechanistic model of disease spread, parametrised using transmission data from the whole of England, rather than on a cluster-by-cluster basis.

This travel network gives the probability for a given proportion $p_{ij}$ of individuals assigned to one region $i$ to be located in another region $j$ at any given point

in time. Given this proportion, we can model new infections occurring in region $i$ as occurring due to three distinct processes. First, infections occur due to contacts amongst the proportion $1-\sum_j p_{ij}$ of individuals assigned to region $i$ who are currently located in region $i$. Second, the portion $\sum_j p_{ij}S_i$ of susceptible individuals assigned to region $i$ who are currently visiting other regions may become infected during their visit. Finally, a number $\sum_j p_{ji}I_j$ of infectious individuals who can cause new infections in region $i$ are currently visiting from other regions. This gives us an expression for the total rate of infections $\nu_i$ in region $i$:

$$\nu_i = \frac{\beta}{N_i} S_i I_i d_i^2 + \frac{\beta}{N_i} S_i \sum_j p_{ij} I_j d_j + \frac{\beta}{N_i} S_i d_i \sum_j p_{ji} I_j, \tag{5.6}$$

where $d_i = 1 - \sum_j p_{ij}$, and $\beta$ is the expected number of potentially infectious contacts per unit time. The rate of removal is simply $r_i = \mu I_i$, where $\mu$ is the sum of the recovery rate and the death rate. For this version of the stochastic SIR metapopulation model (see sections 1.3 and 1.4), the following events can occur in a very small time $\delta t$:

- infection in region $i$, were $S_i \to S_i - 1$ and $I_i \to I_i - 1$, with probability $\nu_i \delta t$;
- recovery in region $i$, where $I_i \to I_i - 1$ and $R_i \to R_i + 1$, with probability $\mu_i I_i \delta t$;
- no event, with probability $1 - \sum_i (\nu_i + \mu_i I_i)\delta t$.

# Parameterisation ⬡ 5.3

To model the dynamics of the COVID-19 epidemic in England, I parameterise the model presented above using data on the spread of SARS-CoV-2. For the pre-lockdown period, I use an estimate of 10 days for the infectious period of SARS-CoV-2, based on the width of the distribution of transmission post infection, to obtain $\mu = 1/10$; and the $R_0$ estimate of 3 to obtain $\beta = 3/10$ [174, 267]. For October and November 2020, estimates for the time-dependent reproduction

number $R_t$ and the growth rate $\gamma$ of the epidemic are available from the UK Health Security Agency (UKHSA) [256]. These can be used to estimate $\beta$ and $\mu$ using the expressions $\beta - \mu = \gamma$ and $\beta S/\mu N = R_t$ as derived in section 1.2. Rearranging yields:

$$\beta \approx \frac{\gamma R_t}{R_t - 1} \tag{5.7}$$

$$\mu \approx \frac{\gamma}{R_t - 1}, \tag{5.8}$$

assuming that $S \approx N$. The estimated parameters for the week ending 2020-10-02 are $R_t = 1.4 \pm 0.2$ and $\gamma = 6 \pm 2$, and for the week ending 2020-11-20 they are $R_t = 1.05 \pm 0.05$ and $\gamma = 1 \pm 1$. Taking the middle value of each range and substituting into equations 5.7 and 5.8, we get $\beta \approx 0.23$ and $\mu \approx 0.16$ for the week ending 2020-10-02, and $\beta \approx 0.21$ and $\mu \approx 0.20$ for the week ending 2020-11-20. I use these values to characterise pathogen spread under medium and high NPIs, respectively.

These values can be used to evaluate the fit to data of the stochastic SIR model as outlined above. I initialise the model using the data on population size and new weekly cases per MSOA, mapped onto Gibbs' movement clusters, for the week ending 2020-08-21 (the first week from which case data on the MSOA level are uniformly available). I use the clusters as defined for the closest available date, the week beginning 2020-08-17. I assume that the initial number of infected individuals per cluster is equal to the number of new cases testing positive over the previous week. I then simulate 100 outbreaks using the direct Gillespie algorithm (see section 1.3) with rates as defined above. For each simulation, I draw a random sample from the posterior for each value of $p_{ij}$, as given by equation 5.5. I draw a sample uniformly from the ranges of $R_t$ and $\gamma$ given by the UKHSA for each week in each sample, and used these to calculate values for $\beta$ and $\mu$, redrawing samples that lead to negative values for $\beta$ or $\mu$. I use the movement model as

defined for 2020-08-17 throughout the simulation, since changing the movement model would involve changing the identities of the clusters, which is not possible in a metapopulation model as defined here.

The results of these simulations are compared to UKHSA case data in figure 5.2 [257]. For the first seven weeks, the median weekly new infections in the simulations, shown by the solid grey line, are very close to the data, shown by black circles. However, by week eight the data fall outside the inter-quartile range (IQR) of the simulations, as indicated by the dashed lines. This shows that the UKHSA parameters alone are not enough to replicate the data out to this range. Week eight is the week ending 2020-10-16, the week in which the tier system of local restrictions was announced [268]. This may mean that the assumption of equal values of $\mu$ and $\beta$ in all regions becomes invalid from week eight onwards, leading to a discrepancy between the simulations and the data. It is also possible that the movement patterns in the UK had changed enough by this time that the data from 2020-08-17 no longer provide a good approximation.

## Prophylaxis ⬡ 5.4

There was no known effective prophylactic antiviral against SARS-CoV-2 when it was originally detected in the UK [174, 269, 270]. In this chapter, I consider an alternative scenario in which such a prophylactic drug does exist. Individuals who have taken a dose of the drug in the last 24 hours are assumed to have their susceptibility to infection reduced by 30%, and their transmissibility while infected reduced by 60%. These figures are taken from estimates of the effectiveness of oseltamivir, an influenza antiviral [271]. These figures have been used previously in modelling studies of pandemic containment [86]. It should be emphasised that there is no evidence that oseltamivir is effective against SARS-CoV-2 [272].

I assume that there exists a stockpile of this prophylactic drug, and that this

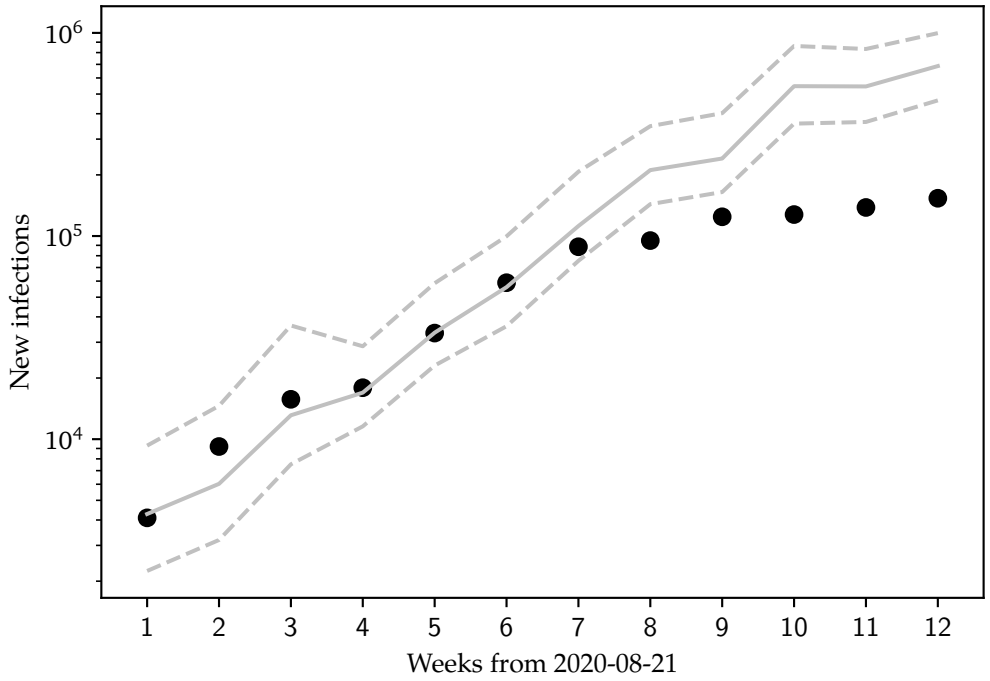



**Figure 5.2** A plot showing the range of simulated weekly new infections with COVID-19 in England (grey lines) as predicted using UKHSA's reported values of $R_t$ and the growth rate, along with the data on weekly cases in England (black circles) as reported by the UKHSA. The solid grey line shows the median of the simulations, and the dashed grey lines show the first and third quartiles. The last five data points lie outside of the predicted range, indicating that the values of $R_t$ and the growth rate as given by the UKHSA are not in themselves sufficient to predict the course of the epidemic out to this range.

entire stockpile can be distributed to a chosen subset of the English population. I model the ensuing spread of SARS-CoV-2 given stockpiles of three different sizes: 15 million, 50 million, or 100 million doses. These sizes are plausible given previous antiviral stockpiling in the UK [273, 274]. Once the invading pathogen is detected, I assume in the base scenario that it takes seven days to deliver the prophylactic drug to all targeted individuals, and compare this with an expedited scenario in which the distribution of the drug takes only one day. A population in which everyone has taken the prophylactic drug within the last 24 hours experiences a $1-0.4\times0.7=72\%$ reduction in $R_t$, which is enough to reduce any original value of $R_t$ under 3.57 to below 1.

## Early detection 5.5

When an invading pathogen is detected early, it could be possible to contain the outbreak. I assume that early detection means that the outbreak is detected when there are only ten infected individuals, all contained within one movement cluster. For consistency, I always assume that this is the cluster containing the city of Manchester.

I assume that all doses are allocated to individuals in the cluster in which the pathogen is first detected. This is the opposite case to the approach in chapter 4, where I assumed that the initially affected region was *never* targeted. I assume a different strategy here because I am modelling a different kind of prophylactic—instead of granting indefinite immunity, the antiviral drug being modelled reduces susceptibility in susceptible individuals, and transmissibility in infective individuals, for a limited time. This means that it is both more effective at slowing an ongoing outbreak in a region where it is distributed, and has very little effect in a region with no ongoing outbreak, since the protection ends once the stockpile runs out.

| Stockpile size ($\times 10^6$) | 0 | 15 | 50 | 100 |
|---|---|---|---|---|
| Low NPIs | 0–4% | 0–4% | 0–4% | 0–4% |
| Medium NPIs | 0.4–2% | 3–5% | 12–16% | 26–32% |
| High NPIs | 47–54% | 46–52% | 80–85% | 89–93% |

**Table 5.1** Extinction probability for an invading pathogen after 90 days, under different NPIs, and with different sizes of prophylactic drug stockpiles, when the stockpile is distributed starting from day seven of the simulation. Extinction is unlikely with low NPIs, but a combination of increased NPIs and a large stockpile of prophylactic drugs can lead to a high probability of extinction.

The population size of the cluster containing the city of Manchester is 2.6 million for the week beginning 2020-03-16, and 2.7 million for the weeks beginning 2020-09-28 and 2020-11-16. I assume that all individuals in the cluster are given a continuous course of the prophylactic drug, so a stockpile of 15 million doses is enough for 5 full days of prophylaxis, a stockpile of 50 million doses is enough for 18 to 19 days, and a stockpile of 100 million doses is enough for 37 to 38 days. Intensive prophylactic use of antivirals, as opposed to a more conservative approach, is supported by the literature on influenza prophylaxis [251, 253].

Assuming that the protection offered by the prophylactic drug begins at the end of day seven of the simulation, I run Gillespie direct method simulations for each scenario, i.e. each combination of NPIs and stockpile size. The resultant probabilities of extinction of the invading pathogen after 90 days are shown in table 5.1. Variation in the outbreaks comes about both from stochastic variation in the simulations, and from stochastic sampling from the posterior of the movement model. The 95% confidence interval for these probabilities is calculated using the Wilson score interval [275]. With low NPIs, the outbreak grows larger than at high NPIs, meaning that the simulation is slower as more intervals between infection events must be calculated. For this reason, I performed 100 simulations for each scenario with low NPIs, and 1000 simulations for each scenario with medium or high NPIs. The results for medium and high NPIs are therefore reported to a higher precision than for low NPIs. The analogous results for the expedited scenario, in which the protection offered by the prophylactic drug begins at the

| Stockpile size ($\times 10^6$) | 0 | 15 | 50 | 100 |
|---|---|---|---|---|
| Low NPIs | 0–4% | 0–4% | 0–4% | 0.6–7% |
| Medium NPIs | 0.4–2% | 6–9% | 27–31% | 48–54% |
| High NPIs | 47–54% | 99-100% | 98-100% | 99–100% |

**Table 5.2** Extinction probability for an invading pathogen after 90 days, under different NPIs, and with different sizes of prophylactic drug stockpiles, when the stockpile is distributed starting from day one of the simulation. This increases the probability containment compared to the values in table 5.1. The entries for stockpile size 0 are the same as those in table 5.1.

end of day one of the simulation, are shown in table 5.2.

The panmictic stochastic SIR model gives the probability that a pathogen will go extinct without causing an epidemic as $(1/R_t)^{I(0)}$ (see equation 1.13 in section 1.3), which in this case is $(1/R_t)^{10}$. For low NPIs, with $R_t = 3$, this probability is 0.0017%; for medium NPIs, with $R_t = 1.4$, it is 3.5%; and for high NPIs, with $R_t = 1.05$, it is 61%. The theoretical prediction for low NPIs matches the value reported in table 5.1, but the probability of extinction in simulations is lower than the theoretical value for medium and high NPIs. This is partially because this model diverges from the panmictic case due to transmission between different regions; partially because the dynamics for low $\beta$ and $\mu$ are slow, so extinction may take longer than 90 days; and partially, in the case of high NPIs, because the value of $R_t$ is very close to 1, which is known to interfere with the validity of theoretical results of the SIR model [56].

Although the prophylactic drug is assumed to be effective enough to bring $R_t$ below 1 when $R_0 = 3$, the probability of extinction under low NPIs remains very low even for large stockpiles of the prophylactic drug. This is because the pathogen is likely to be exported to other regions before the prophylactic drug extinguishes the epidemic, due to the high travel rates and high $R_t$ in other regions. Therefore, containment is very likely to fail. For higher levels of NPIs the pathogen is less likely to be exported out of the initial region in the first one to seven days of the simulation, and so the prophylactic drug can have a greater effect on probability of extinction.

When NPIs are medium or high, increasing the stockpile of the prophylactic drug can greatly increase the probability of extinction. With no prophylactic drug, the probability of extinction with medium NPIs is very low (0.4–2%). When it takes seven days to deliver the drug, the probability of extinction with a small stockpile is only slightly higher (3–5%), whereas a very large (100 million) stockpile of the prophylactic drug can lead to a probability of extinction of 26–32%. With high NPIs the probability of extinction is around a half (47–54%) when no drugs are distributed, and, when distribution takes seven days, it is not detectably different (46–52%) for a small stockpile, but reaches 80–93% for larger stockpiles. The seven-day delay can allow the pathogen to escape containment, as shown by the much higher probabilities of extinction when the drug is distributed on day one (see figure 5.2). A large stockpile of prophylactic drugs can therefore be very helpful to containment efforts, but the probability of success depends on the speed of the response, and on any non-pharmaceutical interventions that are used in concert with the drug. Prophylactic drugs or medium NPIs are very unlikely to achieve pathogen extinction on their own.

## Late detection



5.6

The second scenario that I investigate involves several medium-sized outbreaks ongoing in different locations at the time that the pathogen is detected. I assume that, at the time of detection, there are 10 random clusters containing 50 infective individuals each. In this scenario, containment of the outbreak is very unlikely—under 2% in all cases. However, it is possible to delay the outbreak with the use of prophylactic drugs and NPIs, buying time for other interventions such as vaccines to be brought to bear.

Simulating the epidemic with low or medium NPIs and multiple initially infected regions is very slow, due to the large number of infection events. I therefore

focus on the case with high NPIs. Because the initial regions are randomised, the variance in the simulated epidemics is very high, so more simulations are needed. In this case I use 5120 simulations for each stockpile size. In figure 5.3 I show the cumulative infections over 150 days under high NPIs and a range of stockpile sizes. Although the prophylactic drug is unlikely to enable the containment of the invading pathogen, it can significantly slow its spread, especially when stockpiles are large. After 150 days, the median cumulative size of the outbreak with a stockpile size of 100 million is 7,900 infections (IQR 5,700–13,000). The outbreak with no drugs reaches this size in a median of 58 days (IQR 53–63), while the outbreaks with stockpile sizes of 15 million and 50 million reach it at day 70 (IQR 63–78) and day 97 (IQR 84–112) of the simulation, respectively. So a stockpile of 15 million doses delays the time for the epidemic to reach this median size by 12 days, a stockpile of 50 million doses delays it by 39 days, and a stockpile of 100 million doses delays it by 92 days.

I also simulated the case where the distribution of the prophylactic drug is extremely fast, beginning six days earlier—one day after the outbreak is detected. The results for this case are very similar to the case above—the median size for a stockpile size of 100 million after 150 days is 7,700 (IQR 3,600–17,000), and the median time for the outbreak with no drugs to reach this size is 57 days (IQR 52–62). The delay in this median due to a stockpile size of 15, 50, or 100 million is 10, 53, or 93 days, respectively. The similarity of these figures to those in the previous paragraph suggests that, although acting quickly may be important, a delay of up to a week in distributing the prophylactic drug is not likely to have a large impact on the timing of the epidemic. This contrasts with the power of a very fast response in attempts to *contain* an epidemic, as discussed in section 5.5 above.

The effect of changing the number of clusters that contain a population of infective individuals when the pathogen is first detected is illustrated in figure 5.4.

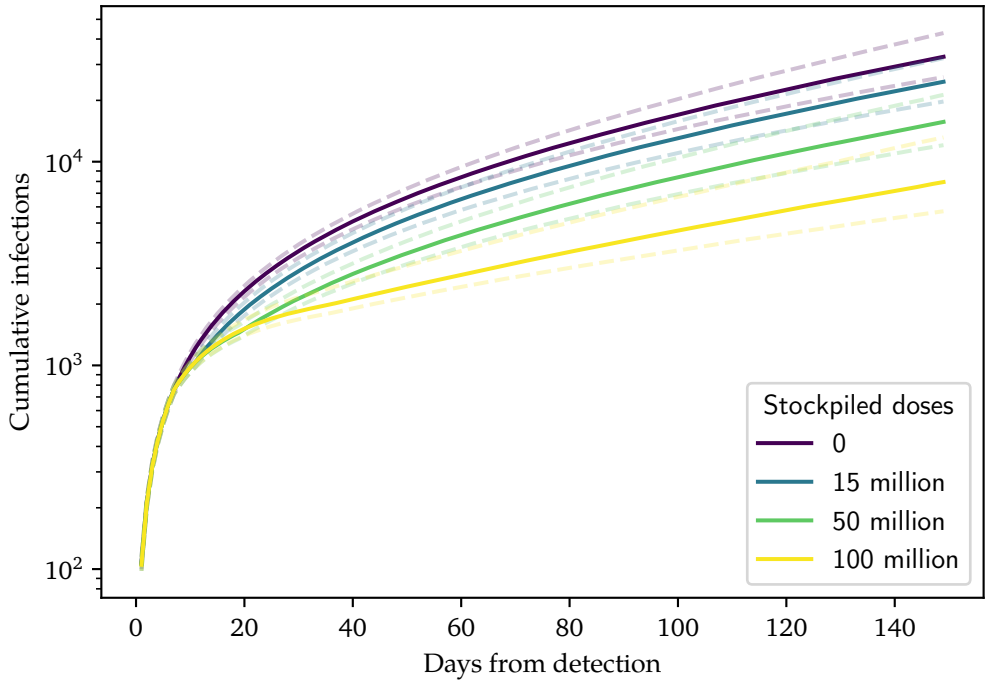



**Figure 5.3** A plot showing cumulative infections in simulated epidemics over time. The simulation starts when the pathogen is first detected. The model parameters are generated from Facebook, UKHSA, and ONS data for the week beginning 2020-11-16, during the November 2020 lockdown in England. The initial ten populations that the pathogen invades are randomised in each iteration of the simulation. The solid lines show the median simulated values under various sizes of prophylactic drug stockpiles. The dashed lines show the first and third quartiles. The drugs are distributed to the regions in which the pathogen was first seeded, beginning at day seven and ending when the stockpile runs out.

When 20 clusters begin with 50 infective individuals each, the prophylactic drug can significantly delay the epidemic, but a larger stockpile is needed to achieve similar results. The median size for a stockpile size of 100 million after 150 days is 30,000 (IQR 24,000–40,000), and the median time for the outbreak with no drugs to reach this size is 92 days (IQR 85–99). The delay in this median due to a stockpile size of 15, 50, or 100 million is 9, 18, or 58 days, respectively. When *all* clusters begin with 50 infective individuals each, then the stockpile sizes considered are far too small to appreciably affect the course of the epidemic, with 100 million doses delaying the outbreak by only eight days. These results demonstrate the importance of disease surveillance during a pandemic—the same stockpile of prophylactic drugs can be used to achieve a greater delay in the epidemic peak when the pathogen is detected before it has spread widely across the country.

## Conclusions ⬡ 5.7

In this chapter I have shown that a stockpile of an effective prophylactic drug could have had a large impact on the spread of SARS-CoV-2 in England, so long as it was implemented together with non-pharmaceutical interventions. For an outbreak detected early, the use of a large drug stockpile could have acted in concert with NPIs to contain the outbreak. For a pathogen detected once it had already spread to several regions, the distribution of a prophylactic drug could have delayed the course of the epidemic. In either case, a stockpile of such a drug would have aided the public health response to the invading pathogen.

Although the model used in this chapter was parameterised using data on the spread of SARS-CoV-2, its results should apply more widely to highly transmissible pandemic pathogens. The particular values of $\beta$ and $\mu$ will vary between pathogens, but the same general conclusions will hold: a combination of pharmaceutical and non-pharmaceutical interventions is required to reliably contain

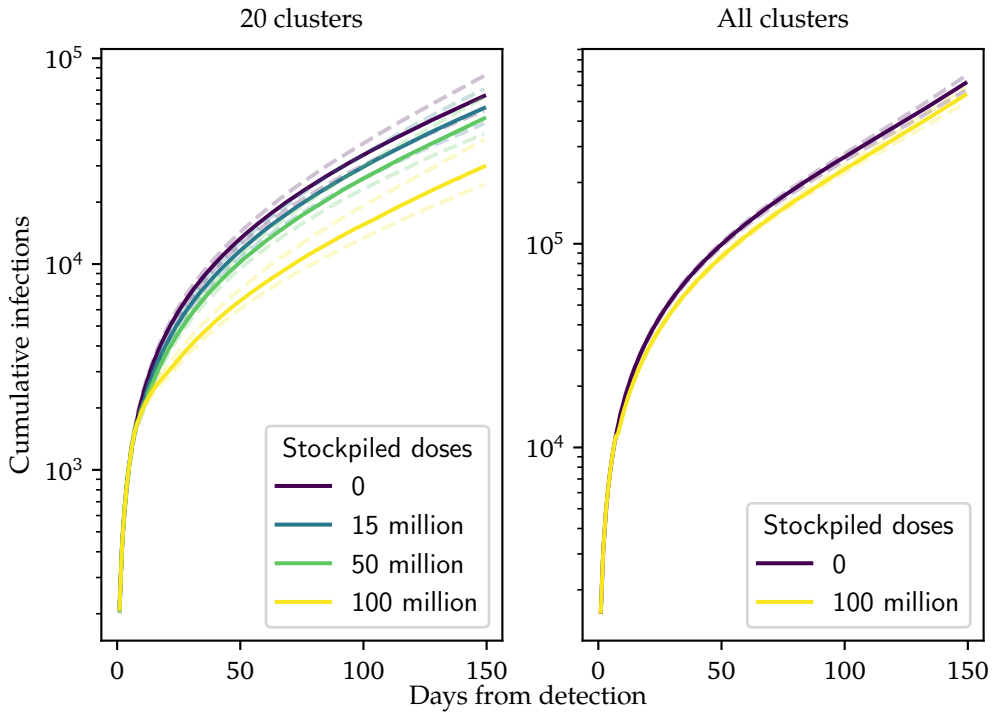



**Figure 5.4** Plots showing cumulative infections in simulated epidemics over time, for different numbers of clusters assigned to contain 50 infective individuals when the pathogen is first detected. The model parameters are generated from Facebook, UKHSA, and ONS data for the week beginning 2020-11-16, during the November 2020 lockdown in England. The solid lines show the median simulated values under various sizes of prophylactic drug stockpiles. The dashed lines show the first and third quartiles. Intermediate stockpile sizes are omitted in the second plot for clarity. The drugs are distributed to the regions in which the pathogen was first seeded, beginning at day seven and ending when the stockpile runs out. The more clusters which have an initial population of infective individuals, the less effective the prophylactic drug is in delaying the epidemic.

outbreaks, and prophylactic drugs can be used to slow epidemics. Some control objectives require acting very quickly, while others can be achieved with a small delay. As discussed in the introduction, many of these results have been previously demonstrated in the context of pandemic influenza. In general, preparation for the invasion of a pandemic pathogen can greatly aid the response to it, including granting extra time for interventions such as vaccines to be developed.

Application of these results to pandemic preparedness in practice requires the availability of effective prophylactic drugs against a variety of PPPs. The effectiveness of oseltamivir against pandemic influenza is disputed, and effective antivirals against SARS-CoV-2 were not developed until late in the pandemic [276–279]. In future, however, it may be possible to speed the development of prophylactic drugs by finding new ways to fast-track clinical development [23]. It may also be possible to formulate broad-spectrum drugs that can be used for prophylaxis against whole categories of pathogens, such as coronaviruses, or, more ambitiously, RNA viruses in general [280]. The results presented in this chapter support the idea that this would improve pandemic preparedness.

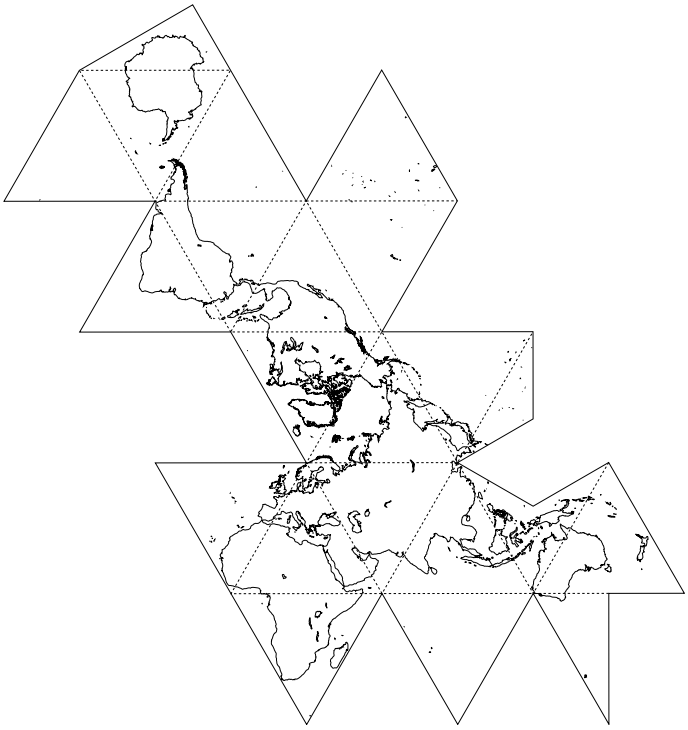
The epidemiological model used in this chapter is a metapopulation model, requiring the assumption that the population is panmictic within a single movement cluster. This is clearly a simplification of the true contact dynamics, and further work is necessary to fully understand how the clustering of human movement relates to pathogen transmission. Some exploration of this topic is performed in the Gibbs et al./ paper for which the clusters used in this chapter were originally calculated [119]. Future work could also incorporate a phylogeographic approach, using viral genetic data to better understand the relationship between movement clusters and transmission clusters [281, 282]. I have also assumed that prophylaxis is distributed to all members of a targeted region, independent of symptoms or contact histories. Studies of influenza prophylaxis have shown that targeting individuals known to have been exposed to the virus is an effective con-

trol strategy [252, 253]. The model presented above could be extended to explore the effect of targeted post-exposure prophylaxis in the case of SARS-CoV-2.

The SIR model used in this chapter assumes that natural immunity granted by SARS-CoV-2 infection is long-lasting. Reinfections with SARS-CoV-2 are, however, known to occur [283]. Using the data sources discussed in this chapter to parameterise an SIRS (susceptible-infective-removed-susceptible) model would allow for an analysis which includes waning natural immunity. The SIR model also assumes that there is no latent period between exposure to SARS-CoV-2 and infectiousness. This could be modified using an SEIR (susceptible-exposed-infective-removed) model, which would likely alter the time taken to reach the outbreak sizes discussed in section 5.6 above. The relative effectiveness of different stockpile sizes would likely be similar in the SEIR model and in the SIR model. The effects of certain NPIs could also differ in the SEIR model, with interventions involving testing allowing the isolation of exposed individuals before infectiousness [284, 285]. This could alter the effects of targeted post-exposure prophylaxis in the model.

In this chapter, I have made extensive use of data sets compiled during the COVID-19 pandemic in England. In general, the UK government and other bodies have gathered a large amount of data on the spread of SARS-CoV-2 through Great Britain, whose island geography, large epidemic, and high vaccination levels make it an interesting case study in the dynamics of pandemic pathogen spread [240, 246, 257, 286]. There is a great deal of further work to be done exploring these sources of data, drawing lessons to support the response to future pandemics.





A Dymaxion map of the world [287]. The Dymaxion map is a projection created by Buckminster Fuller, whereby the sphere of the Earth is projected onto a icosahedron, which is then folded out into its net. This projection is often used to emphasise the continuity of the landmasses of the planet Earth, and the potential for biological populations to migrate between them.

# 6 Conclusion

Mathematical models can help us understand the dynamics of pandemics—both how they arise, and how to prevent or ameliorate them [288–290]. In this thesis, I have used a range of different metapopulation models to investigate these dynamics. Broadly, these models can be used to explore factors contributing to the *risk* from pandemics, and to evaluate methods for their *control*.

## Risk 6.1

In chapters 2 and 3, I discussed the wide range of outbreaks that have been classed as pandemics. Different pandemics have affected different groups of people, and have led to different outcomes in terms of death, illness, and changes to human cultures and civilisations [4, 5, 8–16, 122–124]. To understand pandemic risk, we must be able to characterise this variety of outcomes, and to account for how it comes about.

In chapter 3, I discussed a range of quantitative definitions of the term 'pandemic', and their use in epidemiological modelling. The definition used determines the kinds of outbreak (in terms of size, spatial structure, etc.) that are included in the class of pandemics. We saw that this can affect the interpretation of disease model outputs in important ways. I focussed on pandemic definitions that varied in their application of spatial criteria, since the requirement for a pathogen to become widespread is the only universal element of pandemic definitions [2].

Using a Markov chain method to find the final states of a metapopulation

model of disease outbreaks, I showed that epidemiological parameters such as pre-existing immunity and heterogeneity in transmission have complex effects on the probability of outbreaks with different spatial extensions. This means that the effects of these parameters on the risk of a pandemic can depend sensitively on the definition of 'pandemic'. Importantly, quantitative changes in the definition of 'pandemic' can result in qualitative changes in the effects of epidemiological parameters on pandemic risk.

I discussed one of these parameters, cross-immunity, in detail in chapter 2. The same Markov chain method can be used to show how varying travel rates between regions and varying cross-immunity between strains of a pathogen can give rise to a wide variety of pandemic outcomes. In particular, the presence of cross-immunity can lead to a non-monotonic relationship between pandemic risk and other epidemiological parameters, such as the basic reproduction number. In chapter 3 I extended these results to show that the existence of these non-monotonic effects can depend on pandemic definitions, and that non-monotonic effects can also be caused by heterogeneity in transmission between regions.

Taken together, these chapters show that pandemics constitute a highly complex and heterogeneous class of infectious disease outbreaks. The risk of pandemics is influenced by many factors, and these factors have different effects on the various kinds of outbreaks that make up the broad class of 'pandemic'. These effects are not always monotonic, and it may be difficult to distinguish the effects of one parameter from another. To account for these complexities, pandemic preparedness researchers must communicate clearly about the risk of various different outbreak outcomes, using explicit and consistent definitions of terms wherever possible.

# Control ⬡ 6.2

In chapters 4 and 5 I examined how human intervention can reduce the risk of pandemics, focussing on spatial containment of an invading pathogen. Because the most important criterion distinguishing pandemics from other outbreaks is their wide spatial extension, it is natural to consider spatially targeted control as a method of pandemic prevention or amelioration when the range of a pandemic pathogen is expanding [2]. In chapter 4 I introduced several methods for spatial targeting of prophylaxis, and in chapter 5 I considered a potential application of spatial control in containing an invading pandemic pathogen in England.

Spatial prophylaxis strategies require prioritisation of individuals to receive doses of a prophylactic drug or vaccine based on their location [68, 72, 86, 87, 117, 209–213]. Methods for this prioritisation include strategies that target those regions most at risk of importing a pathogen, and strategies that target regions with high centrality on the network of travel rates between regions. In chapter 4, I used a simple infectious disease epidemic model to compare a risk-targeting strategy to two different centrality-targeting strategies based on betweenness centrality and random walk percolation centrality. I found that the relative effectiveness of these strategies in reducing the total number of infections varies with travel rates, the structure of the travel network, the basic reproduction number of the pathogen, and the availability of the prophylactic drug or vaccine. When a pathogen has high spreading capacity, or when the availability of the prophylactic drug or vaccine is limited, centrality-targeting strategies can be more effective than the more commonly used risk-targeting strategies. However, centrality-targeting strategies proved only to be effective in some kinds of networks. In networks with many weak connections between clusters of regions, each containing many small regions connected to a highly populous region, risk-targeting was always the most effective strategy.

In chapter 5, I used data on the spread of SARS-CoV-2 in England to parameterise a metapopulation model, with the boundaries of different regions and the travel rates between them derived from movement data collected by Facebook [118, 119]. I then simulated the arrival of a pandemic pathogen in England under a variety of situations, and evaluated the effect of distributing a prophylactic drug to prevent the spread of this pathogen in affected regions. I found that a combination of pharmaceutical and non-pharmaceutical interventions was required for the chance to contain the outbreak to exceed 50%, even under the best-case scenario of early pathogen detection. When the pathogen is detected later in the outbreak, I found that the use of a large stockpile of a prophylactic drug can significantly slow the epidemic, allowing extra time for other controls to be implemented. This corroborates existing findings on the use of prophylactic antivirals to combat pandemic influenza [251–255].

This work has shown that spatial control has promise for reducing pandemic risk. With the appropriate strategies, the effectiveness of prophylaxis can be greatly increased when the spatial structure of the population is taken into account. This spatial structure determines which strategies are likely to be effective, and it is important that spatial control be deployed with sensitivity to the spatial dynamics of the target population. In combination with other tools for pandemic preparedness, spatial prophylaxis can be used to prevent the emergence and spread of pandemic pathogens.

## Spatial models and pandemics  6.3

The work presented in the preceding chapters makes use of a range of different epidemiological models. In chapter 1, I introduced the stochastic SIR model of pathogen spread, in both its panmictic and metapopulation formulations. In chapter 2, I developed a novel Markov chain method for calculating the probab-

ilities of various final states of regions in the metapopulation, presenting various use cases for this model in chapters 2 and 3. In chapter 4, I formulated a different stochastic metapopulation model, making a range of simplifying assumptions, based on SIR dynamics, that would allow me to run many simulations of an outbreak and determine the effect of a variety of control strategies. Finally, in chapter 5, I parameterised a full stochastic SIR metapopulation model using epidemiological data, allowing me to produce realistic time series representing the course of an epidemic in England. Each of these models has included different assumptions and parameters in order to achieve different research goals.

In section 1.1, I discussed the roles that models at different levels of complexity can play in theoretical epidemiology. The models used in this thesis have varying levels of complexity. The Markov chain method that I presented in chapter 2 removes all time-dependence from a stochastic SIR metapopulation model, yielding a highly simplified framework conducive to generating theoretical insights of the kind set out in chapters 2 and 3 regarding pandemic risk. On the other end of the spectrum, the rich data available on the spread of COVID-19 in England allowed me to parameterise a full, time-dependent stochastic SIR metapopulation model in chapter 5. This model was well suited for a more specific analysis of the potential impact of various pandemic preparedness policy decisions. The simplified time-dependent model used in chapter 4 has intermediate complexity, and facilitated theoretical work that incorporated several sources of empirical travel data.

Although they have various differences, all of the models used in this thesis are metapopulation models. In section 1.4 I argued that models of pandemics must incorporate geographical structure, and I showed how this can be achieved with a metapopulation model. Metapopulation models represent geographical space as set of distinct regions, usually with simplifying assumptions made about the dynamics within these regions—for instance, that individuals within a single region are panmictic. This representation of space matches well with epidemiological

data, which are often reported in terms of administrative regions such as health zones; and disease control, which is often allocated to the same administrative regions [53, 68, 87, 291–293].

Metapopulation models provide a highly flexible framework for coupling the dynamics of pathogen spread to the spatial structure of the host population. However, host populations can have highly complex spatial structures and dynamics [294, 295]. Communities of organisms on the planet Earth exist on a continuous surface, containing a range of habitats only some of which are suitable for any given organism [296, 297]. Further, organisms have different modes in which to travel over this surface [298, 299]. Humans in particular move in highly complex ways, travelling for many different purposes using many different methods [127, 178, 185]. The dynamics of pathogen spread in different populations of organisms can be coupled, as in the case of vector-borne diseases [68, 300, 301]. Population density can vary at many scales, depending on physical and social geography, and the rate of travel between any two points in space depends on all of the above factors and more [68, 143, 233]. In the face of such complexity, it is futile to expect any single model type, including metapopulation models, to represent every epidemiologically important feature of geographical space. Instead, we must use a range of different models to understand different aspects of the spatial dynamics of pathogens.

Some spatial epidemiological models treat space as fully continuous, or as a lattice meant to approximate continuous space, with the host population represented in terms of its density as a scalar field over geographical space [302–304]. This can produce insights into the epidemiological importance of certain geographical features, such as roads [302]. Models of this type have also contributed to understanding travelling wave-like phenomena in disease dynamics [305]. However, representing the movement of human populations in continuous-space models can be challenging, since human movement diverges considerably from

Brownian motion [306]. This can lead to perverse outcomes from continuous-space models, such as a study which treated the COVID-19 pandemic as a probability cloud spreading continuously from China through Europe to the Americas, thereby including a large population of infective individuals in the middle of the Atlantic ocean [303].

Infectious contacts in humans and other animals are partially determined by social structures. For instance, a person infected with influenza is much more likely to infect their co-workers and members of their household than strangers [307, 308]. Social structures can be represented by networks, with edges representing the relationships between different individuals [309, 310]. Many social relationships are also spatial in nature—two people living in the same city are more likely to be co-workers than people who live far apart, for instance. Social and spatial structures can therefore be represented simultaneously using networks, where the weight of edges between individuals is influenced by both social and spatial variables [311]. These network models provide a highly general structure for understanding pathogen dynamics—for instance, the agent-based models discussed in section 1.4 can be thought of as a kind of network model [92]. Networks can also be amenable to theoretical analysis [98, 312, 313]. The generality of network models can come at the cost of high complexity, with the number of edges in the network scaling with the square of the number of individuals represented [314].

Among these various options, the reasons for using metapopulation models to model pandemics include their tractability and the ease of incorporating certain kinds of epidemiological data. Epidemiologists are likely to continue to use metapopulation models in pandemic research, and my work has generated insights into their structure and potential uses. I showed in chapter 3 that using different pandemic definitions can greatly alter the interpretation of metapopulation models of pandemic risk. Researchers using metapopulation models of pandemics

should clearly state their definition of 'pandemic', and understand the effect of that definition on their conclusions. I demonstrated in chapters 2 and 3 that metapopulation models can capture complex non-linear and non-monotonic effects. This shows that metapopulation models are flexible, and can model a wide range of phenomena—but also that researchers must be aware that the behaviour within one part of parameter space may not generalise monotonically to other parts of parameter space. Finally, I showed in chapter 4 that the structure of the travel network component of a metapopulation model can greatly affect the relative effectiveness of different strategies for disease control in that model. Researchers using metapopulation models to generate recommendations for the spatial control of a pathogen should be aware of the structure of the travel network, and how it enables various control strategies to be implemented more, or less, successfully.

## Future work 6.4

Much of the work presented in this thesis predates the first detection of the novel coronavirus that would go on to be named SARS-CoV-2. The projects that comprise chapters 2 and 3 were both conceived before December 2019. It was the global effort to respond to the COVID-19 pandemic, and the rich sources of data generated by this effort, that inspired the turn to methods of control in chapters 4 and 5. I have, however, focussed on developing pathogen-agnostic methods. The world will need to remain prepared for pandemic pathogens other than SARS-CoV-2, and, as discussed in chapter 1, there is no guarantee that future pandemics will resemble past ones—or, indeed, current ones. This thesis provides a contribution to the modelling toolkit available to those working towards preparedness for the next pandemic pathogen.

Chapter 3 constituted a contribution towards understanding the importance of pandemic definitions in modelling pandemic risk. This could form part of

a larger piece of work aiming to provide comprehensive guidance on various definitions of 'pandemic' and their applicability in different models, including non-metapopulation models. Such a piece of work would be wide in scope, and require collaboration from researchers in theoretical epidemiology, public health, and health governance. Such guidance would constitute a valuable contribution to the field of pandemic preparedness, and could exist alongside existing WHO guidance regarding the PHEIC designation. As noted in chapter 3, pandemic definitions are little discussed in modelling studies, despite their demonstrable importance in interpreting model outputs.

The Markov chain method used in chapters 2 and 3 is a highly general method for finding the final probabilities of epidemics in a stochastic SIR metapopulation model. There are many possible further applications of this method to understanding pandemic risk. Markov chains in general are amenable to mathematical analysis, so this method is a promising subject of future theoretical work [78, 154, 315].

In chapter 4 I showed how travel networks can be used to target prophylactic control of a pandemic pathogen, and how the structure of these networks determines the relative effectiveness of different spatial control strategies. Future work could link the effects of different control strategies to general graph-theoretic properties of the example networks used, which would allow the conclusions of this chapter to be applied to a wider range of disease control problems. The set of control strategies tested could also be expanded, and the model could be coupled to a model of the logistics of prophylaxis distribution in order to generate concrete policy recommendations. These recommendations would have to take into account which bodies could administrate the prophylaxis distribution, and whether they would face legal or political constraints in assigning prophylaxis to specific regions, as the WHO might do in its advice to member states [316, 317].

In chapter 5 I made use of data gathered in the UK during the COVID-19

pandemic. The UK government and various other organisations within the UK have made public a large amount of data on the spread of SARS-CoV-2 in the country [118, 240, 246, 256, 257, 286]. Analysis of these data has the potential to generate many insights into the dynamics of a pandemic pathogen. Future work will be able to investigate a wide range of control strategies other than those examined in chapter 5, using both metapopulation models and other modelling frameworks. The results of these investigations should be made available to bodies such as the UK Health Security Agency, who can use this information to judge which pandemic preparedness systems are likely to be effective, and therefore how to allocate resources for pandemic preparedness.

My research adds to the large body of work suggesting that, as human activity reshapes the global ecosystem, large changes may occur in the dynamics of pandemic pathogens [21–23]. In order to respond to these changes, humanity will need institutions and technologies that can mount rapid responses to complex and developing situations. Much of the material on pandemic preparedness predating the COVID-19 pandemic deals exclusively with pandemic influenza. Many governments had an official pandemic influenza plan, but no public official plan for responding to any other pandemic pathogen [104–106]. With COVID-19, coronaviruses have been placed firmly on the list of pathogens with pandemic potential. However, we must not repeat the mistakes of the past—a world with a plan for pandemic influenza, pandemic coronaviruses, and no other pandemic pathogens will be only slightly better prepared than the world before the COVID-19 pandemic. Instead, a renewed effort must be invested in understanding the dynamics of pandemics in general, and creating robust strategies and technologies that can be used to react to any pandemic pathogen [23, 318].

# Appendices

# A Heat maps with a single scale

The heat map plots in chapters 2 and 3 included a variety of colour bar scales in order to show the variation in risk within each given scenario (e.g. a given value of the cross immunity $\alpha$). In this appendix, I reproduce these plots, starting from those in figure 2.2, with a single colour bar scale ranging from 0 to 0.7. This facilitates the comparison of risk across different scenarios.

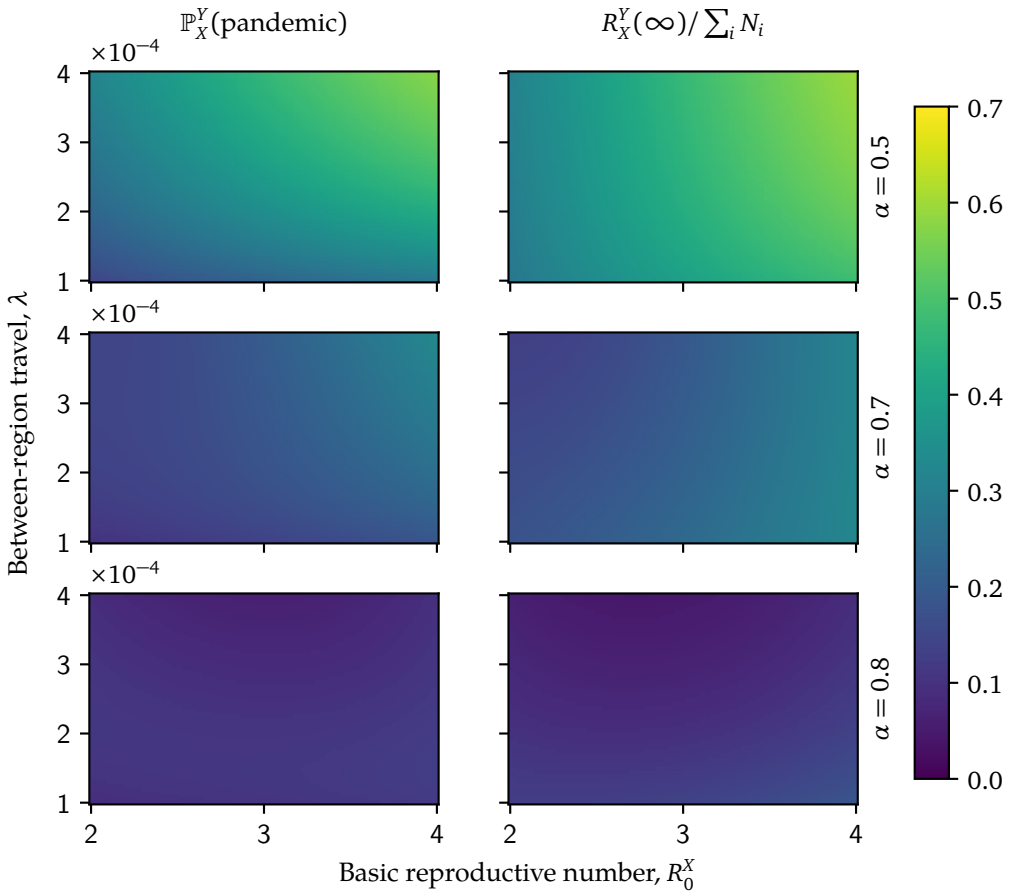



**Figure A.1** Plots showing how the behaviour of strain $Y$ in the two-region model varies with $R_0^X$, $\lambda$, and $\alpha$, when $R_0^Y = \frac{3}{2}R_0^X$, assuming that there has been at least one epidemic of $X$. The first column shows the probability of a pandemic of $Y$ as given by equation 2.24, and the second shows the total proportion of the population who are expected to become infected as given by equation 2.25. Each row gives the results for a different value of the cross-immunity, $\alpha$. For $\alpha = 0.5$, increasing $R_0^X$ or $\lambda$ always increases pandemic probability and expected outbreak size, whereas for $\alpha = 0.7$ increasing $\lambda$ increases pandemic probability while reducing expected outbreak size. For $\alpha = 0.8$ there is non-monotonicity with both parameters.

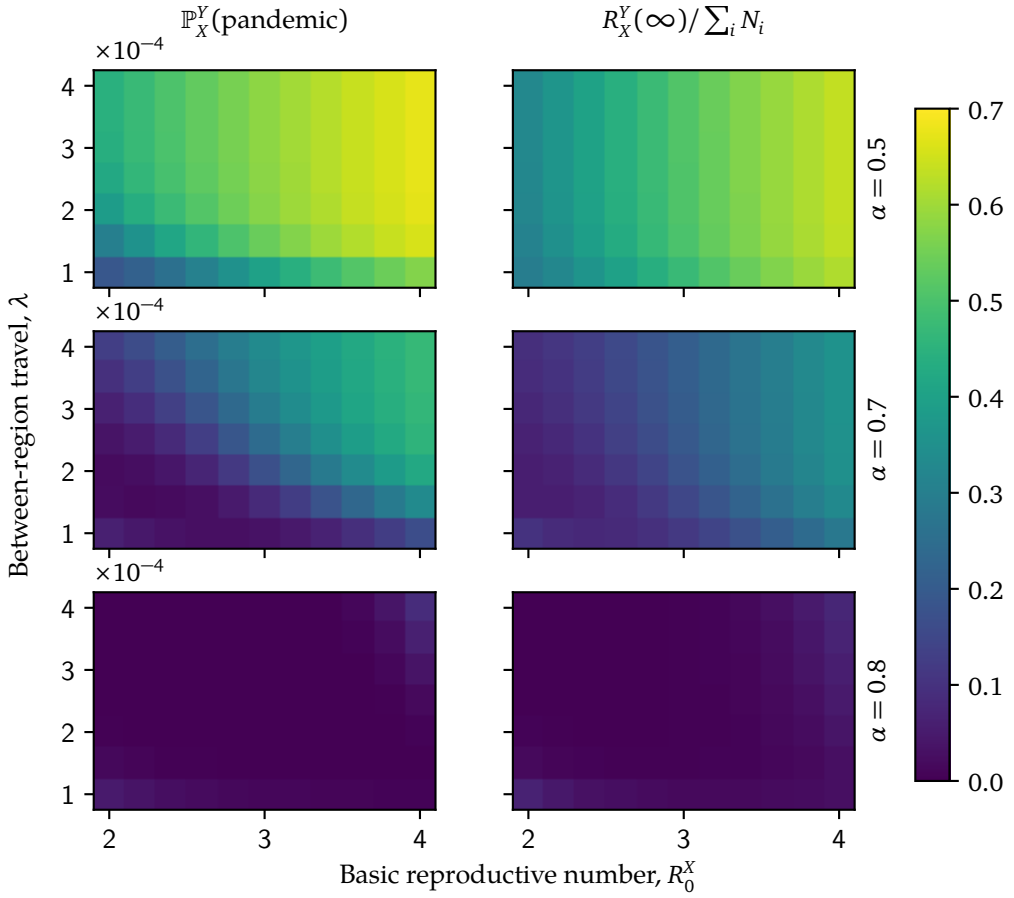



**Figure A.2** Plots showing how the behaviour of strain $Y$ in the ten-region fully connected network varies with $R^X_0$, $\lambda$, and $\alpha$, when $R^Y_0 = \frac{3}{2}R^X_0$, assuming that there has been at least one epidemic of $X$. The first column shows the probability of a pandemic of $Y$, given by the $TT\ldots T$ element of $P^Y_{\text{final};X}$ in equation 2.39. The second column shows the total proportion of the population who are expected to become infected, as given by equation 2.41. Each row gives the results for a different value of the cross-immunity, $\alpha$. Slight non-monotonicity with both $\lambda$ and $R^X_0$ is visible at $\alpha = 0.7$, and becomes much stronger when $\alpha = 0.8$.

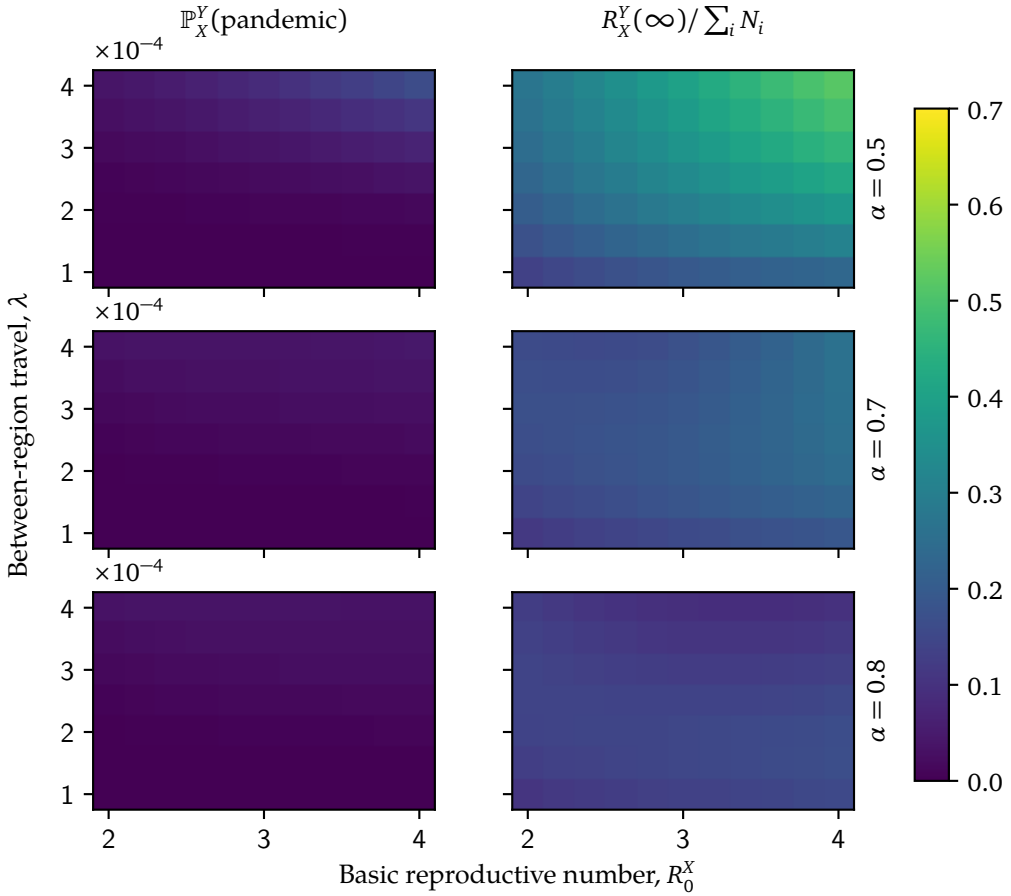



**Figure A.3** Plots showing how the behaviour of strain $Y$ in the ten-region star network varies with $R_0^X$, $\lambda$, and $\alpha$, when $R_0^Y = \frac{3}{2}R_0^X$, assuming that there has been at least one epidemic of $X$. The first column shows the probability of a pandemic of $Y$, given by the $TT\ldots T$ element of $P^Y_{\text{final};X}$ in equation 2.39. The second column shows the total proportion of the population who are expected to become infected, as given by equation 2.41. Each row gives the results for a different value of the cross-immunity, $\alpha$. Pandemic probability always increases with increasing $\lambda$, but the same is not true for the expected infected proportion, which is non-monotonic with $\lambda$ for $\alpha = 0.7$ and $\alpha = 0.8$.

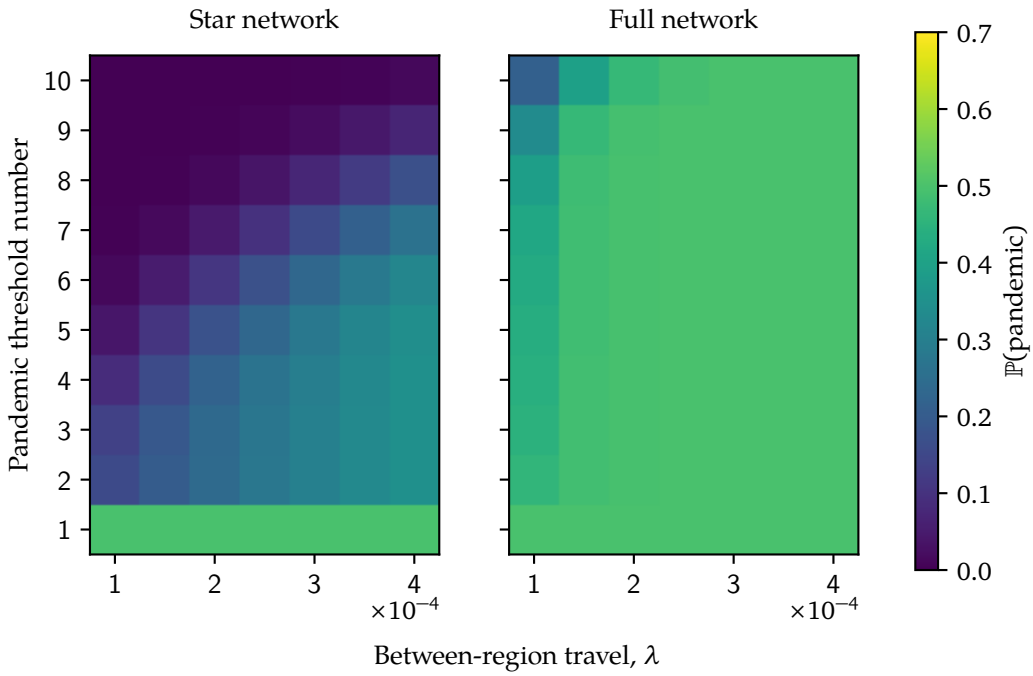



**Figure A.4** Plots showing how pandemic probability varies for a range of between-region travel rates under a range of transregional pandemic definitions, in the 10-region star and fully connected networks. The "pandemic threshold number" refers to the minimum number of regions that must experience epidemics before a pandemic is judged to have occurred. The pandemic probability is, in general, sensitive to the pandemic definition used, but the degree of sensitivity depends on network structure and travel rates. In the star network, the pandemic probability is, in general, highly sensitive to the pandemic definition used, while in the full network this sensitivity is significantly reduced at high travel rates.

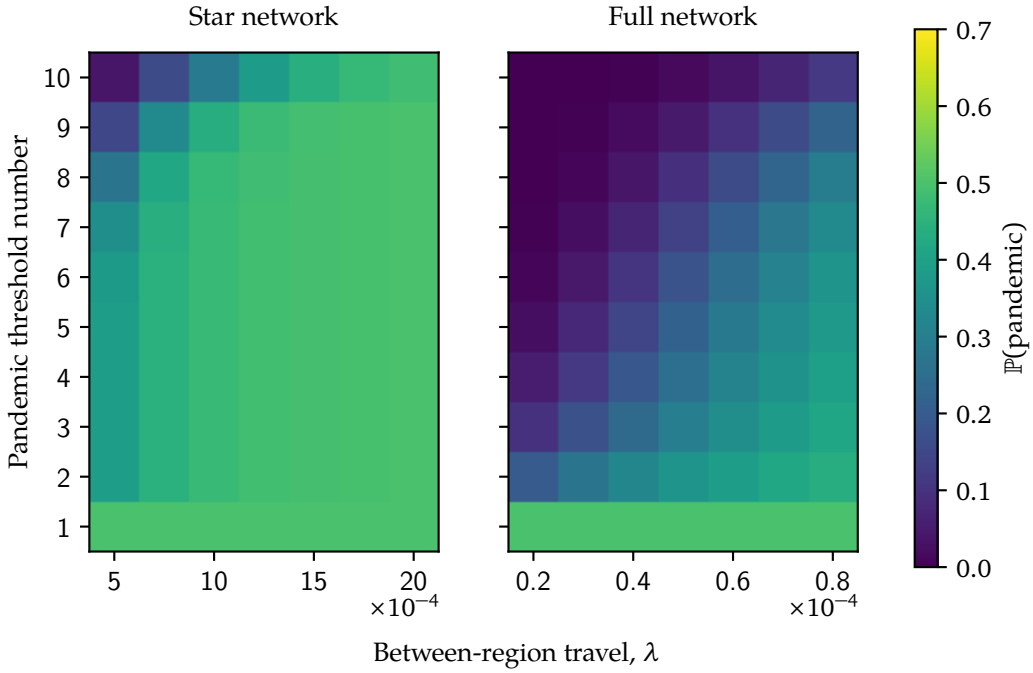



**Figure A.5** Analogous plots to figure 3.1, but with the travel rate for the star network increased by a factor of five, and that for the full network decreased by the same, so as to swap the total volume of travel in the two networks. We see that the star network with high travel rates shows similar behaviour to the full network at medium travel rates, but with more gradual variation as travel increases. The full network at low travel rates shows similar behaviour to the star network at medium travel rates, but with greater changes in pandemic probability as the pandemic threshold number changes.

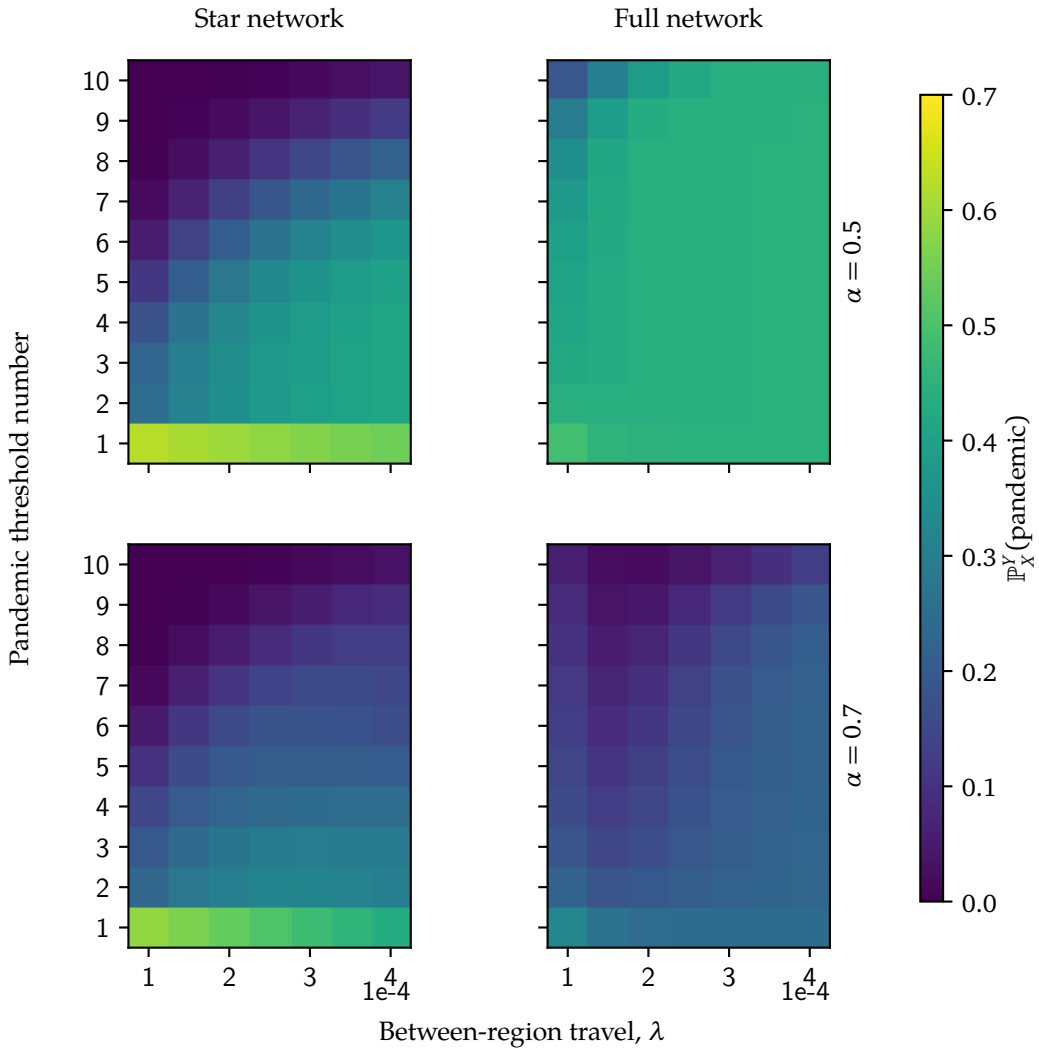



**Figure A.6** Plots showing how the probability $\mathbb{P}_X^Y(pandemic)$ of a pandemic of strain $Y$, given at least one epidemic of strain $X$, varies with travel rate and cross-immunity, under a range of transregional pandemic definitions, in the 10-region star and full networks. At moderate cross-immunity, the plots resemble those in figure A.4, though the pandemic risk is increased for the star network due to the higher value of $R_0^Y$, and decreased in the full network due to wide circulation of $X$. At high cross-immunity, pandemic probability is in general highly suppressed, and there are non-monotonic effects in $\lambda$.

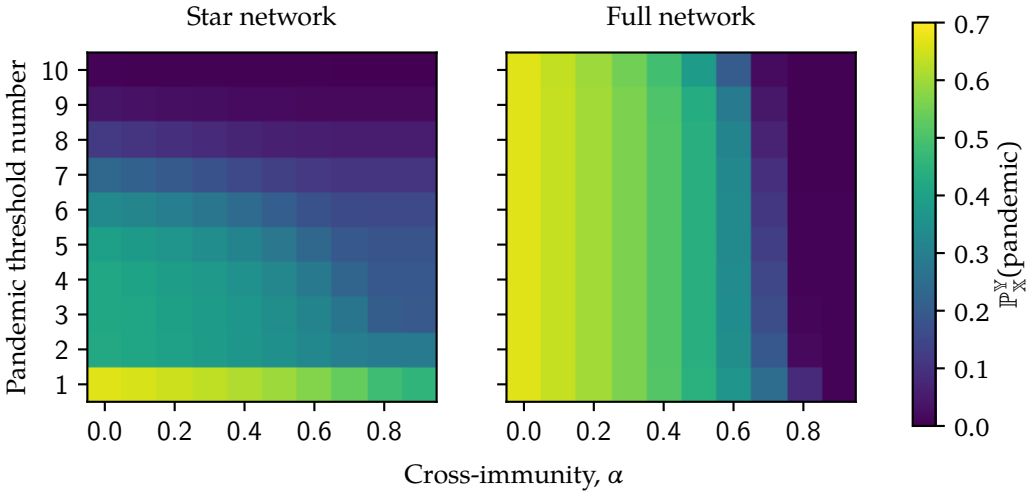



**Figure A.7** Plots showing how the probability $\mathbb{P}_X^Y(pandemic)$ of a pandemic of strain $Y$, given at least one epidemic of strain $X$, varies with cross-immunity under various transregional definitions. In the star network, the pandemic probability is sensitive to changes in definition at all levels of cross-immunity. In the full network the sensitivity of the pandemic probability to the pandemic definition used increases with cross-immunity, until the probability of any epidemic becomes very low.

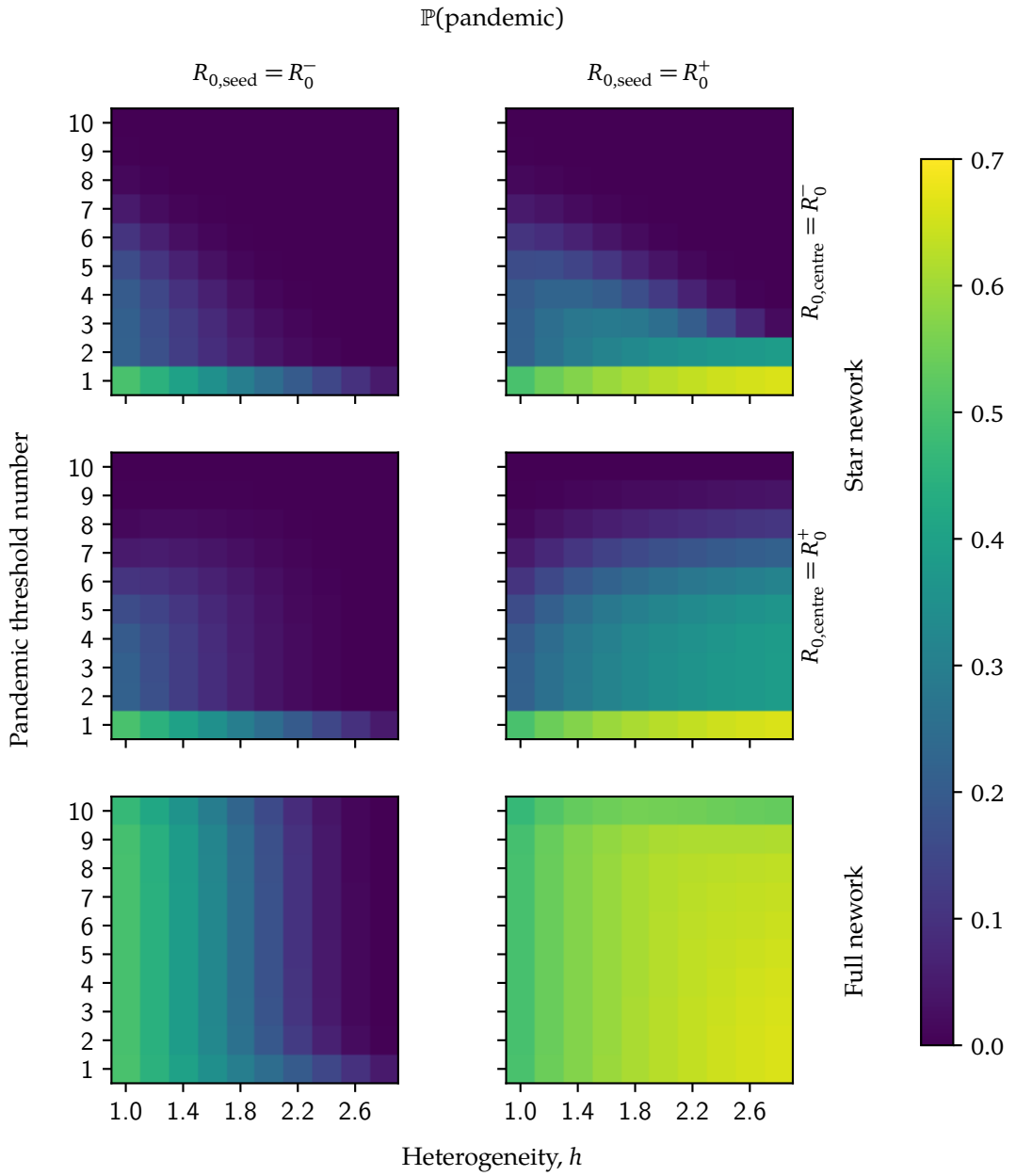



**Figure A.8** Plots showing how pandemic probability varies with heterogeneity $h$ in transmission rates, for different networks and initial states, under a range of different transregional pandemic definitions. When the outbreak is seeded in a lower-transmission region, increased heterogeneity always results in reduced pandemic probability. When the outbreak is seeded in a higher-transmission region, the effect of heterogeneity on pandemic probability depends on the network and on the pandemic definition, with non-monotonic effects occurring in the star network with $R_{0,centre} = R_0^-$, and in the full network when the pandemic threshold number is high.

# B Spatial control in network variations

This appendix contains supplementary figures to chapter 4, including plots showing the effects of prophylaxis allocation strategies in variations on the "joined grids" network, and plots showing the effects of NPI targeting strategies in the "joined grids", "four corners", and Northwest England commuter networks.

## Long-distance travel B.1

Equation 4.17, reproduced below, gives the weight $\lambda_{ij}$ of the edge between two regions within the same grid in the joined grids travel network (see figure 4.1).

$$\lambda_{ij} = \frac{\lambda}{(r_{ij})^a}, \tag{4.17}$$

where $\lambda$ is the travel rate between adjacent regions; $r_{ij} \geq 1$ is the distance between region $i$ and region $j$, as a multiple of the distance between adjacent regions; and $a$ is a gravity model parameter determining how quickly travel falls off with distance. The higher the value of $a$, the less long distance travel occurs. In section 4.4 and figure 4.2 I investigated pathogen control strategies on this travel network when $a = 2$. In figures B.1–B.3, I present similar plots to figure 4.2, but with increased values of $a$, in order to investigate the effect of long-distance travel on the effectiveness of different control strategies.

In general, decreasing the long-distance travel rates within grids decreases the

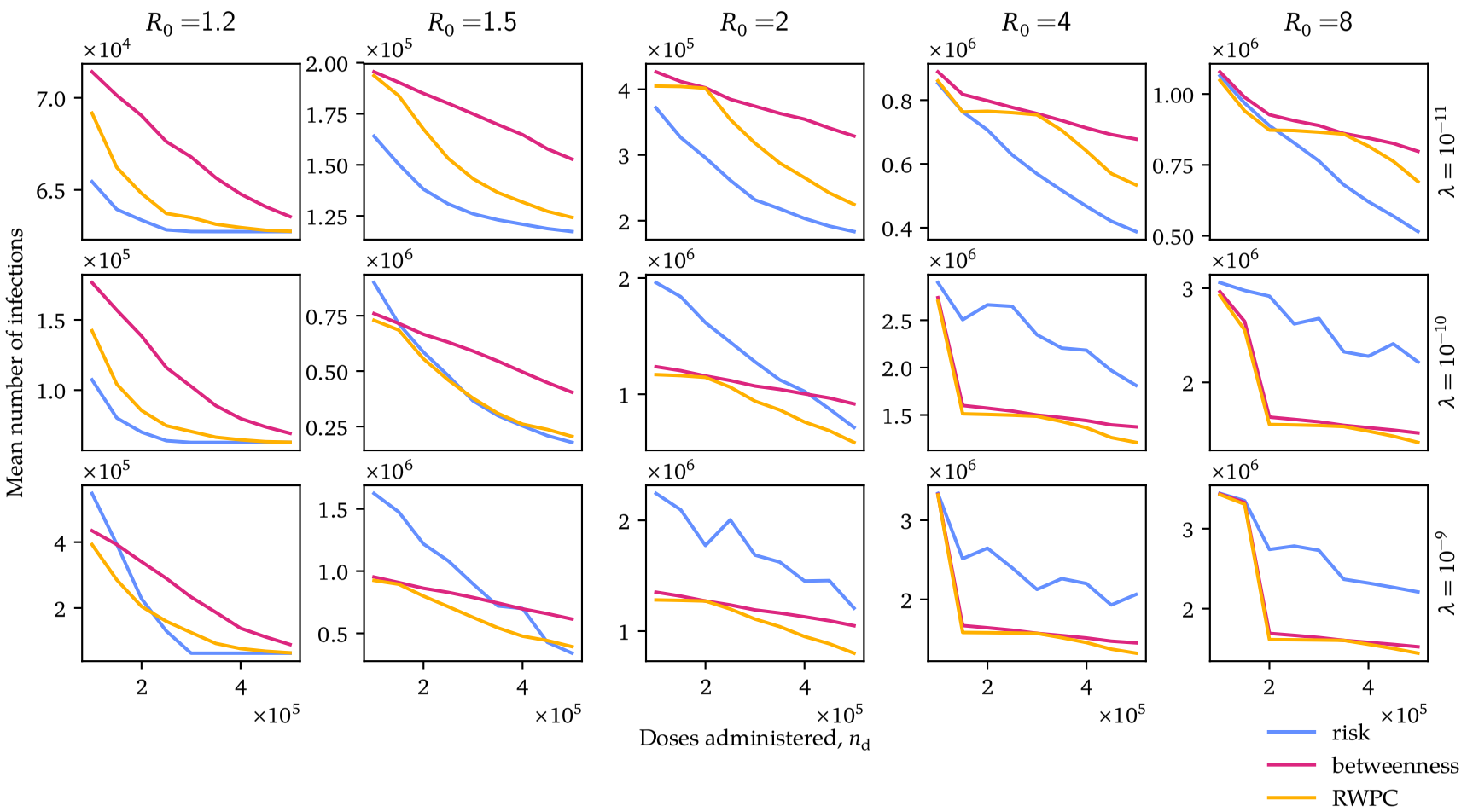



**Figure B.1** Plots of the expected total number of infections due to a disease outbreak against the number of prophylaxis doses administered according to one of three strategies, for five different values of $R_0$ and three different overall travel rates, in the joined grids network with the long-distance travel parameter $a = 3$.

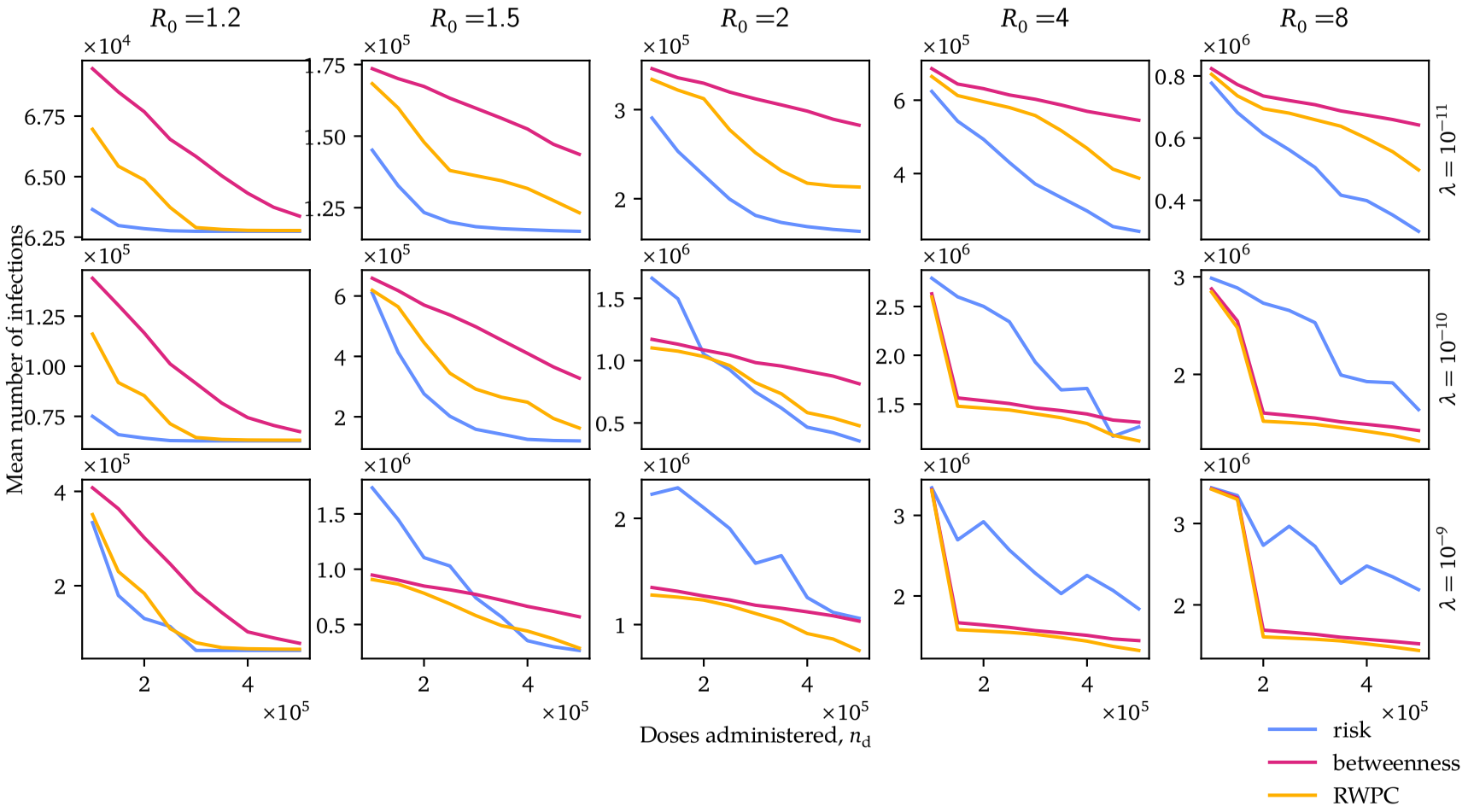



**Figure B.2** Plots of the expected total number of infections due to a disease outbreak against the number of prophylaxis doses administered according to one of three strategies, for five different values of $R_0$ and three different overall travel rates, in the joined grids network with the long-distance travel parameter $a = 6$.

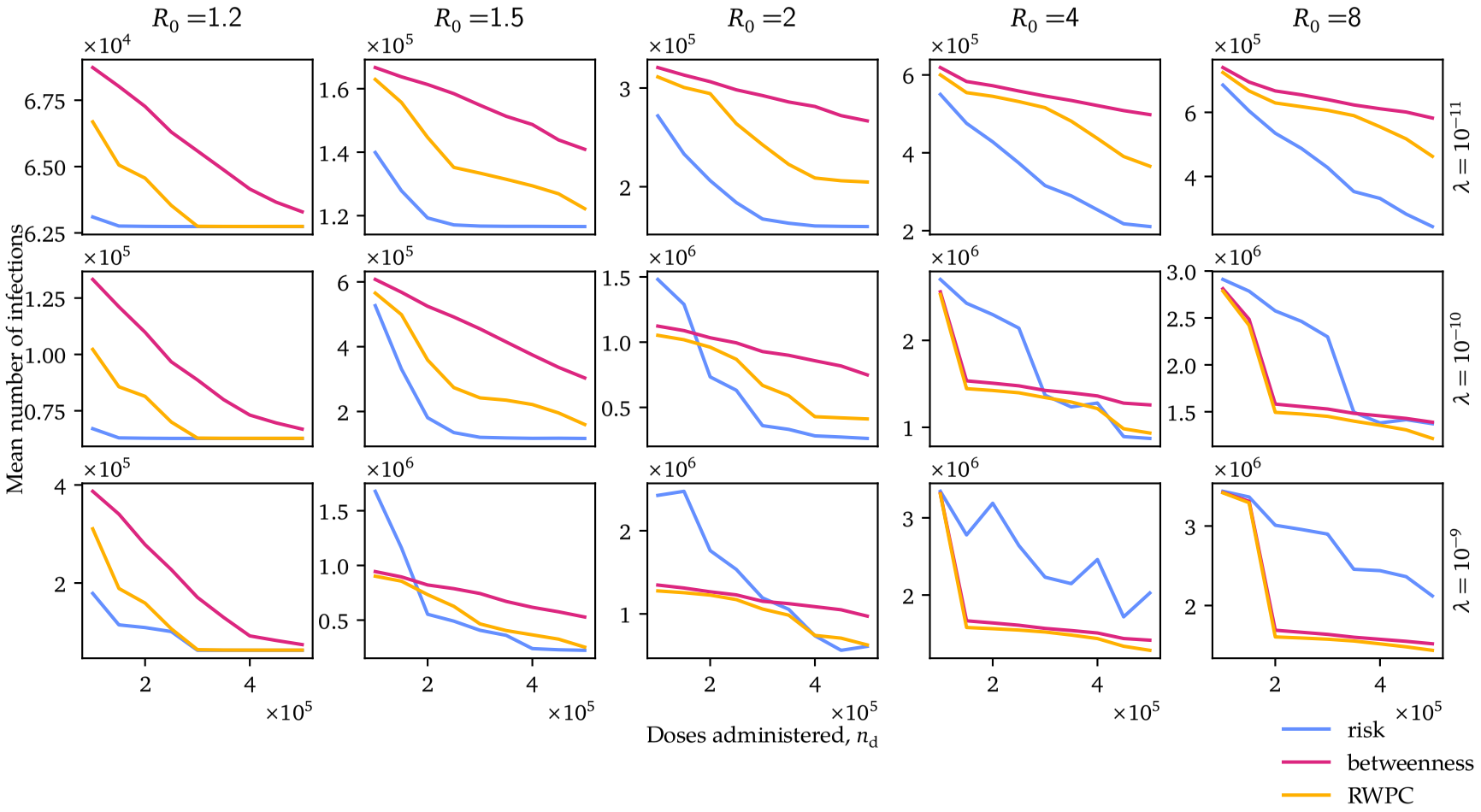



**Figure B.3** Plots of the expected total number of infections due to a disease outbreak against the number of prophylaxis doses administered according to one of three strategies, for five different values of $R_0$ and three different overall travel rates, in the joined grids network with the long-distance travel parameter $a = 12$.

overall expected number of infections. It can also affect the relative performance of different control strategies, increasing the efficacy of the risk-targeting strategy relative to the centrality-targeting strategies. For instance, when $R_0 = 2$, $\lambda = 10^{-10}$, and $a = 3$ (see figure B.1), the risk-targeting strategy is never more effective than the RWPC-targeting strategy—but when $a = 12$ (see figure B.3), the risk-targeting strategy is the most effective at most values of $n_d$, for the same values of $R_0$ and $\lambda$. However, taking each figure as a whole, the same general pattern holds as in figure 4.2—centrality-targeting is more effective than risk-targeting when $R_0$ and $\lambda$ are high, or when $n_d$ is low.

## Joined grids variations B.2

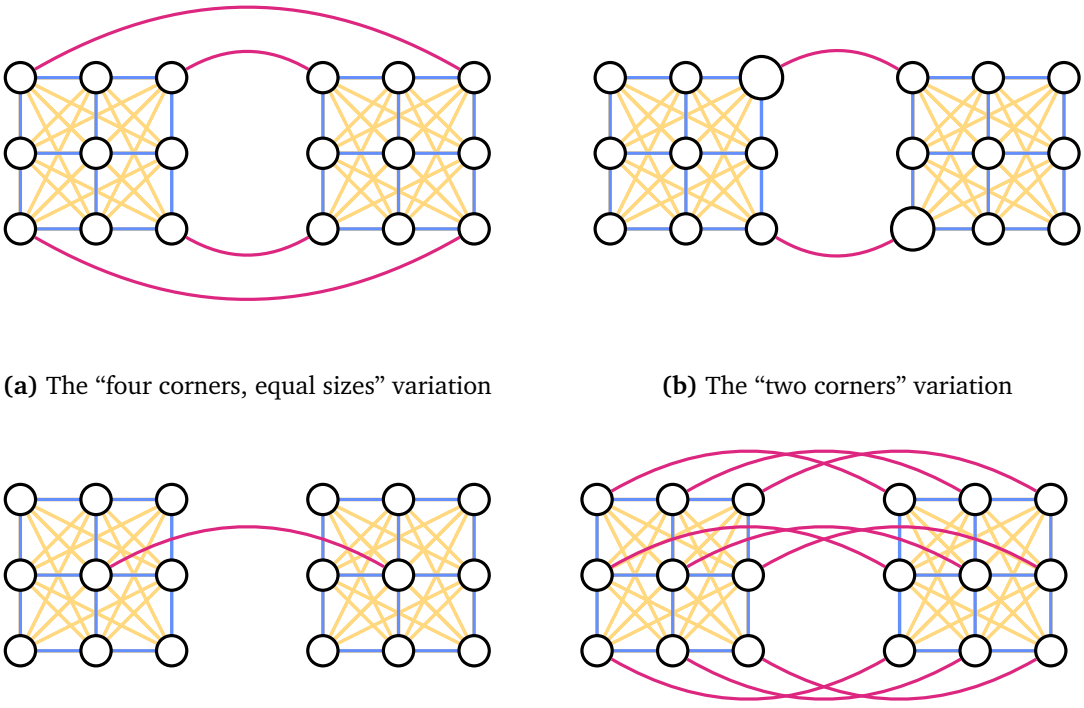
The "four corners" network is a variation on the joined grids network in which risk-targeting is always more effective than centrality-targeting, for the range of parameters considered in section 4.4. In figures B.5–B.8 I consider a range of variations on the joined grids network to illustrate that small changes to this network do not prevent centrality-targeting from being effective, and that the relatively large alterations in the four corners network are necessary to achieve this result.

The networks considered are illustrated in figure B.4. The within-grid travel rates are given by equation 4.18, reproduced below:

$$\lambda_{ij} = \frac{\lambda N_i}{(r_{ij})^a N}, \tag{4.18}$$

The travel rate along each between-grid route is $\lambda_{\text{join}} = \lambda / n_b$, where $n_b$ is the number of between-grid routes in the network, so that the total amount of between-grid travel is constant between the variations.

The variations illustrated in figure B.4a and B.4b have structures in-between that of the joined grids network and the four corners network. The results of prophylaxis strategies in these networks (see figures B.5 and B.6) show that neither



**(a)** The “four corners, equal sizes” variation

**(b)** The “two corners” variation

**(c)** The “centre-to-centre” variation

**(d)** The "nine-to-nine" variation

**Figure B.4** The “joined grids” arrangements of regions (circles), in which the movement of the population happens between adjacent regions in each grid (blue lines), on longer-range travel within a grid (yellow lines), or along inter-grid travel routes (red lines). All regions have a fixed population of 200,000, except for the regions corresponding to the two larger circles in (b), which have fixed populations of 400,000.

multiple connections between the two grids, nor uneven sizes of region populations, is a sufficient change to the joined grids network on its own to make risk-targeting always equal or exceed centrality-targeting in effectiveness—the two modifications must be combined, as in the four corners network.

The results for the centre-to-centre variation, illustrated in figure B.4c, are shown in figure B.7. These results are similar to those for the joined grids network, and illustrate that there is nothing particularly special about joining regions situated at the corners of the grids—joining the central regions produces similar results. The results for the nine-to-nine variation, illustrated in figure B.4d, are shown in figure B.8. Here, the mean number of infections is very similar for the risk- and RWPC-targeting strategies for all parameter values. The nine-to-nine

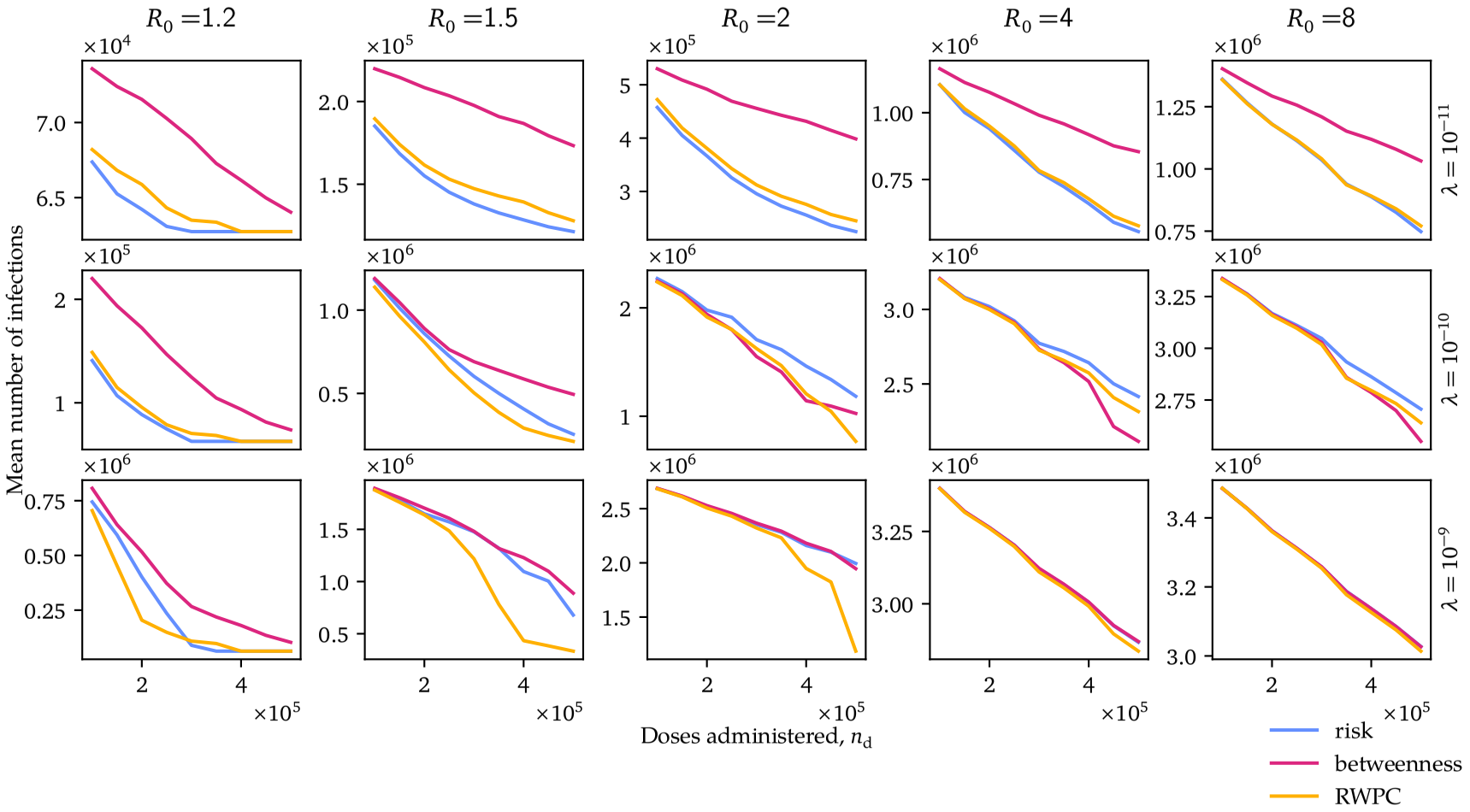



**Figure B.5** Plots of the mean total number of infections due to a disease outbreak against the number of prophylaxis doses administered according to one of three strategies, for five different values of $R_0$ and three different overall travel rates, in the four corners, equal sizes variation of the joined grids network. The long-distance travel parameter $a$ in equation 4.18 is set to 2 in all cases, and $\mu = 0.5$.

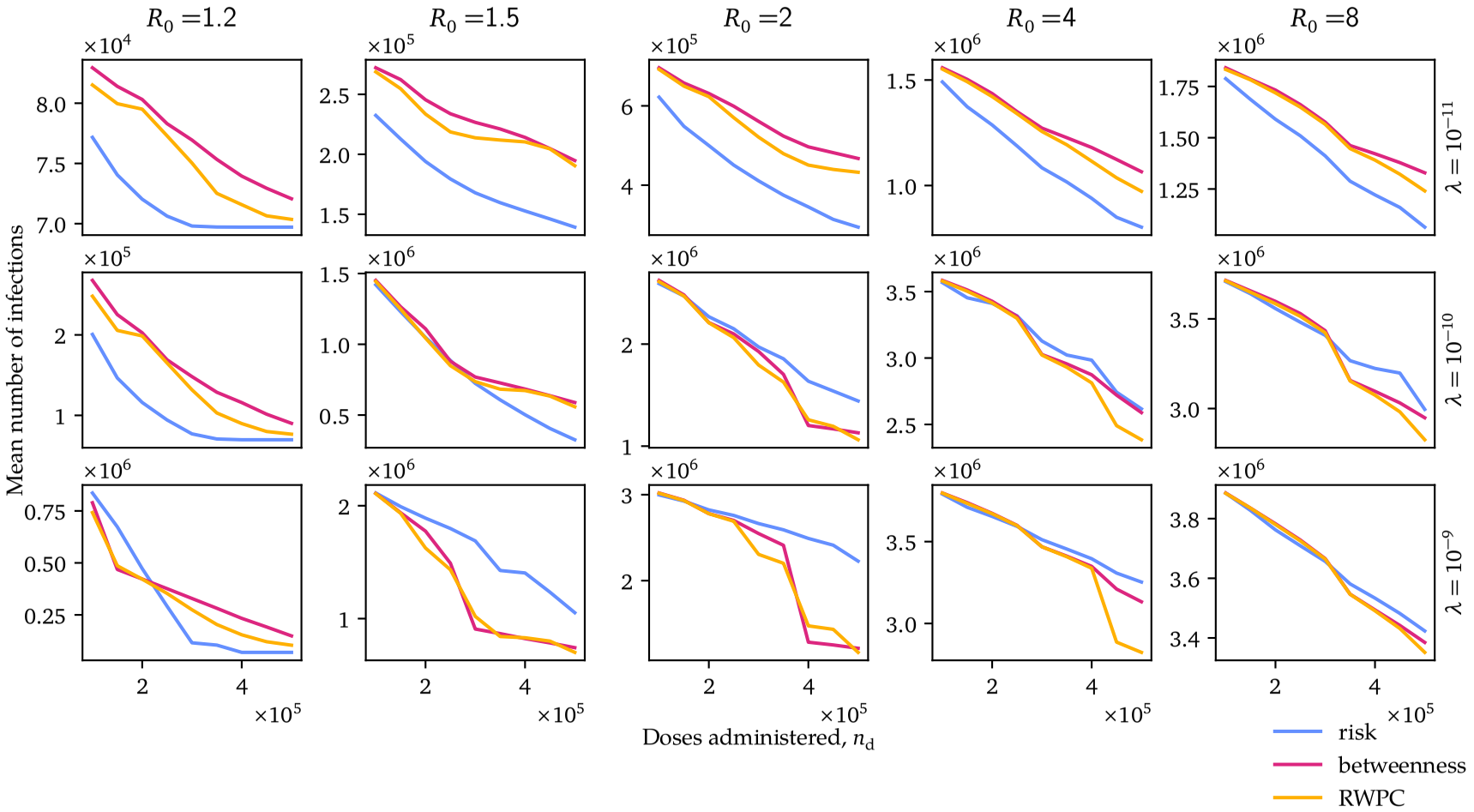



**Figure B.6** Plots of the mean total number of infections due to a disease outbreak against the number of prophylaxis doses administered according to one of three strategies, for five different values of $R_0$ and three different overall travel rates, in the two corners variation of the joined grids network. The long-distance travel parameter $a$ in equation 4.18 is set to 2 in all cases, and $\mu = 0.5$.

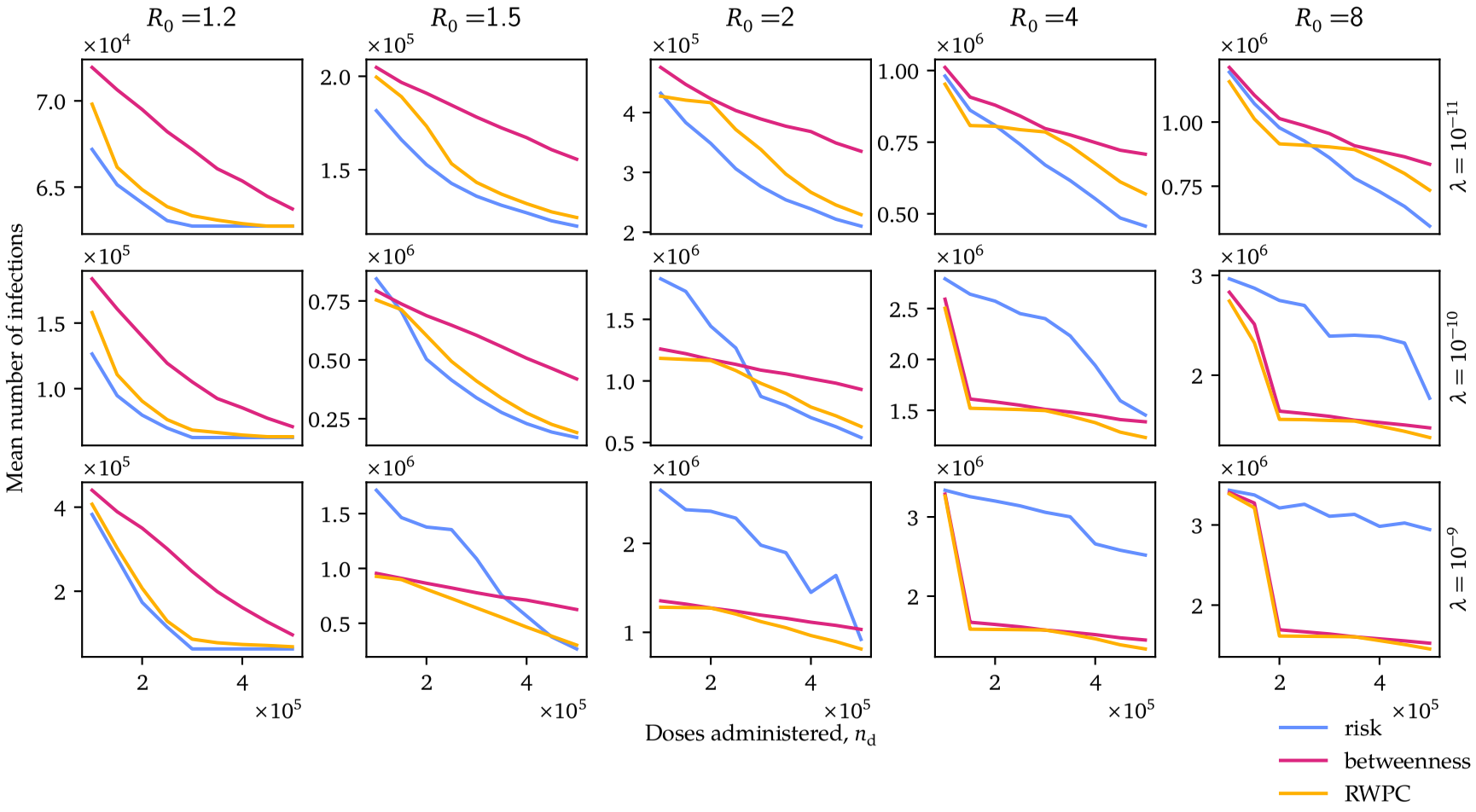



**Figure B.7** Plots of the mean total number of infections due to a disease outbreak against the number of prophylaxis doses administered according to one of three strategies, for five different values of $R_0$ and three different overall travel rates, in the centre-to-centre variation of the joined grids network. The long-distance travel parameter $a$ in equation 4.18 is set to 2 in all cases, and $\mu = 0.5$.

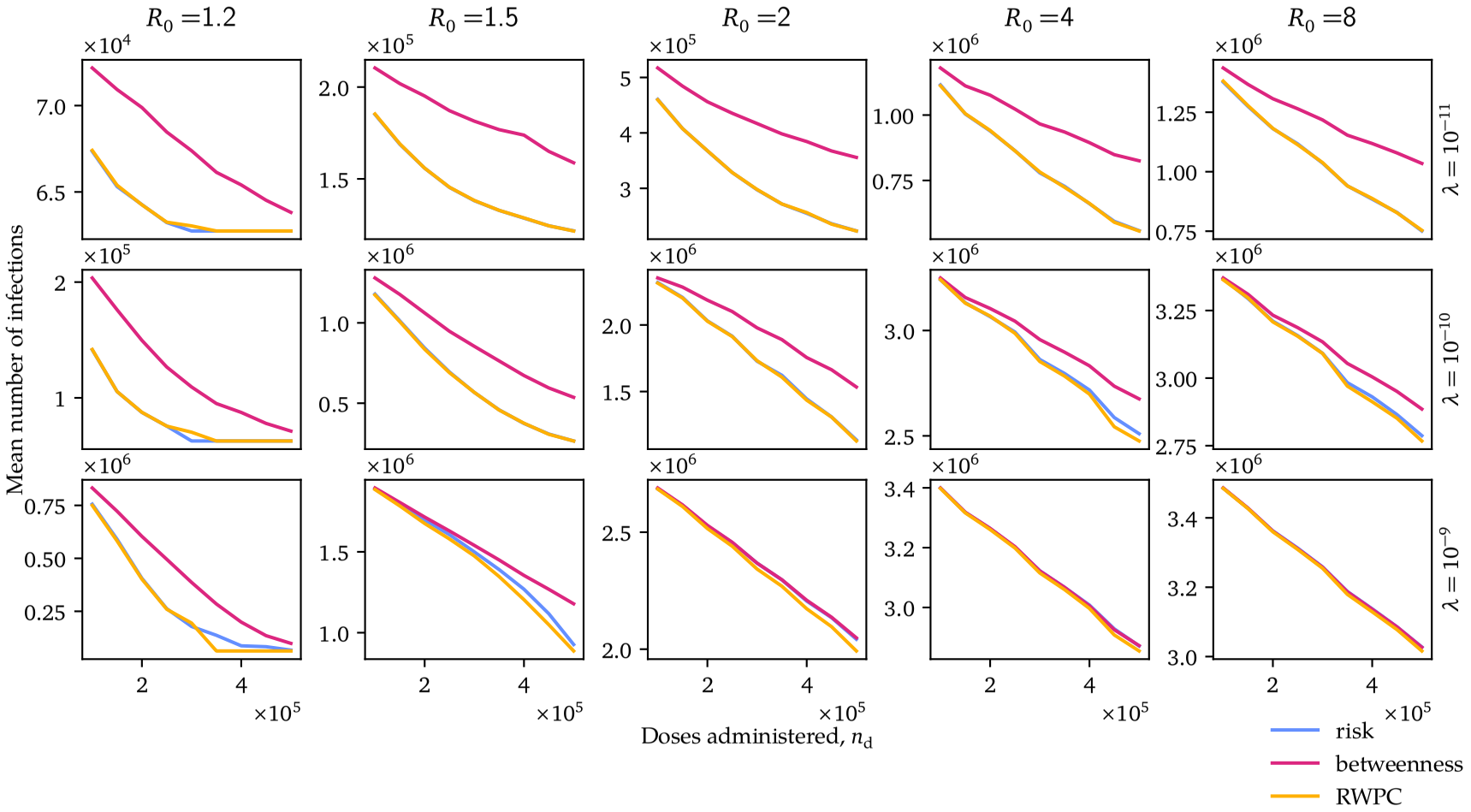



**Figure B.8** Plots of the mean total number of infections due to a disease outbreak against the number of prophylaxis doses administered according to one of three strategies, for five different values of $R_0$ and three different overall travel rates, in the nine-to-nine variation of the joined grids network. The long-distance travel parameter $a$ in equation 4.18 is set to 2 in all cases, and $\mu = 0.5$.

network therefore provides an example of a network in which the choice between risk- and RWPC-targeting strategies is immaterial.

## Non-pharmaceutical interventions ⬡ B.3

In section 4.6 I investigated the effect of targeting non-pharmaceutical interventions (NPIs) according to various different strategies, in the US air travel network as defined in section 4.5 (see figure 4.7). Here, I show analogous results for the other travel networks defined in chapter 4. In general, the conclusions regarding the targeting of NPIs are similar to those for the targeting of a pharmaceutical prophylactic—centrality-targeting is more effective than risk-targeting when $R_0$ and $\lambda$ are high, or when $n_r$ is low, in the joined grids and US air travel network. These plots are less smooth than the corresponding plots in figures 4.1–4.6, due to discontinuities when $n_r$ becomes large enough to place an additional region under restrictions.

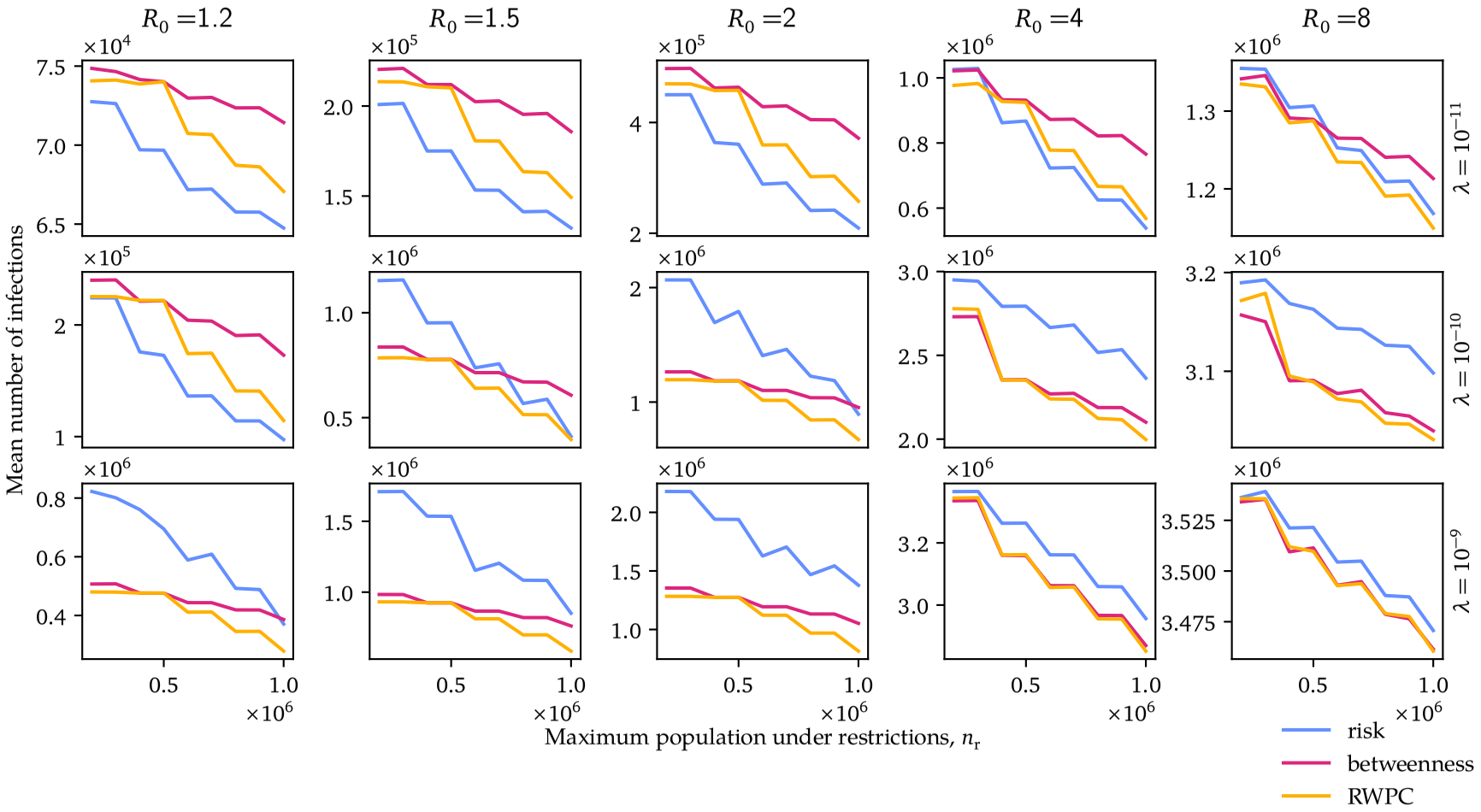



**Figure B.9** Plots of the expected total number of infections due to a disease outbreak against the number of individuals placed under transmission-reducing restrictions, chosen according to one of three NPI strategies, for five different values of $R_0$ and three different overall travel rates, in the "joined grids" network. The overall patterns are similar to those in figure 4.2. The long-distance travel parameter $a$ in equation 4.17 is set to 2 in all cases, and $\mu = 0.5$.

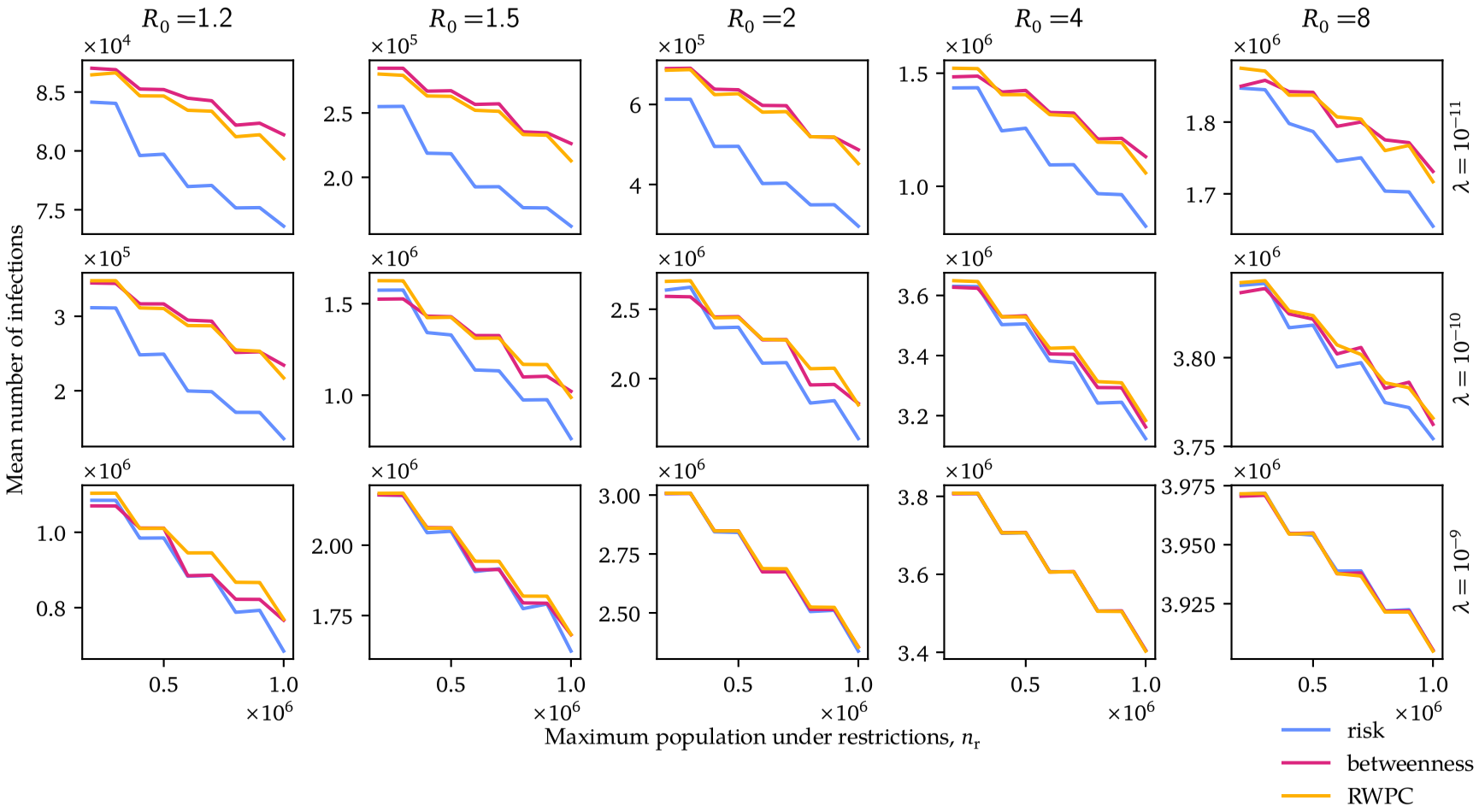



**Figure B.10** Plots of the expected total number of infections due to a disease outbreak against the number of individuals placed under transmission-reducing restrictions, chosen according to one of three NPI strategies, for five different values of $R_0$ for five different values of $R_0$ and three different overall travel rates, in the "four corners" network. The overall patterns are similar to those in figure 4.3. The long-distance travel parameter $a$ in equation 4.17 is set to 2 in all cases, and $\mu = 0.5$.

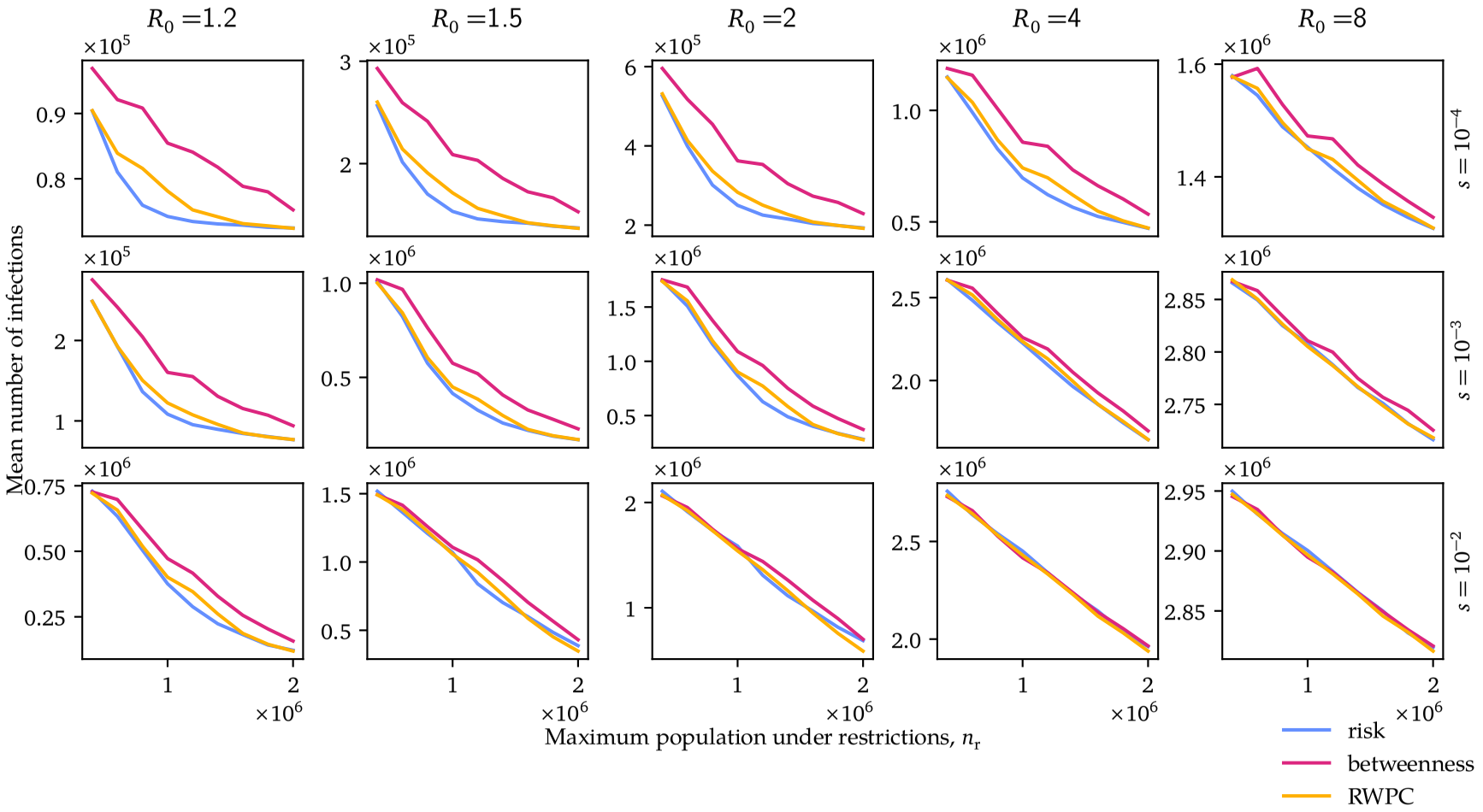



**Figure B.11** Plots of the expected total number of infections due to a disease outbreak against the number of individuals placed under transmission-reducing restrictions, chosen according to one of three NPI strategies, for five different values of $R_0$ and three different values of the travel scaling factor $s$, in the Northwest England commuter network. The overall patterns are similar to those in figure 4.6. The long-distance travel parameter $a$ in equation 4.17 is set to 2 in all cases, and $\mu = 0.5$.

# C Statistical models of the spread of SARS-CoV-2 in England

My original approach to using data on the spread of SARS-CoV-2 in England to parameterise an epidemiological metapopulation model used a statistical model of pathogen spread, in which the number of new cases in a cluster in a given week is drawn from some distribution dependent on the number of new cases in the previous week and on the travel network. I fit this distribution to weekly case data as provided by Public Health England [257], using Bayesian techniques. However, the models parameterised in this way fit the data poorly, either consistently underestimating cases or overestimating them, depending on the kind of distribution used. I therefore chose to use the mechanistic model described in chapter 5 instead. This appendix contains a description of the statistical model, and a discussion of why it provides a poor fit to the data.

## Modelling weekly cases C.1

Data on weekly confirmed cases of COVID-19 in each MSOA are available from Public Health England [257]. In order to fit my transmission model, I use the data on the weekly rolling sum of confirmed cases by specimen date, which are available for all MSOAs from the week ending 2020-08-21. The number of confirmed cases is not the same as the number of infections, so these models are best thought

of as predicting the number of new successfully detected cases per week.

I assume that the expected number of confirmed cases $\mathbb{E}\left(C_{i,w}\right)$ in cluster $i$ for week $w$ is given by:

$$\begin{aligned}\mathbb{E}\left(C_{i,w}\right) &= \gamma_i \left( C_{i,w-1} - \sum_{j \neq i} \bar{p}_{ij} C_{i,w-1} + \sum_{j \neq i} \bar{p}_{ji} C_{j,w-1} \right) \\ &= \gamma_i g_i(C_{w-1}),\end{aligned} \tag{C.1}$$

where $\gamma_i$ is the growth rate, $\bar{p}_{ij}$ is the mean of the proportion of travellers as defined in section 5.1, and $C_{k,w-1}$ is the number of confirmed cases in cluster $k$ in the previous week. The first term in $C_{i,w-1}$, comes about from the assumption that the number of confirmed cases in cluster $i$ is proportional to the number of infectious individuals in that cluster, and that the *serial interval* of SARS-CoV-2 is less than one week, meaning that secondary cases are recorded less than one week after the primary case that caused them. If this is true, and the number of infectious individuals in the cluster is much smaller than the population of the cluster, then we can assume that the expected number of new confirmed cases is proportional to the confirmed cases in the previous week, since the rate of infection $\beta SI$ in the SIR model is proportional to $I$. However, infectious individuals that are not located in cluster $i$ cannot cause infections there, leading to the second term in $C_{i,w-1}$, which gives the total amount of time that cases attributed to cluster $i$ are expected to spend located in other clusters. Cases located in clusters outside of their home cluster can cause infections there, leading to the second sum over $j \neq i$, which gives the total amount of time that cases attributed to clusters $j \neq i$ are expected to spend in cluster $i$. Unlike the model presented in section 5.2, this expression ignores the contribution to new cases from susceptibles who become infected during a visit to another region. This assumption allows the fitting process in each of the clusters to be occur independently.

I assume that $\gamma_i$ is constant during the 11-week period from 2020-08-21 up until

the beginning of the period of nationwide COVID-19 restrictions on 2020-11-05. I use Bayesian statistics to estimate the value of $\gamma_i$ for each cluster in this period, using two alternative forms for the likelihood function—the Poisson distribution and the negative binomial distribution.

## The Poisson model ⬡ C.2

In the Poisson model, the log likelihood $\mathscr{L}_{\mathrm{P}}(C)$ of the array $C$ of new cases by cluster and week, takes the form:

$$\mathscr{L}_{\mathrm{P}}(C) = \sum_{w=2}^{11} \sum_{i} -\gamma_i g_i(C_{w-1}) + C_{i,w} \ln\left(\gamma_i g_i(C_{w-1})\right) - \ln\left(C_{i,w}!\right), \tag{C.2}$$

where the form of $g_i(C_{i,w-1})$ is given by equation C.1. This model assumes that the mean of $C_{i,w}$ is equal to $\gamma_i g_i(C_{w-1})$, and that its variance is equal to its mean. I use a gamma distribution for the value of $\gamma_i$, with shape parameter 1 and scale parameter 1. I use a Metropolis-Hastings (MH) algorithm with a log-normal proposal distribution to avoid sampling values of $\gamma_i$ less than zero. Because $g_i(C_{w-1})$ is independent for each $i$, I can sample fit the value of $\gamma_i$ independently for each cluster. A proposal variance of 0.01 leads to the best performance of the MH algorithm.

One hundred epidemics as generated by this model are illustrated by the grey lines in figure C.1, which give the total new cases across all clusters by week. The black circles give the corresponding value for data to which the model was fit. The model fits the data poorly, due to significantly underestimating sudden growth in case numbers in some clusters, and overestimating growth in others, as illustrated in figure C.2.

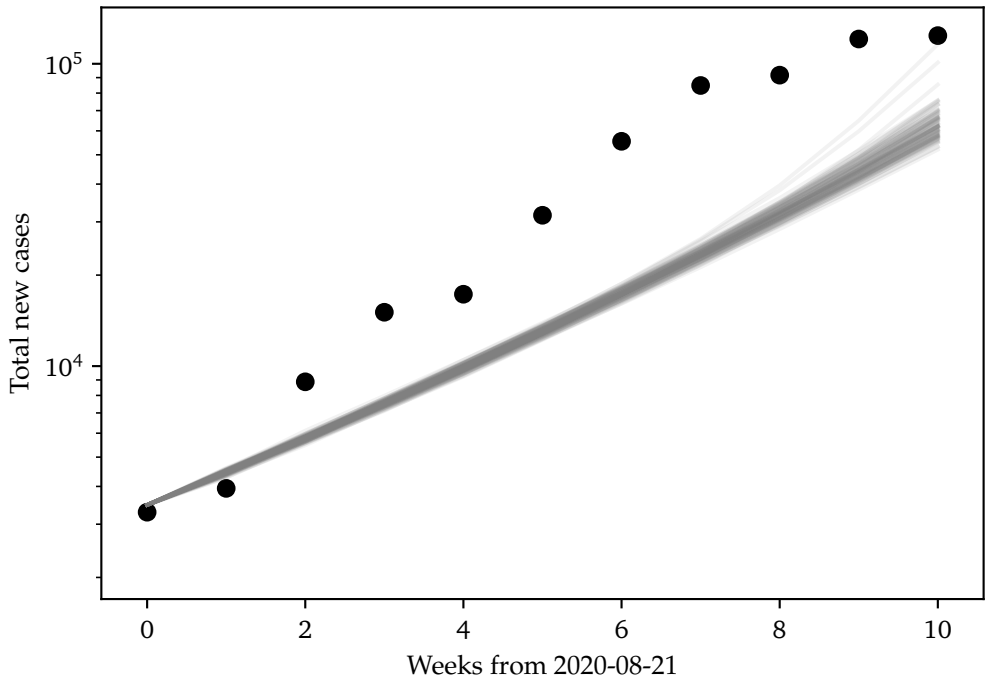



**Figure C.1** A plot of the total number of new cases in each week in the data reported by Public Health England (black circles), and for 100 runs of the Poisson statistical model of SARS-CoV-2 spread in England (grey lines) [257]. The number of new cases is consistently underestimated by the Poisson model.

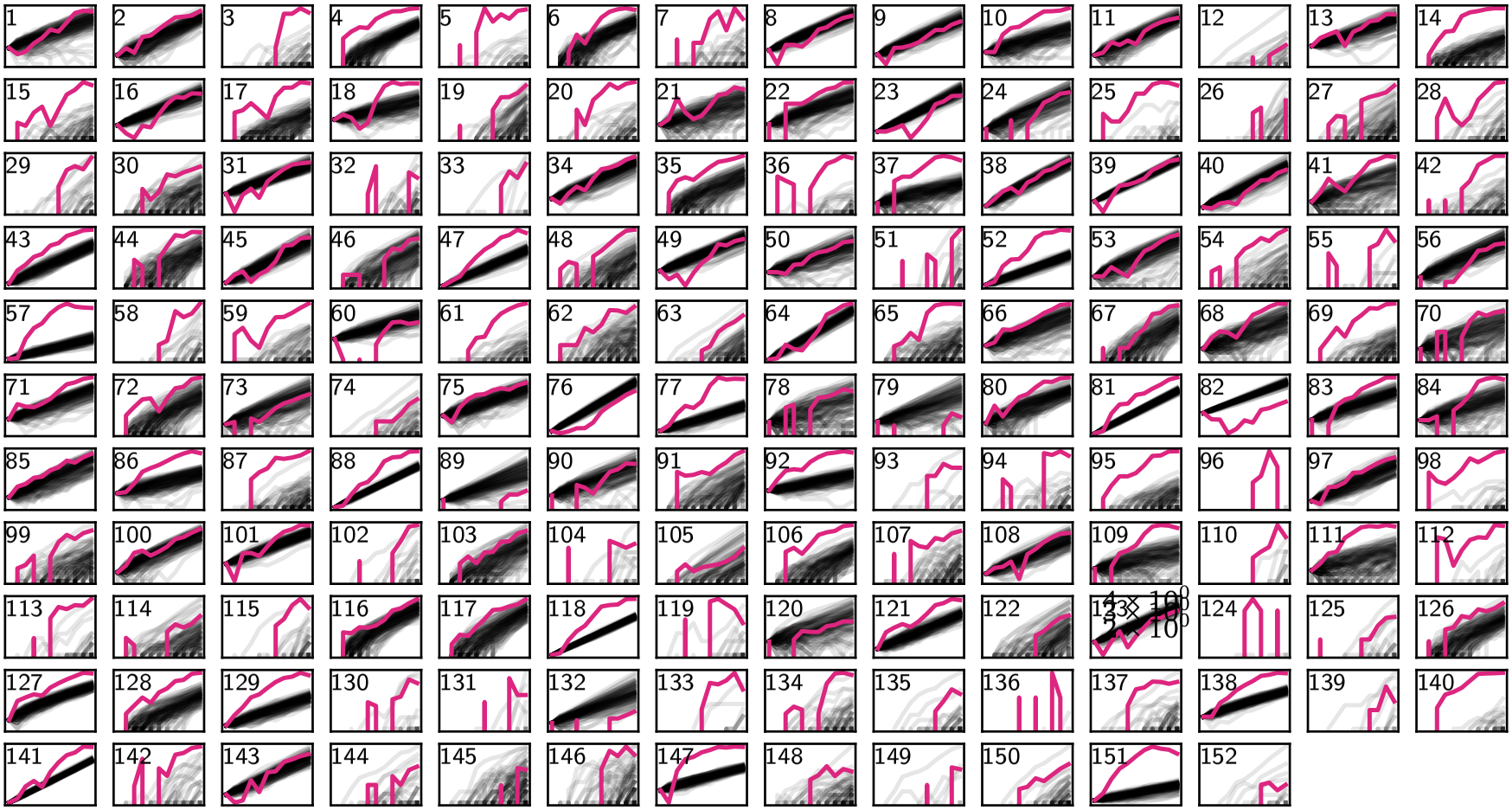



**Figure C.2** Plots showing the new cases in each cluster in each week in the data reported by Public Health England (red lines), and for 100 runs of the Poisson statistical model of SARS-CoV-2 spread in England (black lines) [257]. In some clusters the Poisson model finds a good fit, in most clusters it underestimates the probability of sudden explosions of cases, and in a few clusters it overestimates the growth of the outbreak.

# The negative binomial model ⌬ C.3

The negative binomial distribution is sometimes referred to as the *overdispersed* form of the Poisson distribution, meaning that it has similar properties to the Poisson distribution, but that its variance is greater than its mean. In epidemiological modelling, the negative binomial distribution is thought to capture the role of *superspreaders*—infective individuals who transmit a pathogen much more than the average infective individual, either due to their individual biology, behaviour, or random circumstance. I created a negative binomial model in the hope that it would capture the sudden explosions in new cases that proved difficult to fit using the Poisson model.

In the negative binomial model, the log likelihood $\mathscr{L}_{\text{nb}}(C)$ of the array $C$ of new cases by cluster and week, takes the form:

$$\begin{aligned}\mathscr{L}_{\text{nb}}(C) = \sum_{w=2}^{11}\sum_{i} & C_{i,w}\ln\left(\gamma_i g_i(C_{w-1})\right) - \ln\left(C_{i,w}!\right) \\ & - C_{i,w}\ln\left(\gamma_i g_i(C_{w-1}) + \alpha_i^{-1}\right) - \alpha_i^{-1}\ln(1+\alpha_i\gamma_i g_i(C_{w-1})) \\ & + \ln((\alpha_i^{-1}+C_{i,w})!\,) - \ln(\alpha_i^{-1}!\,),\end{aligned} \tag{C.3}$$

where $\alpha > 0$ is the dispersal parameter, and the form of $g_i(C_{i,w-1})$ is given by equation C.1. This model assumes that the mean of $C_{i,w}$ is equal to $\gamma_i g_i(C_{w-1})$, and that its variance is equal to its $\gamma_i g_i(C_{w-1})(1+\alpha_i\gamma_i g_i(C_{w-1}))$. This variance is always greater than that of the Poisson distribution, and is greater for greater values of $\alpha_i$. I use a Metropolis-Hastings (MH) algorithm with an uncorrelated multivariate log-normal proposal distribution to avoid sampling values of $\gamma_i$ and $\alpha_i$ less than zero.

Using independent gamma distributions for $\gamma_i$ and $\alpha_i$ leads to extremely high values of $\gamma_i$ and $\alpha_i$ in clusters that experienced a local extinction followed by a resurgence. Due to the small amount of data, the best fit for this behaviour is

a distribution with extremely high dispersal and growth rate. To address this behaviour, I use a highly correlated prior. Such a prior encourages the fitting algorithm to search for solutions where different clusters have similar parameters, meaning that the data for one cluster helps inform the fitting process for each other cluster. I use a multivariate normal prior with a mean of 1 and a correlation of 0.9 for each pair of clusters, for both $\gamma_i$ and $\alpha_i$. This leads to fewer extreme values of $\gamma_i$ and $\alpha_i$.

One hundred epidemics as generated by this model are illustrated by the grey lines in figure C.3, which give the total new cases across all clusters by week. The black circles give the corresponding value for data to which the model was fit. The model fits the data poorly, significantly overestimating the growth of the epidemic. Indeed, in many simulations the epidemic reaches a size larger than the population of England. The behaviour of the model in each cluster is illustrated in figure C.4.

## Discussion ⬡ C.4

Despite trying a range of models and fitting methods, I was not able to fit a satisfactory statistical model of the spread of SARS-CoV-2 in the metapopulation defined in section 5.1. Both models fit simulated data from the corresponding distribution well, meaning that they are internally consistent. However, problems arise when trying to fit these models to case data. There are several potential reasons for this.

The cluster-by-cluster case data shows some erratic sudden increases and decreases in case numbers. This could be due to disease dynamics, or it could be due to irregularities in COVID-19 testing and reporting. The Poisson model cannot capture sudden changes like this, and the negative binomial model fit to this data predicts very high levels of dispersal, leading to very fast epidemic growth.

It is also likely that the underlying growth rate varied over the period from

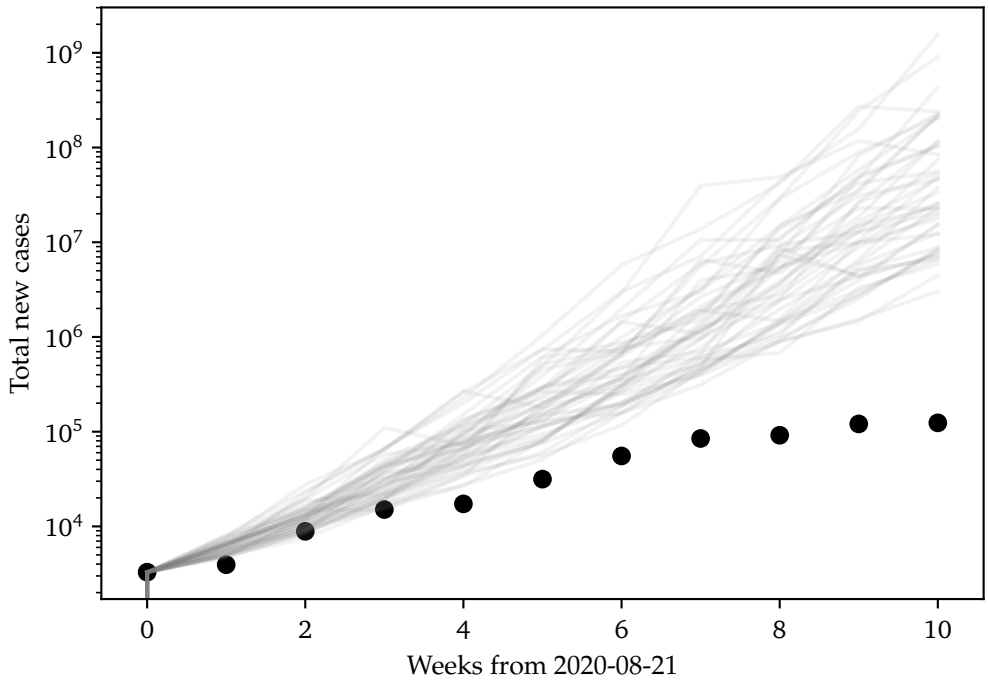



**Figure C.3** A plot of the total number of new cases in each week in the data reported by Public Health England (black circles), and for 100 runs of the negative binomial statistical model of SARS-CoV-2 spread in England (grey lines) [257]. The number of new cases is consistently overestimated by the negative binomial model.




**Figure C.4** Plots showing the new cases in each cluster in each week in the data reported by Public Health England (red lines), and for 100 runs of the negative binomial statistical model of SARS-CoV-2 spread in England (black lines) [257]. The y-axis is on a logarithmic scale. In general, the model overestimates the growth of the epidemic.

2020-08-21 to 2020-11-05, as various restrictions were introduced and lifted in different regions [240]. This means that the assumption that $\gamma_i$ is constant over the course of the simulation could lead to poor fitting. However, fitting a different growth rate $\gamma_{i,w}$ for each week $w$ would lead to fitting nearly as many parameters as there are data points—or more, in the overdispersed case. This would lead to overfitting [266].

The idea of a statistical model of the spread of SARS-CoV-2 in Gibbs' movement cluster remains promising, despite these problems. Such a model could be used to explore a wide range of what-if scenarios regarding the COVID-19 epidemic in the UK, such as testing the effect of different NPI strategies, or the earlier availability of prophylactic drugs or vaccines. Such a model may require more extensive data cleaning, or a sophisticated way to allow the growth rate to change week to week without overfitting the model.

# Bibliography


[1] Porta, M. *A Dictionary of Epidemiology* (Oxford University Press, USA, 2008). .

[2] Morens, D., Folkers, G. & Fauci, A. S. What Is a Pandemic? *JAMA - Journal of the American Medical Association* **321**, 910 (2009). .

[3] World Health Organization. What is a pandemic? (2010). .

[4] Johnson, N. P. A. S. & Mueller, J. Updating the Accounts: Global Mortality of the 1918-1920 "Spanish" Influenza Pandemic. *Bulletin of the History of Medicine* **76**, 105–115 (2002). .

[5] Global Health Data Exchange. Global Burden of Disease Results Tool (2021). .

[6] Cohen, M. L. Changing patterns of infectious disease. *Nature* **406**, 762–767 (2000). .

[7] Michaud, C. M. Global Burden of Infectious Diseases. In Schaechter, M. (ed.) *Encyclopedia of Microbiology*, 444–454 (Elsevier, 2009), third edn. .

[8] White, F. Rethinking Native American Language Revitalization. *American Indian Quarterly* **30**, 91–109 (2006). .

[9] McCaa, R. Spanish and Nahuatl Views on Smallpox and Demographic Catastrophe in Mexico. *The Journal of Interdisciplinary History* **25**, 397–431 (1995). .

[10] Patterson, K. B. & Runge, T. Smallpox and the Native American. *American Journal of the Medical Sciences* **323**, 216–222 (2002). .

[11] Thornton, R., Miller, T. & Warren, J. American Indian Population Recovery Following Smallpox Epidemics. *American Anthropologist* **93**, 28–45 (1991). .

[12] Stover, J. & Bollinger, L. The Economic Impact of AIDS. Tech. Rep. March, The Futures Group International (1999). .

[13] Bonnel, R. HIV/AIDS and Economic Growth : A Global Perspective. *South African Journal of Economics* **68**, 360–379 (2000). .

[14] Meltzer, M. I., Cox, N. J. & Fukuda, K. The Economic Impact of Pandemic Influenza in the United States: Priorities for Intervention. *Emerging Infectious Diseases* **5**, 659–671 (1999). .

[15] Burki, T. Stigmatisation undermining Russia's HIV control efforts. *The Lancet Infectious Diseases* **15**, 881–882 (2015). .

[16] Nicola, M. *et al.* The socio-economic implications of the coronavirus pandemic (COVID-19): A review. *International Journal of Surgery* **78**, 185–193 (2020). .

[17] Guha-Sapir, D. *EM-DAT: The Emergency Events Database* (Université catholique de Louvain (UCL) - CRED, Brussels, Belgium, 2018). .

[18] Millett, P. & Snyder-Beattie, A. Existential Risk and Cost-Effective Biosecurity. *Health Security* **15**, 373–383 (2017). .

[19] Zwetsloot, R. Testing Richardson's Law: A (Cautionary) Note on Power Laws in Violence Data. *SSRN Electronic Journal* (2018). .

[20] Ord, T. *The Precipice: Existential Risk and the Future of Humanity* (Bloomsbury Publishing, London, 2020).

[21] King, D. A., Peckham, C., Waage, J. K., Brownlie, J. & Woolhouse, M. E. Infectious diseases: Preparing for the future. *Science* **313**, 1392–1393 (2006). .

[22] Adalja, A. A., Watson, M., Toner, E. S., Cicero, A. & Inglesby, T. V. The Characteristics of Pandemic Pathogens. Tech. Rep., Johns Hopkins Center for Health Security (2018). .

[23] Simpson, S., Kaufmann, M. C., Glozman, V. & Chakrabarti, A. Disease X: accelerating the development of medical countermeasures for the next pandemic. *The Lancet Infectious Diseases* **20**, e108–e115 (2020). .

[24] Cabinet Office. National Risk Register Of Civil Emergencies 2017 edition. Tech. Rep. (2017). .

[25] UK Health Security Agency. Deaths in United Kingdom (2021). .

[26] Colizza, V., Pastor-Satorras, R. & Vespignani, A. Reaction-diffusion processes and metapopulation models in heterogeneous networks. *Nature Physics* **3**, 276–282 (2007).

[27] Wang, Z. *et al.* Statistical physics of vaccination. *Physics Reports* **664**, 1–113 (2016). .

[28] Cirillo, P. & Taleb, N. N. Tail risk of contagious diseases. *Nature Physics* **16**, 606–613 (2020). .

[29] Anderson, R. & May, R. Population biology of infectious diseases: Part I. *Nature* **280**, 361–367 (1979). .

[30] Earn, D. J., Rohani, P. & Grenfell, B. T. Persistence, chaos and synchrony in ecology and epidemiology. *Proceedings of the Royal Society B: Biological Sciences* **265**, 7–10 (1998). .

[31] Grenfell, B. T. *et al.* Unifying the Epidemiological and Evolutionary Dynamics of Pathogens. *Science* **303**, 327–332 (2004).

[32] Bjørnstad, O. N. & Grenfell, B. T. Hazards, spatial transmission and timing of outbreaks in epidemic metapopulations. *Environmental and Ecological Statistics* **15**, 265–277 (2008).

[33] Newman, M. E., Strogatz, S. H. & Watts, D. J. Random graphs with arbitrary degree distributions and their applications. *Physical Review E - Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics* **64**, 17 (2001).

[34] Bansal, S., Grenfell, B. T. & Meyers, L. A. When individual behaviour matters: homogeneous and network models in epidemiology. *Journal of the Royal Society Interface* **4**, 879–91 (2007).

[35] Rosvall, M. & Bergstrom, C. T. Maps of random walks on complex networks reveal community structure. *PNAS* **105**, 1118–1123 (2008).

[36] Hamra, G., MacLehose, R. & Richardson, D. Markov chain monte carlo: An introduction for epidemiologists. *International Journal of Epidemiology* **42**, 627–634 (2013).

[37] Ross, R. An application of the theory of probabilities to the study of a priori pathometry.—Part I. *Proceedings of the Royal Society A* 204–230 (1916).

[38] Kermack, W. O. & McKendrick, A. G. A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London A* **115**, 700–721 (1927).

[39] Barbour, A. D. The duration of the closed stochastic epidemic. *Biometrika* **62**, 477–482 (1975).

[40] Hethcote, H. W. The Mathematics of Infectious Diseases. *SIAM Review* **42**, 599–653 (2000).

[41] House, T., Ross, J. V. & Sirl, D. How big is an outbreak likely to be? Methods for epidemic final-size calculation. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* **469**, 20120436 (2012). .

[42] van Herwaarden, O. A. & Grasman, J. Stochastic epidemics: major outbreaks and the duration of the endemic period. *Journal of Mathematical Biology* **33**, 581–601 (1995). .

[43] Miller, J. C., Davoudi, B., Meza, R., Slim, A. C. & Pourbohloul, B. Epidemics with general generation interval distributions. *Journal of Theoretical Biology* **262**, 107–115 (2010). .

[44] Juul, J. L., Græsbøll, K., Christiansen, L. E. & Lehmann, S. Fixed-time descriptive statistics underestimate extremes of epidemic curve ensembles. *Nature Physics* **17**, 5–8 (2021). .

[45] Frigg, R. & Hartmann, S. Models in Science (2020). .

[46] Bhattacharyya, S., Gesteland, P. H., Korgenski, K., Bjørnstad, O. N. & Adler, F. R. Cross-immunity between strains explains the dynamical pattern of paramyxoviruses. *Proceedings of the National Academy of Sciences* **112**, 13396–13400 (2015). .

[47] Chang, S. *et al.* Mobility network models of COVID-19 explain inequities and inform reopening. *Nature* **589**, 82–87 (2021). .

[48] Winsberg, E. Computer Simulations in Science (2019). .

[49] Thompson, R. N., Thompson, C. P., Pelerman, O., Gupta, S. & Obolski, U. Increased frequency of travel in the presence of cross-immunity may act to decrease the chance of a global pandemic. *Philosophical Transactions of the Royal Society B: Biological Sciences* **374**, 20180274 (2019). .

[50] Gog, J. R. & Hollingsworth, T. D. Epidemic interventions: Insights from classic results. *Philosophical Transactions of the Royal Society B: Biological Sciences* **376**, 20200263 (2021). .

[51] Danon, L., Brooks-Pollock, E., Bailey, M. & Keeling, M. A spatial model of COVID-19 transmission in England and Wales: Early spread, peak timing and the impact of seasonality. *Philosophical Transactions of the Royal Society B: Biological Sciences* **376**, 20200272 (2021).

[52] Ferguson, N. *et al.* Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand. Tech. Rep. March (2020). .

[53] Thompson, R. N., Gilligan, C. A. & Cunniffe, N. J. Detecting Presymptomatic Infection Is Necessary to Forecast Major Epidemics in the Earliest Stages of Infectious Disease Outbreaks. *PLoS Computational Biology* **12**, 1–18 (2016). .

[54] Funk, S. *et al.* Assessing the performance of real-time epidemic forecasts: A case study of the 2013-16 Ebola epidemic. *PLoS Computational Biology* **15**, 177451 (2018). .

[55] Mastin, A. J., Van Den Bosch, F., Van Den Berg, F. & Parnell, S. R. Quantifying the hidden costs of imperfect detection for early detection surveillance. *Philosophical Transactions of the Royal Society B: Biological Sciences* **374**, 3–8 (2019). .

[56] Thompson, R., Gilligan, C. & Cunniffe, N. Will an outbreak exceed available resources for control? Estimating the risk from invading pathogens using practical definitions of a severe epidemic. *Journal of the Royal Society Interface* **17**, 20200690 (2020). .

[57] Bonds, M. H., Keenan, D. C., Rohani, P. & Sachs, J. D. Poverty trap formed by the ecology of infectious diseases. *Proceedings of the Royal Society B: Biological Sciences* **277**, 1185–1192 (2010). .

[58] McKendrick, A. G. On certain mathematical aspects of malaria. *Paludism* **1**, 54–66 (1912).

[59] De Quadros, C. A. *et al.* Measles Elimination in the Americas: Evolving Strategies. *Journal of the American Medical Association* **275**, 224–229 (1996). .

[60] Anderson, R. M. & May, R. M. Vaccination and herd immunity to infectious diseases. *Nature* **318**, 323–329 (1985). .

[61] Gay, N. J. The Theory of Measles Elimination: Implications for the Design of Elimination Strategies. *Journal of Infectious Diseases* **189**, 27–35 (2004). .

[62] Keeling, M. J. & Rohani, P. Controlling Infectious Diseases. In *Modeling Infectious Diseases in Humans and Animals*, chap. 8, 291–334 (Princeton University Press, Princeton, 2008).

[63] Jones, D. & Helmreich, S. A history of herd immunity. *The Lancet* **396**, 810–811 (2020). .

[64] Hogan, A. B. *et al.* Report 33: Modelling the allocation and impact of a COVID-19 vaccine. Tech. Rep., Imperial College COVID-19 response team (2020). .

[65] Pei, S., Kandula, S., Yang, W. & Shaman, J. Forecasting the spatial transmission of influenza in the United States. *Proceedings of the National Academy of Sciences of the United States of America* **115**, 2752–2757 (2018). .

[66] Keeling, M. J. & Rohani, P. *Modeling Infectious Diseases in Humans and Animals* (Princeton University Press, 2008).

[67] Chowell, G., Sattenspiel, L., Bansal, S. & Viboud, C. Mathematical models to characterize early epidemic growth: A review. *Physics of Life Reviews* **18**, 66–97 (2016). .

[68] Kraemer, M. U. *et al.* Spread of yellow fever virus outbreak in Angola and the Democratic Republic of the Congo 2015–16: a modelling study. *The Lancet Infectious Diseases* **17**, 330–338 (2017). .

[69] Gomes, M. F. C. *et al.* Assessing the International Spreading Risk Associated with the 2014 West African Ebola Outbreak. *PLoS Currents Outbreaks* **1** (2014). .

[70] Ferretti, L. *et al.* Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. *Science* **368**, eabb6936 (2020). .

[71] Reichert, T. A. *et al.* The Japanese experience with vaccinating school children against influenza. *New England Journal of Medicine* **344**, 889–896 (2001). .

[72] Keeling, M. J., Woolhouse, M. E., May, R. M., Davies, G. & Grenfell, B. T. Modelling vaccination strategies against foot-and-mouth disease. *Nature* **421**, 136–142 (2003). .

[73] Graham, M. *et al.* Measles and the canonical path to elimination. *Science* **364**, 584–587 (2019). .

[74] Chowell, G., Nishiura, H. & Bettencourt, L. M. Comparative estimation of the reproduction number for pandemic influenza from daily case notification data. *Journal of the Royal Society Interface* **4**, 154–166 (2007). .

[75] Cori, A., Ferguson, N. M., Fraser, C. & Cauchemez, S. A new framework and software to estimate time-varying reproduction numbers during epidemics. *American Journal of Epidemiology* **178**, 1505–1512 (2013). .

[76] Goeyvaerts, N. *et al.* Estimating dynamic transmission model parameters for seasonal influenza by fitting to age and season-specific influenza-like illness incidence. *Epidemics* **13**, 1–9 (2015). .

[77] Gillespie, D. T. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *Journal of Computational Physics* **22**, 403–434 (1976). .

[78] Norris, J. R. *Markov Chains* (Cambridge University Press, 1997). .

[79] Jagers, P. *Branching processes with biological applications* (Wiley, London, 1975).

[80] Anderson, D. & Watson, R. On the spread of a disease with gamma distributed latent and infectious periods. *Biometrika* **67**, 191–198 (1980). .

[81] Svensson, A. On the Asymptotic Size and Duration of a Class of Epidemic Models. *Journal of Applied Probability* **32**, 11–24 (1995). .

[82] Craft, M. E., Beyer, H. L. & Haydon, D. T. Estimating the Probability of a Major Outbreak from the Timing of Early Cases: An Indeterminate Problem? *PLoS ONE* **8**, e57878 (2013). .

[83] Watts, D. J., Muhamad, R., Medina, D. C. & Dodds, P. S. Multiscale, resurgent epidemics in a hierarchical metapopulation model. *Proceedings of the National Academy of Sciences* **102**, 11157–11162 (2005). .

[84] Keeling, M. J. & Rohani, P. Spatial Models. In *Modelling Infectious Diseases in Humans and Animals*, chap. Seven, 232–290 (Princeton University Press, 2008). .

[85] Miller, J. C. A Note on the Derivation of Epidemic Final Sizes. *Bulletin of Mathematical Biology* **74**, 2125–2141 (2012). .

[86] Ferguson, N. M. *et al.* Strategies for containing an emerging influenza pandemic in Southeast Asia. *Nature* **437**, 209–214 (2005). .

[87] Merler, S. *et al.* Containing Ebola at the Source with Ring Vaccination. *PLoS Neglected Tropical Diseases* **10**, e0005093 (2016). .

[88] Peak, C. M., Childs, L. M., Grad, Y. H. & Buckee, C. O. Comparing nonpharmaceutical interventions for containing emerging epidemics. *Proceedings of the National Academy of Sciences of the United States of America* **114**, 4023–4028 (2017). .

[89] Ferguson, N. M. *et al.* Strategies for mitigating an influenza pandemic. *Nature* **442**, 448–452 (2006). .

[90] Ajelli, M. *et al.* The 2014 Ebola virus disease outbreak in Pujehun, Sierra Leone: Epidemiology and impact of interventions. *BMC Medicine* **13**, 281 (2015). .

[91] Ball, F. *et al.* Seven challenges for metapopulation models of epidemics, including households models. *Epidemics* **10**, 63–67 (2015). .

[92] Riley, S., Eames, K., Isham, V., Mollison, D. & Trapman, P. Five challenges for spatial epidemic models. *Epidemics* **10**, 68–71 (2015). .

[93] Levins, R. Some Demographic and Genetic Consequences of Environmental Heterogeneity for Biological Control. *Bulletin of the Entomological Society of America* **15**, 237–240 (1969). .

[94] Xia, Y., Bjørnstad, O. N. & Grenfell, B. T. Measles metapopulation dynamics: A gravity model for epidemiological coupling and dynamics. *American Naturalist* **164**, 267–281 (2004). .

[95] Colizza, V., Barrat, A., Barthelemy, M., Valleron, A. J. & Vespignani, A.

Modeling the worldwide spread of pandemic influenza: Baseline case and containment interventions. *PLoS Medicine* **4**, 95–110 (2007). .

[96] Rowthorn, R. E., Laxminarayan, R. & Gilligan, C. A. Optimal control of epidemics in metapopulations. *Journal of the Royal Society Interface* **6**, 1135–1144 (2009). .

[97] Gilarranz, L. J. & Bascompte, J. Spatial network structure and metapopulation persistence. *Journal of Theoretical Biology* **297**, 11–16 (2012). .

[98] Rock, K., Brand, S., Moir, J. & Keeling, M. J. Dynamics of infectious diseases. *Reports on Progress in Physics* **77**, 026602 (2014). .

[99] Venkatramanan, S. *et al.* Optimizing spatial allocation of seasonal influenza vaccine under temporal constraints. *PLoS Computational Biology* **15**, e1007111 (2019). .

[100] Keeling, M. J. & Rohani, P. Stochastic Dynamics. In *Modeling Infectious Diseases in Humans and Animals*, chap. Six, 190–231 (Princeton University Press, 2008). .

[101] Newman, M. E. J. *Networks: An Introduction* (Oxford University Press, Oxford, 2010).

[102] Hufnagel, L., Brockmann, D. & Geisel, T. Forecast and control of epidemics in a globalized world. *Proceedings of the National Academy of Sciences* **101**, 15124–15129 (2004). .

[103] Colizza, V., Barrat, A., Barthelemy, M. & Vespignani, A. The role of the airline transportation network in the prediction and predictability of global epidemics. *Proceedings of the National Academy of Sciences* **103**, 2015–2020 (2006). .

[104] Department of Health. UK Influenza Pandemic Preparedness Strategy 2011. Tech. Rep. November (2011). .

[105] Homeland Security Council. National Strategy for Pandemic Influenza. Tech. Rep. November (2005).

[106] Government of Italy. National Plan for Preparedness and Response To an Influenza Pandemic. Tech. Rep. (2010).

[107] Heymann, D. L. & Rodier, G. R. Hot spots in a wired world: WHO surveillance of emerging and re-emerging infectious diseases. *Lancet Infectious Diseases* **1**, 345–353 (2001).

[108] Coninx, J. K. *et al.* (eds.) *Pandemic influenza preparedness and response: a WHO guidance document* (WHO Press, 2009). .

[109] on prioritisation of pathogens, W. Blueprint for R&D preparedness and response to public health emergencies due to highly infectious pathogens. Tech. Rep. December (2015). .

[110] World Health Organisation. International Health Regulations - Third Edition. Tech. Rep., World Health Organisation (2005). .

[111] Doshi, P. The elusive definition of pandemic influenza. *Bulletin of the World Health Organization* **89**, 532–538 (2011). .

[112] Nebehay, S. WHO says it no longer uses 'pandemic' category, but virus still emergency (2020). .

[113] The Independent Panel for Pandemic Preparedness and Response. Second report on progress. Tech. Rep. January (2021).

[114] Duff, J. H. *et al.* A global public health convention for the 21st century. *The Lancet. Public health* **2667**, 1–6 (2021). .

[115] Thompson, R. N. & Brooks-Pollock, E. Preface to theme issue 'Modelling infectious disease outbreaks in humans, animals and plants: Epidemic forecasting and control'. *Philosophical Transactions of the Royal Society B: Biological Sciences* **374**, 20190038 (2019). .

[116] Singer, B., Thompson, R. & Bonsall, M. The effect of the definition of 'pandemic' on quantitative assessments of infectious disease outbreak risk. *Scientific Reports* **11**, 2547 (2021). .

[117] Tuite, A. R., Fisman, D. N., Kwong, J. C. & Greer, A. L. Optimal pandemic influenza vaccine allocation strategies for the Canadian population. *PLoS ONE* **5**, e10520 (2010). .

[118] Facebook. Disease Prevention Maps (2021).

[119] Gibbs, H. *et al.* Detecting behavioural changes in human movement to inform the spatial scale of interventions against COVID-19. *PLoS Computational Biology* **17**, e1009162 (2021). .

[120] Mei, Y.-F., Harrach, B. & Wadell, G. *Mastadenovirus*, 33–48 (Springer New York, New York, NY, 2011). .

[121] Zandi, R., Reguera, D., Bruinsma, R. F., Gelbart, W. M. & Rudnick, J. Origin of icosahedral symmetry in viruses. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 15556–15560 (2004). .

[122] Simonsen, L. *et al.* Global Mortality Estimates for the 2009 Influenza Pandemic from the GLaMOR Project: A Modeling Study. *PLoS Medicine* **10**, e1001558 (2013). .

[123] Viboud, C. *et al.* Global mortality impact of the 1957-1959 influenza pandemic. *Journal of Infectious Diseases* **212**, 738–745 (2015). .

[124] Center for Systems Science and Engineering (CSSE) at Johns Hopkins University. COVID-19 Data Repository (2021). .

[125] Saker, L., Lee, K., Cannito, B., Gilmore, A. & Campbell-Lendrum, D. Globalization and infectious diseases: A review of the linkages. Tech. Rep. Special topics no. 3, World Health Organization (2004). .

[126] Collin, J. & Lee, K. Globalisation and transborder health risk in the UK: case studies in tobacco control and population mobility. Tech. Rep., Centre on Global Change and Health, LSHTM (2003). .

[127] International Tourism Highlights. Tech. Rep., World Tourism Organization (2019).

[128] Gushulak, B. D. & Macpherson, D. W. Population Mobility and Infectious Diseases: The Diminishing Impact of Classical Infectious Diseases and New Approaches for the 21st Century. *Clinical Infectious Diseases* **31**, 776–780 (2000). .

[129] Brownstein, J. S., Wolfe, C. J. & Mandl, K. D. Empirical evidence for the effect of airline travel on inter-regional influenza spread in the United States. *PLoS Medicine* **3**, 1826–1835 (2006). .

[130] Quinn, S. C. & Kumar, S. Health inequalities and infectious disease epidemics: A challenge for global health security. *Biosecurity and Bioterrorism* **12**, 263–273 (2014). .

[131] Chen, J. Pathogenicity and Transmissibility of 2019-nCoV—A Quick Overview and Comparison with Other Emerging Viruses. *Microbes and infection* (2020). .

[132] Mills, C. E., Robins, J. M. & Lipsitch, M. Transmissibility of 1918 pandemic influenza. *Nature* **432**, 904–906 (2004). .

[133] Patz, J. A., Campbell-Lendrum, D., Holloway, T. & Foley, J. A. Impact of regional climate change on human health. *Nature* **438**, 310–317 (2005). .

[134] Jones, K. E. *et al.* Global trends in emerging infectious diseases. *Nature* **451**, 990–993 (2008). .

[135] Shanks, G. D. *et al.* Variable mortality from the 1918–1919 influenza pandemic during military training. *Military Medicine* **181**, 878–882 (2016). .

[136] Brauner, J. M. *et al.* Inferring the effectiveness of government interventions against COVID-19. *Science* **371**, eabd9338 (2021). .

[137] Hatchett, R. J., Mecher, C. E. & Lipsitch, M. Public health interventions and epidemic intensity during the 1918 influenza pandemic. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 7582–7587 (2007). .

[138] Koonin, L. M. & Patel, A. Timely antiviral administration during an influenza pandemic: Key components. *American Journal of Public Health* **108**, S215–S220 (2018). .

[139] Opatowski, L., Baguelin, M. & Eggo, R. M. Influenza interaction with cocirculating pathogens and its impact on surveillance, pathogenesis, and epidemic profile: A key role for mathematical modelling. *PLoS Pathogens* **14**, e1006770 (2018). .

[140] Gagnon, A. *et al.* Age-Specific Mortality During the 1918 Influenza Pandemic: Unravelling the Mystery of High Young Adult Mortality. *PLoS ONE* **8**, e69586 (2013). .

[141] Kucharski, A. J., Andreasen, V. & Gog, J. R. Capturing the dynamics of pathogens with many strains. *Journal of Mathematical Biology* **72**, 1–24 (2016). .

[142] Penman, B. S., Gupta, S. & Shanks, G. D. Rapid mortality transition of Pacific Islands in the 19th century. *Epidemiology and Infection* **145**, 1–11 (2016). .

[143] Marshall, J. M. *et al.* Mathematical models of human mobility of relevance to malaria transmission in Africa. *Scientific Reports* **8**, 7713 (2018). .

[144] Bazzini, A. A., Asurmendi, S., Hopp, H. E. & Beachy, R. N. Tobacco mosaic virus (TMV) and potato virus X (PVX) coat proteins confer heterologous interference to PVX and TMV infection, respectively. *Journal of General Virology* **87**, 1005–1012 (2005). .

[145] Huang, A. *et al.* A systematic review of antibody mediated immunity to coronaviruses: kinetics, correlates of protection, and association with severity. *Nature Communications* **11**, 4704 (2020). .

[146] Koelle, K., Pascual, M. & Yunus, M. Serotype cycles in cholera dynamics. *Proceedings of the Royal Society B: Biological Sciences* **273**, 2879–2886 (2006). .

[147] Masumu, J., Marcotty, T., Geerts, S., Vercruysse, J. & Van den Bossche, P. Cross-protection between Trypanosoma congolense strains of low and high virulence. *Veterinary Parasitology* **163**, 127–131 (2009). .

[148] Chowell, G., Bettencourt, L. M., Johnson, N., Alonso, W. J. & Viboud, C. The 1918-1919 influenza pandemic in England and Wales: Spatial patterns in transmissibility and mortality impact. *Proceedings of the Royal Society B: Biological Sciences* **275**, 501–509 (2008). .

[149] Andreasen, V., Viboud, C. & Simonsen, L. Epidemiologic characterization of the 1918 influenza pandemic summer wave in Copenhagen: Implications for pandemic control strategies. *Journal of Infectious Diseases* **197**, 270–278 (2008). .

[150] Valkenburg, S. A. *et al.* Immunity to seasonal and pandemic influenza A viruses. *Microbes and Infection* **13**, 489–501 (2013). .

[151] Hancock, K. *et al.* Cross-Reactive Antibody Responses to the 2009 Pandemic H1N1 Influenza Virus. *New England Journal of Medicine* **361**, 1945–1952 (2009). .

[152] Carter, D. M. *et al.* Sequential Seasonal H1N1 Influenza Virus Infections Protect Ferrets against Novel 2009 H1N1 Influenza Virus. *Journal of Virology* **87**, 1400–1410 (2013). .

[153] Luk, J., Gross, P. & Thompson, W. W. Observations on mortality during the 1918 influenza pandemic. *Clinical Infectious Diseases* **33**, 1375–1378 (2001). .

[154] Chung, K. L. *Markov Chains with Stationary Transition Probabilities* (Springer, Berlin, Heidelberg, 1960).

[155] Peeters, M., D'Arc, M. & Delaporte, E. The origin and diversity of human retroviruses. *AIDS Review* **16**, 23–34 (2014). .

[156] Maartens, G., Celum, C. & Lewin, S. R. HIV infection: Epidemiology, pathogenesis, treatment, and prevention. *The Lancet* **384**, 258–271 (2014). .

[157] Tan, C.-W. *et al.* Pan-Sarbecovirus Neutralizing Antibodies in BNT162b2-Immunized SARS-CoV-1 Survivors. *New England Journal of Medicine* **385**, 1401–1406 (2021). .

[158] Hasan, S., Jamdar, S. F., Alalowi, M. & Al Ageel Al Beaiji, S. M. Dengue virus: A global human threat: Review of literature. *Journal of International Society of Preventive and Community Dentistry* **6**, 1–6 (2016). .

[159] Carmona, P. & Gandon, S. Winter is coming: Pathogen emergence in seasonal environments. *PLoS Computational Biology* **16**, e1007954 (2020). .

[160] Jancloes, M. *et al.* Climate services to improve public health. *International Journal of Environmental Research and Public Health* **11**, 4555–4559 (2014). .

[161] The Metropolitan Museum of Art. Twenty-sided die (icosahedron) with faces inscribed with Greek letters. .

[162] Osman, A. S. HIV / AIDS in the last 10 years. *Eastern Mediterranean Health Journal* **14**, 90–96 (2007). .

[163] Cohen, M. S. *et al.* The spread, treatment, and prevention of HIV-1: evolution of a global pandemic. *The Journal of Clinical Investigation* **118**, 1244–1254 (2008). .

[164] Breban, R., Riou, J. & Fontanet, A. Interhuman transmissibility of Middle East respiratory syndrome coronavirus: Estimation of pandemic risk. *The Lancet* **382**, 694–699 (2013). .

[165] Webster, R. G. Predictions for Future Human Influenza Pandemics. *The Journal of Infectious Diseases* **176**, S14–S19 (1997). .

[166] Towers, S. & Feng, Z. Pandemic H1N1 influenza: predicting the course of a pandemic and assessing the efficacy of the planned vaccination programme in the United States. *Eurosurveillance* **14**, 19358 (2009). .

[167] Coburn, B. J., Wagner, B. G. & Blower, S. Modeling influenza epidemics and pandemics: insights into the future of swine flu (H1N1). *BMC Medicine* **7**, 30 (2009). .

[168] Mateus, A. L., Otete, H. E., Beck, C. R., Dolan, G. P. & Nguyen-Van-Tam, J. S. Effectiveness of travel restrictions in the rapid containment of human

influenza: a systematic review. *Bulletin of the World Health Organization* 868–880D (2014). .

[169] Hossain, M. P. *et al.* The effects of border control and quarantine measures on the spread of COVID-19. *Epidemics* **32**, 100397 (2020). .

[170] Walker, P. G. T. *et al.* The impact of COVID-19 and strategies for mitigation and suppression in low- and middle-income countries. *Science* **422**, 413–422 (2020). .

[171] Merino, J. G. Response to Ebola in the US: Misinformation, fear, and new opportunities. *BMJ* **349**, g6712 (2014). .

[172] WHO. Ebola virus disease - United Kingdom (2014). .

[173] Gandhi, M., Yokoe, D. S. & Havlir, D. V. Asymptomatic transmission, the achilles' heel of current strategies to control Covid-19. *New England Journal of Medicine* **382**, 2158–2160 (2020). .

[174] Lai, C. C., Shih, T. P., Ko, W. C., Tang, H. J. & Hsueh, P. R. Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and coronavirus disease-2019 (COVID-19): The epidemic and the challenges. *International Journal of Antimicrobial Agents* **55**, 105924 (2020). .

[175] Sirleaf, H. E. J. & Clark, R. H. H. Losing time: End this pandemic and secure the future. Progress six months after the report of the Independent Panel for Pandemic Preparedness and Response. Tech. Rep. (2021). .

[176] Ritchie, H. *et al.* Coronavirus Pandemic (COVID-19). *Our World in Data* (2020).

[177] Tao, K. *et al.* The biological and clinical significance of emerging SARS-CoV-2 variants. *Nature Reviews Genetics* **22**, 757–773 (2021).

[178] Khan, K. *et al.* Infectious disease surveillance and modelling across geographic frontiers and scientific specialties. *The Lancet Infectious Diseases* **12**, 222–230 (2012). .

[179] Poletto, C., Tizzoni, M. & Colizza, V. Human mobility and time spent at destination : Impact on spatial epidemic spreading. *Journal of Theoretical Biology* **338**, 41–58 (2013). .

[180] Barbarossa, M. V. *et al.* Transmission dynamics and final epidemic size of ebola virus disease outbreaks with varying interventions. *PLoS ONE* **10**, e0131398 (2015). .

[181] Nishiura, H., Chowell, G., Safan, M. & Castillo-Chavez, C. Pros and cons of estimating the reproduction number from early epidemic growth rate of influenza A (H1N1) 2009. *Theoretical Biology and Medical Modelling* **7**, 1 (2010). .

[182] Apolloni, A., Poletto, C., Ramasco, J. J., Jensen, P. & Colizza, V. Metapopulation epidemic models with heterogeneous mixing and travel behaviour. *Theoretical Biology and Medical Modelling* **11**, 3 (2014). .

[183] Meakin, S. R., Tildesley, M. J., Davis, E. & Keeling, M. J. A metapopulation model for the 2018 Ebola outbreak in Equateur province in the Democratic Republic of the Congo. *bioRxiv* (2018). .

[184] Baguelin, M. *et al.* Control of equine influenza: Scenario testing using a realistic metapopulation model of spread. *Journal of the Royal Society Interface* **7**, 67–79 (2009). .

[185] Balcan, D. *et al.* Multiscale mobility networks and the spatial spreading of infectious diseases. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 21484–21489 (2009). .

[186] Bajardi, P. *et al.* Human mobility networks, travel restrictions, and the global spread of 2009 H1N1 pandemic. *PLoS ONE* **6**, e16591 (2011). .

[187] Gonçalves, B., Balcan, D. & Vespignani, A. Human mobility and the worldwide impact of intentional localized highly pathogenic virus release. *Scientific Reports* **3**, 810 (2013). .

[188] Bonneux, L. & Van Damme, W. Health is more than influenza. *Bulletin of the World Health Organization* **89**, 539–540 (2011). .

[189] Coronavirus Disease (COVID-19) Press Conference 17 February 2020 (2020). .

[190] Thompson, R. N., Gilligan, C. A. & Cunniffe, N. J. Control fast or control smart: When should invading pathogens be controlled? *PLoS Computational Biology* **14**, e1006014 (2018). .

[191] Kyoon-achan, G. & Wright, L. Community-based pandemic preparedness: COVID-19 procedures of a Manitoba First Nation community. *Journal of Community Safety and Well-Being* **5**, 45–50 (2020). .

[192] Marston, H. D., Paules, C. I. & Fauci, A. S. The Critical Role of Biomedical Research in Pandemic Preparedness. *JAMA* **318**, 1757 (2017). .

[193] Khor, S. K. & Heymann, D. L. Pandemic preparedness in the 21st century: which way forward? *The Lancet Public Health* **2667**, 8–9 (2021). .

[194] Morens, D. M., Taubenberger, J. K., Folkers, G. K. & Fauci, A. S. Pandemic Influenza's 500th Anniversary. *Clinical Infectious Diseases* **51**, 1442–1444 (2010). .

[195] Fox, S. J., Miller, J. C. & Meyers, L. A. Seasonality in risk of pandemic influenza emergence. *PLoS Computational Biology* **13**, e1005749 (2017). .

[196] Thompson, R. Pandemic potential of 2019-nCoV. *The Lancet Infectious Diseases* **3099**, 30068 (2020). .

[197] WHO. WHO Director-General's opening remarks at the media briefing on COVID-19 - 11 March 2020 (2020). .

[198] Kopecki, D., Lovelace Jr., B., Feuer, W. & Higgins-Dunn, N. WHO declares the coronavirus outbreak a global pandemic (2020). .

[199] Boseley, S. WHO declares coronavirus pandemic (2020). .

[200] Wan, W. What is pandemic? Why did WHO just declare one? (2020). .

[201] Villarreal, P. A. Pandemic: Building a Legal Concept for the Future. *Washington University Global Studies Law Review* **20**, 611–626 (2021). .

[202] Abeysinghe, S. *Pandemics, Science and Policy: H1N1 and the World Health Organization* (Palgrave Macmillan, New York, 2015). .

[203] NHS. Flu vaccine - NHS (2020). .

[204] Nichols, B. E., Boucher, C. A., van der Valk, M., Rijnders, B. J. & van de Vijver, D. A. Cost-effectiveness analysis of pre-exposure prophylaxis for HIV-1 prevention in the Netherlands: a mathematical modelling study. *The Lancet Infectious Diseases* **16**, 1423–1429 (2016). .

[205] Grulich, A. E. *et al.* Population-level effectiveness of rapid, targeted, high-coverage roll-out of HIV pre-exposure prophylaxis in men who have sex with men: the EPIC-NSW prospective cohort study. *The Lancet HIV* **5**, e629–e637 (2018). .

[206] Kirby, T. PrEP finally approved on NHS in England. *The Lancet* **395**, 1025 (2020). .

[207] WHO SAGE Working Group on COVID-19 Vaccines. Roadmap for Prioritizing Population Groups for Vaccines against COVID-19. Tech. Rep. October, World Health Organisation (2020). .

[208] Moore, S., Hill, E. M., Dyson, L., Tildesley, M. J. & Keeling, M. J. Modelling optimal vaccination strategy for SARS-CoV-2 in the UK. *PLoS Computational Biology* **17**, e1008849 (2021). .

[209] Tildesley, M. J. *et al.* Optimal reactive vaccination strategies for a foot-and-mouth outbreak in the UK. *Nature* **440**, 83–86 (2006). .

[210] Longini, I. M. *et al.* Containing pandemic influenza at the source. *Science* **309**, 1083–1087 (2005). .

[211] Piraveenan, M., Prokopenko, M. & Hossain, L. Percolation Centrality: Quantifying Graph-Theoretic Impact of Nodes during Percolation in Networks. *PLoS ONE* **8**, e53095 (2013). .

[212] Tanaka, G., Urabe, C. & Aihara, K. Random and targeted interventions for epidemic control in metapopulation models. *Scientific Reports* **4**, 5522 (2014). .

[213] Holme, P. Three faces of node importance in network epidemiology: Exact results for small graphs. *Physical Review E* **96**, 062305 (2017). .

[214] Barrat, A., Barthélemy, M., Pastor-Satorras, R. & Vespignani, A. The architecture of complex weighted networks. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 3747–3752 (2004). .

[215] Lawyer, G. Understanding the influence of all nodes in a network. *Scientific Reports* **5**, 8665 (2015). .

[216] Latora, V., Nicosia, V. & Russo, G. *Complex Networks: Principles, Methods and Applications* (Cambridge University Press, Cambridge, 2017).

[217] Zhang, Q., Karsai, M. & Vespignani, A. Link transmission centrality in large-scale social networks. *EPJ Data Science* **7**, 33 (2018). .

[218] Freeman, L. C. A Set of Measures of Centrality Based on Betweenness. *Sociometry* **40**, 35–41 (1977). .

[219] Brandes, U. A faster algorithm for betweenness centrality. *Journal of Mathematical Sociology* **25**, 163–177 (2001). .

[220] Hagberg, A. A., Schult, D. A. & Swart, P. J. Exploring network structure, dynamics, and function using NetworkX. *7th Python in Science Conference (SciPy 2008)* 11–15 (2008). .

[221] Kitsak, M. *et al.* Identification of influential spreaders in complex networks. *Nature Physics* **6**, 888–893 (2010). .

[222] Ahajjam, S. & Badir, H. Identification of influential spreaders in complex networks using HybridRank algorithm. *Scientific Reports* **8**, 11932 (2018). .

[223] Christley, R. M. *et al.* Infection in social networks: Using network analysis to identify high-risk individuals. *American Journal of Epidemiology* **162**, 1024–1031 (2005). .

[224] Enright, J. & Kao, R. R. Epidemics on dynamic networks. *Epidemics* (2018). .

[225] Borgatti, S. P. Centrality and network flow. *Social Networks* **27**, 55–71 (2005).

[226] Newman, M. E. A measure of betweenness centrality based on random walks. *Social Networks* **27**, 39–54 (2005). .

[227] Brandes, U. & Fleischer, D. Centrality measures based on current flow. *Lecture Notes in Computer Science* **3404**, 533–544 (2005). .

[228] Donnell, D. *et al.* HIV protective efficacy and correlates of tenofovir blood concentrations in a clinical trial of PrEP for HIV prevention. *Journal of Acquired Immune Deficiency Syndromes* **66**, 340–348 (2014). .

[229] Spinner, C. D. *et al.* HIV pre-exposure prophylaxis (PrEP): a review of current knowledge of oral systemic HIV PrEP in humans. *Infection* **44**, 151–158 (2016). .

[230] Zipf, G. K. The P1 P2 / D Hypothesis : On the Intercity Movement of Persons. *American Sociological Review* **11**, 677–686 (1946). .

[231] Xia, Y., Bjørnstad, O. N. & Grenfell, B. T. Measles metapopulation dynamics: A gravity model for epidemiological coupling and dynamics. *American Naturalist* **164**, 267–281 (2004). .

[232] Viboud, C. *et al.* Synchrony, waves, and spatial hierarchies in the spread of influenza. *Science* **312**, 447–451 (2006). .

[233] Truscott, J. & Ferguson, N. M. Evaluating the Adequacy of Gravity Models as a Description of Human Mobility for Epidemic Modelling. *PLoS Computational Biology* **8**, e1002699 (2012). .

[234] Local_Profil. Blank US Map, Mainland with no States (2007). .

[235] Nilfanion. English ceremonial counties 2010 (2010). .

[236] Bureau of Transport Statistics. Air Carriers: T-100 Domestic Market (2021).

[237] U.S. Census Bureau. Summary File 1 (2000). .

[238] Office for National Statistics. 2011 Census: Special Workplace Statistics (United Kingdom) (2011).

[239] Huang, C. Y., Chin, W. C. B., Wen, T. H., Fu, Y. H. & Tsai, Y. S. EpiRank: Modeling Bidirectional Disease Spread in Asymmetric Commuting Networks. *Scientific Reports* **9**, 5415 (2019). .

[240] Jarvis, C. I. *et al.* The impact of local and national restrictions in response to COVID-19 on social contacts in England: a longitudinal natural experiment. *BMC Medicine* **19**, 52 (2021). .

[241] Kribs-Zaleta, C. M. & Velasco-Hernández, J. X. A simple vaccination model with multiple endemic states. *Mathematical Biosciences* **164**, 183–201 (2000). .

[242] Russell, C. A., Real, L. A. & Smith, D. L. Spatial control of rabies on heterogeneous landscapes. *PLoS ONE* **1**, e27 (2006). .

[243] Carter, R., Mendis, K. N., Miller, L. H., Molineaux, L. & Saul, A. Malaria transmission-blocking vaccines - How can their development be supported? *Nature Medicine* **6**, 241–244 (2000). .

[244] Préziosi, M. P. & Halloran, M. E. Effects of pertussis vaccination on transmission: Vaccine efficacy for infectiousness. *Vaccine* **21**, 1853–1861 (2003). .

[245] Smith, R. J. & Blower, S. M. Could disease-modifying HIV vaccines cause population-level perversity? *The Lancet Infectious Diseases* **4**, 636–639 (2004). .

[246] Harris, R. J. *et al.* Effect of Vaccination on Household Transmission of SARS-CoV-2 in England. *New England Journal of Medicine* **385**, 759–760 (2021). .

[247] Heffernan, J. M. & Keeling, M. J. Implications of vaccination and waning immunity. *Proceedings of the Royal Society B: Biological Sciences* **276**, 2071–2080 (2009). .

[248] Crellen, T. *et al.* Dynamics of SARS-CoV-2 with waning immunity in the UK population. *Philosophical Transactions of the Royal Society B: Biological Sciences* **376**, 20200274 (2021). .

[249] Saad-Roy, C. M. *et al.* Epidemiological and evolutionary considerations of SARS-CoV-2 vaccine dosing regimes. *Science* **372**, 363–370 (2021). .

[250] Baldwin, C. L. & Runkle, R. S. Biohazards Symbol: Development of a Biological Hazards Warning Signal. *Science* **158**, 264–265 (1967). .

[251] Arinaminpathy, N. & McLean, A. R. Antiviral treatment for the control of pandemic influenza: Some logistical constraints. *Journal of the Royal Society Interface* **5**, 545–553 (2008). .

[252] McCaw, J. M. & McVernon, J. Prophylaxis or treatment? Optimal use of an antiviral stockpile during an influenza pandemic. *Mathematical Biosciences* **209**, 336–360 (2007). .

[253] McVernon, J., McCaw, J. M. & Nolan, T. M. Modelling strategic use of the national antiviral stockpile during the CONTAIN and SUSTAIN phases of an Australian pandemic influenza response. *Australian and New Zealand Journal of Public Health* **34**, 113–119 (2010). .

[254] Greer, A. L. & Schanzer, D. Using a Dynamic Model to Consider Optimal Antiviral Stockpile Size in the Face of Pandemic Influenza Uncertainty. *PLoS ONE* **8**, e67253 (2013). .

[255] Greer, A. L. Can informal social distancing interventions minimize demand for antiviral treatment during a severe pandemic? *BMC Public Health* **13**, 669 (2013). .

[256] UK Health Security Agency. Reproduction number (R) and growth rate: methodology. Tech. Rep. (2021). .

[257] Public Health England. About the data (2021). .

[258] Prem, K. *et al.* The effect of control strategies that reduce social mixing on outcomes of the COVID-19 epidemic in Wuhan, China. *The Lancet Public Health* **2667**, e261–e270 (2020). .

[259] du Plessis, L. *et al.* Establishment and lineage dynamics of the SARS-CoV-2 epidemic in the UK. *Science* **371**, 708–712 (2021).

[260] Schwartz, J. Bing Maps Tile System (2018).

[261] Office for National Statistics. Census geography (2021). .

[262] Office for National Statistics. Population estimates by output areas, electoral, health and other geographies, England and Wales: mid-2019. Tech. Rep. (2020). .

[263] Office for National Statistics. Middle Layer Super Output Areas (December 2011) Population Weighted Centroids (2019). .

[264] Bayes, T. An Essay towards solving a Problem in the Doctrine of Chances. *Philosophical Transactions of the Royal Society* **53**, 370–418 (1763). .

[265] Kruschke, J. K. *Doing Bayesian Data Analysis* (Elsevier, 2015), second edn.

[266] Lambert, B. *A Student's Guide to Bayesian Statistics* (SAGE Publications, London, 2018).

[267] Ferretti, L. *et al.* The timing of COVID-19 transmission. *medRxiv* (2020). .

[268] Prime Minister's Office. Prime Minister announces new local COVID Alert Levels (2020). .

[269] Wu, Z. & McGoogan, J. M. Characteristics of and Important Lessons From the Coronavirus Disease 2019 (COVID-19) Outbreak in China: Summary

of a Report of 72 314 Cases From the Chinese Center for Disease Control and Prevention. *JAMA* **323**, 1239–1242 (2020). .

[270] Lee, S. Y., Lei, B. & Mallick, B. Estimation of COVID-19 spread curves integrating global data and borrowing information. *PLoS ONE* **15**, e0236860 (2020). .

[271] Yang, Y., Longini Jr., I. M. & Halloran, E. M. Design and Evaluation of Prophylactic Interventions Using Infectious Disease Incidence Data from Close Contact Groups. *J R Stat Soc Ser C Appl Stat.* **55**, 317–330 (2006). .

[272] Simonis, A., Theobald, S. J., Fätkenheuer, G., Rybniker, J. & Malin, J. J. A comparative analysis of remdesivir and other repurposed antivirals against SARS-CoV-2. *EMBO Molecular Medicine* **13**, e13105 (2021). .

[273] PharmaTimes. UK Government stockpiling Tamiflu (2005). .

[274] GP. UK to boost Tamiflu stockpile to tackle swine flu (2009). .

[275] Wilson, E. B. Probable Inference, the Law of Succession, and Statistical Inference. *Journal of the American Statistical Association* **22**, 209–212 (1927). .

[276] Jefferson, T. *et al.* Neuraminidase inhibitors for preventing and treating influenza in adults and children. *Cochrane Database of Systematic Reviews* (2014). .

[277] O'Dowd, A. Government says it would stockpile Tamiflu again. *BMJ* **349**, g6386 (2014). .

[278] Jack, A. Tamiflu: "A nice little earner". *BMJ* **348**, 13–16 (2014). .

[279] Frediansyah, A., Tiwari, R., Sharun, K., Dhama, K. & Harapan, H. Antivirals for COVID-19: A critical review. *Clinical Epidemiology and Global Health* **9**, 90–98 (2021). .

[280] Adalja, A. & Inglesby, T. Broad-Spectrum Antiviral Agents: A Crucial Pandemic Tool. *Expert Review of Anti-Infective Therapy* **17**, 467–470 (2019). .

[281] Kinganda-Lusamaki, E. *et al.* Integration of genomic sequencing into the response to the Ebola virus outbreak in Nord Kivu, Democratic Republic of the Congo. *Nature Medicine* **27**, 710–716 (2021). .

[282] Bousali, M. *et al.* SARS-CoV-2 molecular transmission clusters and containment measures in ten european regions during the first pandemic wave. *Life* **11**, 219 (2021).

[283] Stokel-Walker, C. What we know about covid-19 reinfection so far. *The BMJ* **372**, n99 (2021). .

[284] Richard, Q., Alizon, S., Choisy, M., Sofonea, M. T. & Djidjou-Demasse, R. Age-structured non-pharmaceutical interventions for optimal control of COVID-19 epidemic. *PLoS Computational Biology* **17**, e1008776 (2021).

[285] Galanis, G., Di Guilmi, C., Bennett, D. L. & Baskozos, G. The effectiveness of non-pharmaceutical interventions in reducing the COVID-19 contagion in the UK, an observational and modelling study. *PLoS ONE* **16**, 1–13 (2021). .

[286] Wymant, C. *et al.* The epidemiological impact of the NHS COVID-19 App. *Nature* **594**, 408–412 (2021). .

[287] Moussa, N. & Gaba, E. File:Fuller projection rotated (land only) (2011). .

[288] Heesterbeek, H. *et al.* Modeling infectious disease dynamics in the complex landscape of global health. *Science* **347**, aaa4339 (2015). .

[289] Walters, C. E., Meslé, M. M. & Hall, I. M. Modelling the global spread of diseases: A review of current practice and capability. *Epidemics* (2018). .

[290] Thompson, R. N. Epidemiological models are important tools for guiding COVID-19 interventions. *BMC Medicine* **18**, 152 (2020). .

[291] Kraemer, M. U. *et al.* Dynamics of conflict during the Ebola outbreak in the Democratic Republic of the Congo 2018-2019. *BMC Medicine* **18**, 113 (2020). .

[292] Venkatramanan, S. *et al.* Spatio-Temporal Optimization of Seasonal Vaccination Using a Metapopulation Model of Influenza. *Proceedings - 2017 IEEE International Conference on Healthcare Informatics, ICHI 2017* 134–144 (2017). .

[293] O'Regan, S. M. *et al.* Multi-scale model of epidemic fade-out: Will local extirpation events inhibit the spread of white-nose syndrome? *Ecological Applications* **25**, 621–633 (2015). .

[294] Balk, D. *et al.* Determining Global Population Distribution: Methods, Applications and Data. *Advances in Parasitology* **62**, 119–156 (2006). .

[295] Weiss, D. J. *et al.* A global map of travel time to cities to assess inequalities in accessibility in 2015. *Nature* **553**, 333–336 (2018). .

[296] Paton, R. S. & MatthIopoulos, J. Defining the scale of habitat availability for models of habitat selection. *Ecology* **97**, 1113–1122 (2016). .

[297] Schulze-Makuch, D., Airo, A. & Schirmack, J. The adaptability of life on earth and the diversity of planetary habitats. *Frontiers in Microbiology* **8**, 2011 (2017). .

[298] Pedgley, D. E. Windborne Spread of Insect-Transmitted Diseases of Animals and Man. *Philosophical Transactions of the Royal Society of London. Series B, Biological sciences* **302**, 463–470 (1983). .

[299] West, J. S. Plant Pathogen Dispersal. *eLS* 1–7 (2014). .

[300] Barrios, J. M. *et al.* Using the gravity model to estimate the spatial spread of vector-borne diseases. *International Journal of Environmental Research and Public Health* **9**, 4346–4364 (2012). .

[301] Lourenço, J. & Recker, M. Natural, Persistent Oscillations in a Spatial Multi-Strain Disease System with Application to Dengue. *PLoS Computational Biology* **9**, e1003308 (2013). .

[302] Berestycki, H., Roquejoffre, J. M. & Rossi, L. Propagation of Epidemics Along Lines with Fast Diffusion. *Bulletin of Mathematical Biology* **83**, 2 (2021). .

[303] Beneduci, R., Bilotta, E. & Pantano, P. A unifying nonlinear probabilistic epidemic model in space and time. *Scientific Reports* **11**, 13860 (2021). .

[304] Grave, M. & Coutinho, A. L. Adaptive mesh refinement and coarsening for diffusion–reaction epidemiological models. *Computational Mechanics* **67**, 1177–1199 (2021). .

[305] Grenfell, B. T., Bjørnstad, O. N. & Kappey, J. Travelling waves and spatial hierarchies in measles epidemics. *Nature* **414**, 716–723 (2001). .

[306] Brockmann, D., Hufnagel, L. & Geisel, T. The scaling laws of human travel. *Nature* **439**, 462–465 (2006). .

[307] Viboud, C. *et al.* Risk factors of influenza transmission in households. *British Journal of General Practice* **54**, 684–689 (2004). .

[308] Ahmed, F., Zviedrite, N. & Uzicanin, A. Effectiveness of workplace social distancing measures in reducing influenza transmission: A systematic review. *BMC Public Health* **18**, 518 (2018). .

[309] Killworth, P. D., Johnsen, E. C., Bernard, H., Ann Shelley, G. & McCarty,

C. Estimating the size of personal networks. *Social Networks* **12**, 289–312 (1990). .

[310] Borgatti, S. P., Mehra, A., Brass, D. J. & Labianca, G. Network Analysis in the Social Sciences. *Science* **323**, 892–896 (2009). .

[311] Keeling, M. J. & Eames, K. T. D. Networks and epidemic models. *Journal of the Royal Society, Interface* **2**, 295–307 (2005). .

[312] Barthélemy, M. Spatial networks. *Physics Reports* **499**, 1–101 (2011). .

[313] Lang, J. C., De Sterck, H., Kaiser, J. L. & Miller, J. C. Analytic models for SIR disease spread on random spatial networks. *Journal of Complex Networks* **6**, 948–970 (2018). .

[314] Pellis, L. *et al.* Eight challenges for network epidemic models. *Epidemics* **10**, 58–62 (2015). .

[315] Parzen, E. Markov chains: discrete parameter. In *Stochastic Processes*, chap. Six, 187–275 (Courier Dover Publications, 2015).

[316] Wenham, C. What we have learnt about the world health organization from the ebola outbreak. *Philosophical Transactions of the Royal Society B: Biological Sciences* **372** (2017).

[317] Reddy, S. K., Mazhar, S. & Lencucha, R. The financial sustainability of the World Health Organization and the political economy of global health governance: A review of funding proposals. *Globalization and Health* **14**, 1–11 (2018).

[318] Group of Chief Scientific Advisors to the European Comission, European Group on Ethics in Science and New Technologies (EGE) & Special advisor to President Ursula von der Leyen on the response to the coronavirus and

COVID-19. Improving pandemic preparedness and management. Tech. Rep. (2020). .