

Optimising the surveillance and control of communicable diseases: the effects of population heterogeneity on early outbreak transmission dynamics

Francesca Lovell-Read
Merton College

A thesis submitted for the degree of
Doctor of Philosophy

September 2023

This thesis is dedicated to my father,
Gordon Lovell-Read
who always told me to Believe

Acknowledgements

First and foremost, I would like to say a massive thank you to my wonderful supervisor, Robin Thompson. Robin – thank you so much for all of your guidance, support and advice over the past four years, for your endless patience in reading and commenting on my many draft manuscripts and chapters, and for your commitment to supervising my DPhil despite moving to another university half way through. I couldn't have asked for a better supervisor, and I'll miss our weekly Zoom!

Thank you also to Philip Maini, for providing valuable advice and encouragement throughout.

Further thanks are due to each of the following collaborators, who contributed to the publications that this thesis is based on: Sebastian Funk, Uri Obolski, and Christl Donnelly (Chapter 3); Silvia Shen (Chapter 4); and Stephen Parnell and Nik Cunniffe (Chapter 6). My work has greatly benefitted from their valuable input and expertise.

I would also like to acknowledge the UKRI-BBSRC for providing the funding that supported my DPhil; the Oxford Interdisciplinary Bioscience Doctoral Training Partnership for the initial research training and ongoing academic support; the Mathematical Institute for the teaching and outreach opportunities; and Merton College, for being the most wonderful environment in which to live and learn for the past eight years.

Another big thank you must go to all my friends for making the past four years so much more than just an academic experience – to Andrew, Charlie, David, Georgia, Iona, James, Jenny, Silvia, Lachlan, Lucas, Tom G, and Tom H; and most particularly to Antonia, for being the best and most supportive friend anyone could ask for.

Finally, thank you to my mum, Kate, to my sister, Stephanie, and to Mathias, for all of your belief, support, and love – I couldn't have done it without you.

Abstract

Mathematical models are increasingly used in epidemiological settings to understand pathogen transmission and to guide the implementation of surveillance and control strategies. In this thesis, we focus on modelling transmission dynamics in the very early stages of an infectious disease outbreak, using mathematical and computational techniques to design control mechanisms that facilitate early case detection and local containment.

When an infectious pathogen first arrives in a susceptible population, the initial imported case(s) may lead to a major disease outbreak, or may instead fade out with only a few infections ever observed. In Part I of this thesis, we use SARS-CoV-2 as a case study to demonstrate how the risk that a major outbreak occurs can be calculated analytically and used as a metric for assessing the effectiveness of control interventions that aim to interrupt transmission. Initially, we consider the impact of transmission occurring from nonsymptomatic infectious individuals, and demonstrate that surveillance targeting individuals who are not showing symptoms can be crucial for outbreak containment. Then, we extend our model to account for age-dependent factors that affect SARS-CoV-2 transmission, including patterns of social contacts, variation in susceptibility to infection, and the likelihood of developing clinical symptoms. Using this age-structured model, we compare the effects of non-pharmaceutical control interventions such as school closures, workplace closures, and social distancing policies, that affect individuals differently according to their age.

For pathogens with the potential to cause a widespread outbreak, early detection through effective surveillance of the susceptible population can be crucial for the success of subsequent control measures. In Part II of this thesis, we turn to the context of plant disease to explore how early detection monitoring for invasive pathogens can be improved. Surveillance for plant pathogens is often hindered by a long incubation period during which plants may be infectious but not displaying visible symptoms. To overcome this, 'sentinel' plants – alternative susceptible host species that display visible symptoms of infection more rapidly – could be introduced to at-risk populations and included in monitoring programmes to act as early warning beacons for infection. Here, we use *Xylella fastidiosa* infection in *Olea europaea* (European olive) as a current high-profile case study to inform the construction of a computational model of pathogen transmission and surveillance. Using this model, we demonstrate that including sentinel plants in monitoring programmes could reduce the expected prevalence of infection upon outbreak detection, thereby reducing the cost of post-detection control and increasing the feasibility of local containment and eradication.

Contents

1 Introduction **1**
1.1 Mathematical modelling of infectious disease outbreaks 3
1.1.1 Compartmental models of pathogen transmission 4
1.1.2 Stochastic simulations of compartmental models 8
1.2 Early interventions for outbreak control 9
1.3 The structure of this thesis 11

I Using outbreak risk calculations to assess the optimal targeting of control interventions for COVID-19 **15**
Publication details and author contributions 16

2 Introduction to Part I **17**
2.1 What do we mean by a 'major outbreak' of a disease? 19
2.2 Major outbreaks in stochastic compartmental models 20
2.3 First-step analysis method for estimating outbreak risks 22
2.4 Case study: the COVID-19 pandemic 26
2.5 The contribution of Part I to the literature 29

3 Interventions targeting nonsymptomatic cases can be important to prevent major outbreaks **33**
3.1 Introduction 33
3.2 Methods 36
3.2.1 Model 36
3.2.2 Reproduction number 37
3.2.3 Baseline values of model parameters 39
3.2.4 Probability of a major outbreak 39
3.2.5 Computational approach for optimising surveillance enhancement . . 42
3.2.6 Code availability 43
3.3 Results 45
3.3.1 Probability of a major outbreak 45

3.3.2 Optimising surveillance enhancement 46
3.4 Discussion 49

A Supplementary material for Chapter 3 **53**
A.1 Robustness of results to parameter values used 53
A.2 The expected time to isolation, given that isolation occurs before recovery . 53
A.3 Probability of a major outbreak 54
A.4 Supplementary Figures 56

4 Estimating outbreak risks and the effects of non-pharmaceutical interventions in age-structured populations **67**
4.1 Introduction 67
4.2 Methods 69
4.2.1 Mathematical model 69
4.2.2 Reproduction number 71
4.2.3 Model parameterisation 72
4.2.4 Probability of a major outbreak 74
4.2.5 Code availability 78
4.3 Results 78
4.3.1 Effect of the age of the index case on the risk of a major outbreak . 78
4.3.2 Effect of the target age group on NPI effectiveness 80
4.3.3 Mixed strategies for reducing the major outbreak risk 82
4.3.4 Effect of surveillance on the contact-reducing NPIs required for major outbreak control 82
4.4 Discussion 85

B Supplementary material for Chapter 4 **89**

Part I: Summary and concluding remarks **103**

II ‘Sentinel’ plants for surveillance: Improving early detection monitoring for invasive plant pathogens **107**
Publication details and author contributions 108

5 Introduction to Part II **109**
5.1 Mathematical modelling for post-detection control 111
5.2 Mathematical modelling for early detection 113
5.3 Presymptomatic transmission and ‘sentinel’ plants 115
5.4 Case study: *Xylella fastidiosa* 116
5.5 The contribution of Part II to the literature 120

6 'Sentinel' plants as an early warning system for *Xylella fastidiosa* infection in *Olea europaea* 123
6.1 Introduction . . . 123
6.2 Methods . . . 125
6.2.1 Transmission model . . . 125
6.2.2 The 'baseline' case – reduced model in the absence of sentinel plants 128
6.2.3 Sampling and detection . . . 128
6.2.4 Model parameterisation . . . 129
6.2.5 Computational implementation . . . 130
6.2.6 Bayesian optimisation and its application to this thesis . . . 132
6.2.7 Code availability . . . 134
6.3 Results . . . 134
6.3.1 The baseline case – a monitoring programme without sentinel plants 134
6.3.2 Introducing sentinel plants – choosing P_S and N_S carefully is critical 135
6.3.3 Optimising the number of sentinel plants included in the population 140
6.3.4 Robustness of the results to the parameter values used . . . 144
6.4 Discussion . . . 145

C Supplementary material for Chapter 6 149
C.1 Comparing random sampling and repeated sampling . . . 149
C.2 Varying the sensitivity of visual inspection . . . 150
C.3 Variation in parameter values . . . 151
C.4 Variation in the objective function . . . 153
C.5 Supplementary Figures . . . 155

7 Seasonal surveillance for *Xylella fastidiosa*: Optimising resource allocation for early detection 171
7.1 Introduction . . . 171
7.2 Methods . . . 175
7.2.1 Incorporating seasonality into the transmission model . . . 175
7.2.2 Incorporating seasonality in symptom expression . . . 178
7.2.3 The seasonal baseline – model in the absence of sentinel plants . . . 178
7.2.4 Allowing for 'window' sampling strategies . . . 179
7.2.5 Sampling effort . . . 180
7.2.6 Application of the seasonal model . . . 181
7.2.7 Code availability . . . 184
7.3 Results . . . 184
7.3.1 The effect of seasonal factors on the conclusions of Chapter 6 . . . 184
7.3.2 Optimising the sampling window for the seasonal baseline model . . 186
7.3.3 Optimising the sampling window for the full seasonal model . . . 189

7.4 Discussion 193

D Supplementary material for Chapter 7 **197**

D.1 Construction of the seasonal scaling factor for the rate of symptom development, $\psi(t)$ 197

D.2 Construction of the seasonal detection sensitivity, $\chi(t)$ 197

D.3 Sensitivity to changes in the seasonal scaling factors $\omega(t)$, $\psi(t)$, and $\chi(t)$ . . 199

D.4 Supplementary Figures 202

Part II: Summary and concluding remarks **207**

8 Conclusion **211**

8.1 Key results of this thesis and areas for future research 213

8.1.1 Summary of Part I 214

8.1.2 Part I: potential areas for future research 217

8.1.3 Summary of Part II 220

8.1.4 Part II: potential areas for future research 223

8.2 Concluding remarks 228

Bibliography **229**

Chapter 1

Introduction

Outbreaks of infectious diseases in human, animal, and plant populations present an ongoing and increasing threat to global public health and biosecurity [1–5]. Dangerous diseases that affect humans such as COVID-19 and Ebola naturally attract the focus of public attention, but the economic and environmental consequences of disease in livestock, crops, and wild species are also severe, with particular implications for food security and the balance of natural ecosystems [4–6]. Over the last few centuries our understanding of infectious pathogens, how they are transmitted, and how we can protect ourselves against them has developed enormously, with significant breakthroughs including Edward Jenner's work on immunisation against smallpox in the late 18^{th} century, often credited as the foundation of modern principles of vaccination [7], and Sir Alexander Fleming's discovery of penicillin in 1928, which heralded the beginning of the antibiotic era [8]. However, despite such advancements the challenges associated with controlling disease outbreaks remain substantial, highlighted by the fact that only two infectious diseases have ever been successfully eradicated as a result of human intervention – smallpox, which was declared eradicated in 1980 [9], and rinderpest, a dangerous disease of cattle that was declared eradicated in 2011 [10]. Despite long-standing efforts to control them, a multitude of dangerous pathogens continue to circulate in human, animal, and plant populations – and in addition to these familiar threats, the risk of novel pathogens emerging is ever present. The recent COVID-19 pandemic has served as a stark illustration of the threat posed by pathogens that emerge in the human population as a result of zoonotic spillover, and the risk of such emergence events will only increase as human activities encroach ever further into the natural landscape [11]. For governments and other organisations responsible for public, animal, and plant health, continuing to develop an improved understanding of how to control the spread of infectious pathogens and minimise the negative effects of disease outbreaks is therefore critical.

A wide range of potential interventions may be deployed to mitigate the spread of an infectious pathogen, depending on its mode of transmission and the types of host it infects. For directly transmitted pathogens causing severe disease in humans, containment measures

commonly include the isolation of infected individuals, the tracing and subsequent isolation of people they have come into contact with, and public awareness campaigns that provide information about how to reduce the risk of infection [12–16]. The rollout of vaccination programmes is also a key strategy for reducing pathogen transmission and the burden of disease, and the development of effective vaccines remains a important focus of research for a variety of infectious pathogens [17–19]. In addition to human diseases, vaccination may also be used as a control strategy for infectious diseases of livestock; indeed, the development of an effective vaccine against rinderpest in the 1960s played an instrumental role in the successful eradication of that virus [10]. Control measures for livestock diseases also commonly include the culling of animals on premises that have detected infections or lie within a specified distance of such premises. 'Ring-culling' strategies of this kind have been considered in the context of outbreaks of foot-and-mouth disease and avian influenza, for example [20]. For plant pathogens, control interventions may similarly include the removal and destruction (roguing) of infected hosts and nearby at-risk plants to reduce transmission [21], or may alternatively rely on the use of chemical treatments such as pesticides and fungicides [22]. In any situation, effective surveillance of potential hosts for signs of infection is a crucial component of disease control strategies, since early detection of an outbreak can be critical for the success of any containment measures that are subsequently implemented [23, 24].

Although interventions such as those described above can be effective at interrupting pathogen transmission, determining the most appropriate control strategy to use in a particular scenario can be challenging. Typically, deploying an intervention requires a policy maker to make a decision regarding the scale, duration, or stringency of that measure. The potential benefits and drawbacks of different possible choices must be carefully evaluated to ensure that the optimal decision is made. For example, a policy maker seeking to implement a ring-culling strategy to contain the spread of a livestock pathogen must determine an appropriate 'cull radius', which defines the area surrounding an infected farm in which culling is to be performed [20]. The larger the choice of cull radius, the greater the number of infected hosts that will be removed from the population, and thus the greater the reduction in transmission to other hosts outside that radius. On the other hand, a very large cull radius may result in a large number of healthy animals being culled unnecessarily, which is ethically questionable and likely to result in substantial economic losses for farmers [20]. A policy maker must weigh up these advantages and disadvantages, with the optimal choice of cull radius depending on their specific goals and priorities.

A further consideration for policy makers when designing a control strategy is the level of resource that is required to implement it. Since the scale and intensity of control interventions are typically limited by economic and/or logistical constraints, it is important for policy makers to understand how to efficiently use the resources at their disposal to achieve the

maximum possible reduction in pathogen transmission and minimise the cost of an outbreak. The judicious targeting of control interventions becomes particularly important when there is heterogeneity in the population, and policy makers must determine how to allocate limited resources for control between groups that may contribute to transmission or experience infection in different ways. For example, when a vaccination programme is rolled out, the number of doses that can be administered per day is likely to be restricted by factors such as the efficiency of the supply chain and the number of medical staff who are available to deliver the vaccine. In that scenario, a policy maker must determine which individuals in the population should be prioritised for vaccination, weighing up the benefits of directly protecting those with the highest risk of experiencing severe disease, or indirectly protecting the population as a whole by targeting those who contribute the most to transmission [25].

Selecting appropriate interventions to deploy in a given disease control scenario is therefore a complex process, with many potentially conflicting factors to account for. Mathematical modelling provides an invaluable tool for projecting the effects of control measures on pathogen spread, and is increasingly used to evaluate the advantages and disadvantages of different control strategies and to guide the optimal deployment of limited resources for control.

1.1 Mathematical modelling of infectious disease outbreaks

Mathematical modelling of infectious disease outbreaks dates back to at least the mid-18$^{\text{th}}$ century, when the well-known Swiss mathematician and physicist Daniel Bernoulli performed a mathematical analysis to predict the effects of widespread variolation against smallpox on childhood mortality [26]. Thought to have originated in China or India in the mid-16$^{\text{th}}$ century, the process of variolation involved deliberately infecting people with a small amount of the variola (smallpox) virus, with the intention of triggering a mild case of smallpox that would confer lasting immunity against future, more severe, infection [27, 28]. However, this process was mistrusted due to the risk of variolation inducing a severe case of disease, as well as the possibility that the induced cases could infect other individuals in the population. Bernoulli's analysis demonstrated that when these factors were taken into account, widespread variolation would nevertheless result in a net benefit, increasing the average life expectancy at birth by over three years [26, 28]. This early example highlights the power of mathematical analyses for evaluating the potential effects of disease control strategies and weighing up their benefits and drawbacks.

Since then, the field of mathematical epidemiology has developed into a well-studied and diverse area of research, with wide-ranging applications in disease monitoring, forecasting, and control [29]. The valuable role mathematical modelling can play in public health policy has

been clearly illustrated by its widespread use during the recent COVID-19 pandemic, when global governments relied on quantitative predictions of pathogen transmission to guide the implementation of interventions such as travel restrictions, social distancing, and lockdowns [30]. Mathematical modelling is also widely used in the context of livestock diseases to predict the effects of interventions such as vaccination, ring-culling, and restrictions on the movement of animals [31–33]. For plant diseases, modelling approaches have been applied to explore the effects of different possible roguing strategies, and to optimise the spatial targeting of surveillance resources for invasive pathogens by identifying areas at high risk of infection [34, 35].

In the following subsections, we outline some of the key concepts in epidemiological modelling that we will make use of throughout this thesis.

1.1.1 Compartmental models of pathogen transmission

First described by McKendrick in 1914 [36], 'compartmental modelling' has since become one of the most widely used techniques in mathematical epidemiology. In a compartmental model, every individual in the population is assigned to one of several possible compartments, which can be thought of as 'buckets' that contain all individuals with a specific infection or symptom status. Individuals move between compartments when their status changes, at specified rates that are informed by the epidemiology of the pathogen being considered.

The most well known example of a compartmental model is the 'SIR' model, in which individuals in the population are classified as either Susceptible (S), Infected (I), or Removed (R) (where 'Removed' may refer to individuals who have recovered, died, or been isolated, depending on the specific context being considered). The numbers of individuals in each of these compartments at time t are represented by $S(t)$, $I(t)$, and $R(t)$, which evolve as individuals move between the model compartments. The total population size is given by $N = S + I + R$. In the simplest case, N remains constant, but N may also be allowed to depend on time – for example, to account for births or non-disease-related deaths in the population.

When a Susceptible individual becomes Infected, they move from the S compartment into the I compartment; and when an Infected individual recovers, dies, or is isolated, they move from the I compartment into the R compartment. Every Infected individual is assumed to be removed at a constant rate μ, so that the total rate at which individuals transition from the Infected compartment to the Removed compartment is $\mu I(t)$. However, the rate at which individuals transition from the Susceptible compartment to the Infected compartment is a little more nuanced. New infections occur at rate $c\nu I(t)/N$ per susceptible individual (i.e. at total rate $c\nu S(t)I(t)/N$) where c represents the rate at which contacts occur between individuals in the population, $I(t)/N$ represents the probability that a contact is with an

infected individual, and ν represents the rate at which such contacts result in transmission occurring. The appropriate form for the contact rate c depends on the type of transmission being considered: 'density dependent' or 'frequency dependent' transmission.

- **Density dependent transmission.** As the name suggests, in a density dependent transmission model the contact rate c depends on the population density, N/A, where N is the population size and A is the area the population covers. Typically, this is assumed to be a linear relationship, and is represented by a contact rate $c = kN/A$ for some constant k. In that case, the total rate at which infections occur becomes $k\nu S(t)I(t)/A$; that is, the infection rate is $\beta_d S(t)I(t)$ for some constant β_d.
- **Frequency dependent transmission.** In a frequency dependent transmission model, the contact rate does not depend on the population density. In such a case, the contact rate c is typically represented by a constant independent of N, so that infections occur at rate $\beta_f S(t)I(t)/N$ for some constant $\beta_f \neq \beta_d$.

A modeller seeking to apply a compartmental model to a given pathosystem must determine which of these is the most appropriate formulation for that context. In a density dependent model, the infection rate increases as the population grows. A model of this type will therefore be appropriate for a pathosystem in which an increase in the population density increases the transmission potential, as is the case for many respiratory or vector borne pathogens, for example. On the other hand, in a frequency dependent model the infection rate does not scale with the population size. This formulation may be well suited to modelling the spread of a sexually transmitted pathogen, since in that case an increase in the population size is unlikely to substantially affect the number of contacts an infected host has with susceptible individuals. Alternatively, a frequency dependent model could be appropriate when the population is assumed to grow in space at the same time as it grows in size, so that the overall population density does not change. Of course, if the population size N is assumed constant, the density dependent and frequency dependent models are exactly equivalent. In that case, a modeller may select either formulation, and must simply take care to ensure that β_d or β_f is of the correct order of magnitude when parameterising the model.

A schematic of the (frequency dependent) SIR model showing its compartments and the rates at which transitions between compartments occur is shown in Fig 1.1A.

Compartmental models that describe more complex pathosystems can be constructed straightforwardly by introducing additional model compartments and specifying the corresponding transition rates. Commonly used model structures include the Susceptible-Exposed-Infectious-Removed (SEIR) model (in which the additional 'E' compartment represents 'Exposed' individuals who are infected but not yet able to transmit the pathogen to others)

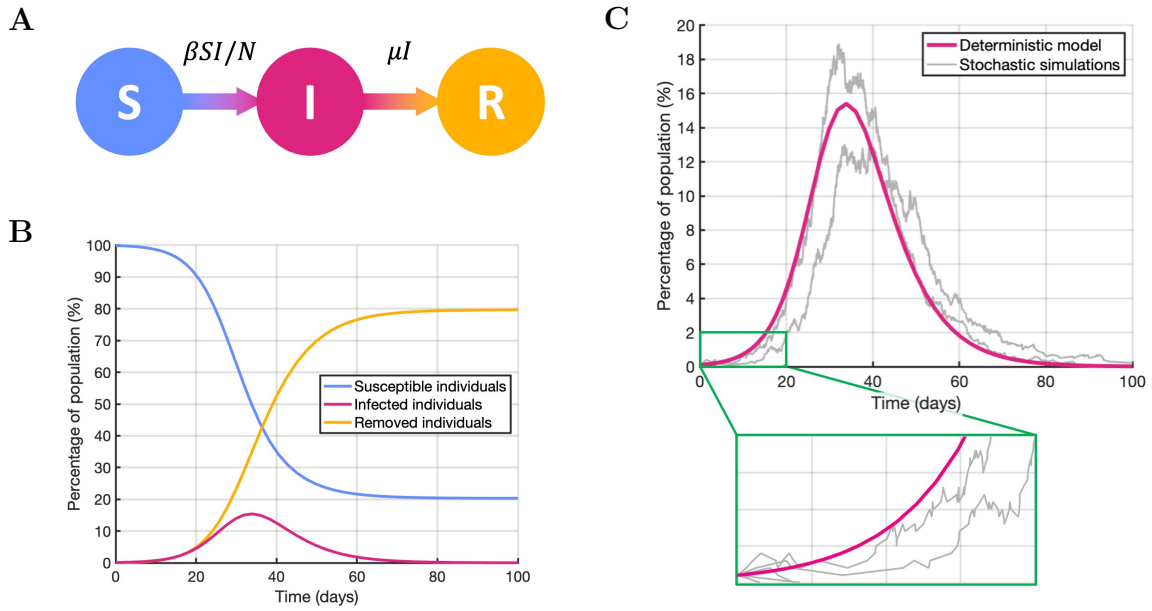


Figure 1.1: The SIR model. A. Schematic illustrating the compartments of the SIR model and the rates at which transitions between those compartments occur. The S, I, and R compartments respectively represent individuals who are susceptible to infection, individuals who are currently infected, and individuals who have been removed from the model system (due to recovery, death, or isolation). The quantity β/N represents the rate at which contacts occur between each possible pair of infected and susceptible individuals, multiplied by the probability that a contact leads to a new infection occurring. The parameter μ represents the per capita rate at which infected individuals are removed from the model system. **B.** The deterministic solution to the system of ordinary differential equations (ODEs) describing the SIR model (Equations (1.1)), showing how the numbers of susceptible, infected, and removed individuals evolve over time. In this figure, $N = 1000$, $S(0) = 999$, $I(0) = 1$, $R(0) = 0$, $\beta = 0.4$, and $\mu = 0.2$. **C.** The deterministic solution for $I(t)$ derived from Equations (1.1) (as shown in panel B; pink line), and individual realisations of $I(t)$ obtained through five separate stochastic simulations of the model system (grey lines). All model parameters are the same as in panel B. The magnified area marked by the green box shows that only two of the five simulated stochastic trajectories took off, with the other three dying out very rapidly. This demonstrates that, unlike deterministic models, stochastic simulations are able to capture the inherent randomness associated with pathogen transmission when the number of infected individuals is small.

[37, 38], and the SECIR model, in which Susceptible (S) individuals first become Exposed (E; infected, but not infectious or symptomatic), then Cryptic (C; infectious, but not symptomatic), then Infected (I; infectious and symptomatic), before finally being Removed (R) [34, 39]. Compartmental models can similarly be extended to account for heterogeneities in the population such as age or existing partial immunity to infection [40, 41], or to explicitly incorporate vector transmission by tracking vector population dynamics and host-vector interactions [42].

The time evolution of a compartmental model can be represented by a system of ordinary differential equations (ODEs), each of which describes the rate at which individuals enter

and leave a given compartment. For example, the dynamics of the frequency dependent SIR model are described by the following system of ODEs:

$$\frac{\mathrm{d}S(t)}{\mathrm{dt}} = -\frac{\beta S(t)I(t)}{N}, \quad \frac{\mathrm{d}I(t)}{\mathrm{dt}} = \frac{\beta S(t)I(t)}{N} - \mu I(t), \quad \frac{\mathrm{d}R(t)}{\mathrm{dt}} = \mu I(t), \tag{1.1}$$

with initial conditions $S(0) = S^*$, $I(0) = I^*$, and $R(0) = R^*$ (the density dependent formulation is almost identical, but without the factors of $1/N$ scaling the transmission rates in the first two equations). Given values for the parameters N, β, and μ, this ODE system can be solved numerically to determine how the numbers of Susceptible, Infected, and Removed individuals in the population change over time (see Fig 1.1B for an example).

Compartmental models are commonly used to predict important quantities relating to a disease outbreak, such as the total number of individuals who will ever be infected, the overall duration of the outbreak, the peak incidence (the maximum number of individuals who will ever be infected simultaneously), and the timing of that peak [43, 44]. As well as being useful for epidemic preparedness and planning, estimates of these quantities may crucially be used to investigate the effects of different possible control measures on the size and duration of the outbreak, and to guide the timing of their deployment and removal [37, 45, 46]. For example, Prem *et al.* [37] used an age-structured SEIR model of SARS-CoV-2 transmission to investigate the effects of interventions that reduce the numbers of contacts between individuals on the progression of COVID-19 outbreaks, demonstrating how such interventions could be used to delay the epidemic peak and reduce its incidence. Similarly, Chowell *et al.* [38] used an SEIR model of Ebola virus transmission in Uganda and the Democratic Republic of the Congo to investigate how the timing of control interventions influenced the total number of individuals likely to be infected in an Ebola outbreak, showing that the final size of an outbreak increases exponentially with the time taken to implement control interventions.

A key quantity that is commonly used to express the transmission potential of a pathogen is the 'basic reproduction number', R_0, which represents the expected number of secondary infections that will be generated by a single infected individual in a fully susceptible population. An expression for R_0 can typically be written down for a given compartmental model based on the relative rates at which infections occur and at which infectious individuals are removed from the system. For example, for the SIR model the basic reproduction number is given by

$$R_0 = \beta/\mu. \tag{1.2}$$

The value of the basic reproduction number is a key indicator of how an outbreak will evolve, with a crucial threshold being whether R_0 is greater than or less than one [47]. If $R_0 > 1$, the average number of secondary cases generated by each infected individual is more than one, and the number of infected hosts in the population will increase. If instead $R_0 < 1$,

the overall number of infections will decrease, as infectious hosts are removed from the population more rapidly than new infections are generated. The basic reproduction number is typically estimated for a given pathogen based on observed data from real outbreaks.

1.1.2 Stochastic simulations of compartmental models

As an alternative to using the deterministic formulation, the time evolution of a compartmental model system may instead be obtained by performing stochastic simulations of pathogen transmission. One commonly used approach for doing this is the 'direct method' version of the Gillespie Stochastic Simulation Algorithm (SSA; see Algorithm 1.1) [48]. This algorithm was originally developed in the context of simulating reactions in chemical systems, but can also be used to simulate the progression of infectious disease outbreaks in an analogous way. Instead of considering different types of chemical reactants and products, we consider the individuals belonging to each of the different model compartments, and instead of tracking the occurrence and outcomes of chemical reactions, we track the infections, recoveries, and other epidemiological events that may occur within the model system. Beginning from the initial state of the system, at each step the Gillespie SSA considers the different possible compartmental transition events that could occur in the model, and calculates the propensity of each of these events to occur based on the number of individuals in each compartment at that time. Then, in essence, it tosses a dice weighted by these propensities to select which of the possible transitions occurs next, and updates the number of individuals in each compartment accordingly, as well as the time at which the next event occurs. The algorithm outputs simulated trajectories for the time evolution of the number of individuals in each compartment, with every simulation leading to a different trajectory (Fig 1.1C).

A key advantage of using stochastic simulations to generate epidemic trajectories is that they are able to account for the inherent randomness in transmission that occurs as a result of variation in individual behaviour and response to infection. This randomness may substantially influence the behaviour of the system, particularly at the very beginning of an outbreak when the number of infected individuals in the population is small. In that scenario, the existing cases could lead to a large scale outbreak, but randomness in individual behaviour could alternatively mean that the pathogen dies out after only a few cases have occurred, without ever causing a large outbreak. Unlike deterministic models, stochastic simulations can reproduce these two types of behaviour for a single set of assumed parameter values, as can be observed in Fig 1.1C. Therefore, although performing a large number of stochastic simulations of a model system is more computationally intensive than numerically solving a system of ODEs, it provides a more nuanced picture of the transmission dynamics. In this thesis, we are interested in pathogen transmission and control at the very beginning of infectious disease outbreaks, and will therefore use the stochastic simulation approach throughout.

The Gillespie Stochastic Simulation Algorithm
Consider a compartmental epidemiological system in which n different possible events can occur, denoted $R_1, R_2, \ldots, R_n$. Let $\alpha_i(t)$ be the propensity function of each event R_i at time t; that is, $\alpha_i(t)\mathrm{d}t$ represents the probability that event R_i occurs in the time interval $[t, t+\mathrm{d}t]$.

Set the initial time $t = 0$ and specify the initial numbers of individuals in each compartment within the model system. Define a stopping condition T_{stop} for the simulation. Until the stopping condition is fulfilled, repeat the following:

1. Generate two uniformly distributed random numbers $r_1, r_2 \overset{i.i.d.}{\sim} U[0,1]$.
2. Compute the propensity function $\alpha_i(t)$ of each event R_i at time t.
3. Compute the total event propensity, α_0, as
$$\alpha_0 = \sum_{i=1}^{n} \alpha_i(t).$$
4. Compute the time τ until the next event occurs in the system as
$$\tau = \frac{1}{\alpha_0} \ln\left(\frac{1}{r_1}\right).$$
5. Determine which event R_j occurs at time $t + \tau$ by finding j such that
$$\frac{1}{\alpha_0} \sum_{i=1}^{j-1} \alpha_i(t) \leqslant r_2 < \frac{1}{\alpha_0} \sum_{i=1}^{j} \alpha_i(t)$$
(note that the left hand sum is taken by convention to be 0 when $j = 1$).
6. Update the numbers of individuals in each model compartment following the occurrence of event R_j accordingly.
7. Set $t = t + \tau$.

Algorithm 1.1: The direct method version of the Gillespie Stochastic Simulation Algorithm (SSA) for simulating the time evolution of a compartmental epidemiological model [48, 49].

1.2 Early interventions for outbreak control

Once a pathogen has invaded a population and widespread transmission has occured, it can be very difficult for containment measures to successfully bring the outbreak under control. Therefore, there is clear motivation for policy makers to implement control interventions in the very early stages of an outbreak, with the intention of eradicating or containing the

spread of the pathogen before it has reached a high prevalence in the population. If this can be achieved, the burden of disease on the population and the economic and societal costs of the outbreak will be reduced.

In some cases, the introduction of a pathogen to a population may be anticipated. For example, at the beginning of the COVID-19 pandemic when observed cases were limited to China, there was a clear risk that SARS-CoV-2 would soon be introduced to other countries worldwide as a result of international travel [13, 50]. Similarly, since plant pathogens are commonly transported from regions of known infection into new locations due to the global trade in live plant material, plant growers may be on high alert for signs of pathogens that are known to be spreading in other regions and may have arrived via infected imported stock [51–54]. When a population is known to be at risk of invasion by a particular infectious pathogen, policy makers may choose to deploy pre-emptive surveillance programmes that aim to identify the pathogen soon after it has arrived in the population. In the case of COVID-19, for example, countries worldwide attempted to catch imported infections by performing temperature checks on passengers arriving at airports [55]. Similarly, in many countries phytosanitary regulations require health inspections to be carried out on imported plant material in an effort to identify infected hosts and prevent them from progressing further through the trade network [52, 54].

In other instances the arrival of a pathogen in a population may occur more unexpectedly. For example, a novel infectious pathogen may emerge without warning in a human population as a result of zoonotic spillover, which occurs when a pathogen that is circulating within an animal reservoir makes the jump to a human host [3, 56]. High profile examples of pathogens whose initial emergence has been traced back to animal origins include HIV, which is thought to have originated in chimpanzees [57], the Ebola virus, thought to have originated in bats [58], and the coronaviruses responsible for the 2002-2003 SARS outbreak and the COVID-19 pandemic, which are also thought to have their origins in bats [59, 60]. Unexpected invasions of this kind are unlikely to be identified as quickly as anticipated threats, making the subsequent rapid deployment of control interventions that aim to disrupt the transmission of the pathogen particularly important if the outbreak is to be contained during its early stages.

Despite the benefits of early detection and the subsequent rapid containment of invading pathogens, in practice this can be difficult to achieve. Structured programmes of intensive surveillance that aim to detect a pathogen whilst it is still at a low prevalence in the population can be expensive and logistically challenging to implement. When designing a surveillance strategy, a policy maker must therefore determine how best to make use of limited resources. As mentioned previously, this is particularly important given the heterogeneity that is

present within populations, which should be accounted for when deciding how to allocate resources optimally between different groups. For example, in the case of a vector-borne pathogen, surveillance could be targeted towards either potential hosts, potential vectors, or a combination of both [61]. The effects of different possible control measures that may be deployed to prevent the further transmission of the pathogen once it has been detected in the population must also be carefully evaluated if containment is to be successful. Although it is crucial to make sure that the measures used are sufficient to prevent the wider spread of the pathogen, it is also important to ensure that the interventions deployed are not unnecessarily stringent, or their costs – both economic and societal – may not be justified. Once again, heterogeneities in the population may influence the effectiveness of different types of control intervention, and this must be considered when planning measures that aim to limit transmission.

In this DPhil thesis, we will use mathematical models to explore how limited resources for surveillance and control can be optimally deployed against the spread of an infectious pathogen in the very early stages of a disease outbreak. Using two major current infectious disease threats as case studies, selected to represent two very different host-pathogen systems, we will consider how heterogeneities in the disease-related characteristics of the population influence the early outbreak transmission dynamics. In particular, we will investigate how those heterogeneities can be exploited when designing targeted control interventions that aim to identify and contain an infectious pathogen soon after it has invaded a population.

1.3 The structure of this thesis

This thesis is composed of two main parts, each corresponding to a key question relating to the optimal deployment of resources for pathogen surveillance and control in the early stages of an outbreak. Here, we provide a brief overview of the structure of this thesis and the main points addressed in each section. The overall thesis structure is summarised in Fig 1.2.

Part I – Using outbreak risk calculations to assess the optimal targeting of control interventions for COVID-19

When an infectious pathogen first arrives in a susceptible population, there is a possibility that sustained transmission will occur, leading to a major outbreak of disease. However, it is also possible that the initial cases will fade out rapidly, with only a small number of infections ever observed. Which of these two possibilities occurs depends on the epidemiology and transmission characteristics of the pathogen in question, the disease-related and behavioural characteristics of the initially infected hosts, and any control interventions that are already in place. By quantifying the probability that the initial cases will indeed trigger a major outbreak, the need for further measures that aim to reduce this risk and their

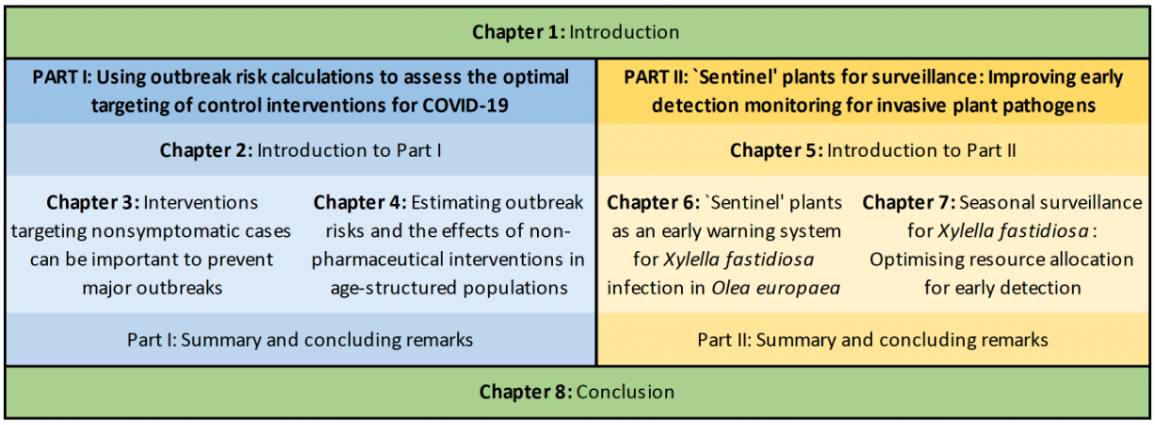


Figure 1.2: The structure of this thesis.

relative effectiveness can be assessed. In Part I of this thesis, we will use SARS-CoV-2 (the pathogen responsible for the recent COVID-19 pandemic) as a case study to demonstrate how the probability that an imported pathogen leads to a major disease outbreak can be calculated, and to explore how this outbreak risk can be used as a metric to evaluate the efficacy of different types of control interventions in the early outbreak. In particular, we will consider how heterogeneities associated with SARS-CoV-2 transmission influence the outbreak probability, and explore the optimal targeting of control interventions in the population based on those heterogeneities.

We begin in Chapter 2 by defining what we mean by a 'major outbreak' of disease, and demonstrating how the probability that a single infectious host triggers a major outbreak can be calculated analytically for a specified compartmental model of pathogen transmission. In Chapter 3, we apply this methodology to a model of SARS-CoV-2 transmission, in which we explicitly account for infections arising from individuals who are not showing clinical symptoms. By estimating the probability that a major outbreak occurs starting from a single symptomatic or nonsymptomatic infected host, we investigate how limited surveillance resources can be optimally allocated between these two groups to minimise the risk that a major outbreak occurs. In Chapter 4, we then extend our model to account for the age structure of the population and the associated heterogeneities in susceptibility to SARS-CoV-2 infection, the probability of developing clinical symptoms, and expected patterns of social contacts. Using this age-stratified model, we explore how the risk of triggering a major outbreak varies depending on the age of the initial infected host, and assess the efficacy of control interventions such as school closures and workplace closures that target different age groups within the population. The key messages identified in these research chapters are summarised in a concluding section at the end of the Part.

Part II – 'Sentinel' plants for surveillance: Improving early detection monitoring for invasive plant pathogens

When a population is at risk of invasion from a known infectious pathogen, effective surveillance for infected hosts is crucial to ensure that potential outbreaks are identified early and brought under control before the pathogen spreads widely. For infectious pathogens that affect plant populations, surveillance strategies typically rely on the visual inspection of potential hosts for symptoms of infection, since large scale laboratory testing is prohibitively expensive. However, many plant diseases have a long incubation period (time from infection to symptom onset), during which infected hosts can be infectious without displaying visible symptoms. This period of asymptomatic infection can substantially hinder the effectiveness of monitoring programmes that seek to achieve early detection. In Part II of this thesis, we will investigate a possible strategy for overcoming this problem. Specifically, we will consider how 'sentinel' plants – alternative susceptible species that display visible symptoms of infection more rapidly – could be introduced to at-risk populations and included in surveillance programmes to act as early warning beacons for infection. Using *Xylella fastidiosa* (one of the most dangerous plant pathogens worldwide) as a case study, we will explore whether induced population heterogeneity of this kind could be used to improve the perfomance of surveillance strategies that aim to detect plant disease outbreaks in their very early stages.

We begin in Chapter 5 by discussing the importance of effective surveillance for invasive plant pathogens and describing the potential advantages and disadvantages of using sentinel plants as a strategy to overcome the challenges posed by asymptomatic infection. Additionally, we provide background and motivation for our choice of case study – *Xylella fastidiosa* infection in European Olive (*Olea europaea*). In Chapter 6, we then construct a compartmental transmission model that accounts for two different types of plants – crops and sentinels – and perform stochastic simulations of pathogen transmission and surveillance to explore whether sentinel based surveillance strategies could be used to achieve detection of an invading pathogen at a lower prevalence than traditional, non-sentinel based approaches. Finally, in Chapter 7 we extend our model to account for seasonal factors that lead to variation in the rate of pathogen transmission, the rate of symptom development, and the strength of symptom expression throughout the year. As well as investigating whether our previous conclusions regarding the usefulness of sentinel plants to facilitate early detection still hold when these seasonal factors are accounted for, we additionally explore how limited resources for surveillance can be optimally allocated throughout the year based on the resulting seasonal variation in the probability of pathogen detection. Part II concludes with a brief summary of the key results generated in these chapters.

In Chapter 8, we discuss the main results of this thesis, and consider the open questions in early outbreak modelling that remain for future research to address. Supplementary material

for each of the four research chapters presented here can be found in the appendices at the end of the thesis.

Although the case studies used in the two parts of this thesis – SARS-CoV-2 and *Xylella fastidiosa* – may seem disparate at first glance, close parallels may be drawn between the modelling approaches applied and the objectives addressed in each context. In both cases, we seek to understand the optimal targeting of resources for pathogen surveillance and control in the early stages of an outbreak, within a population characterised by heterogeneities in the transmission risk between infectious hosts and susceptible individuals. Accounting for and overcoming the challenge of asymptomatic and/or presymptomatic infection is also a connecting theme. By considering case studies from both human and plant disease epidemiology in this thesis, we aim to demonstrate the broad applicability of early outbreak modelling techniques in different epidemiological settings.

Part I

Using outbreak risk calculations to assess the optimal targeting of control interventions for COVID-19

Publication details and author contributions

The work presented in Part I of this thesis has been adapted for publication in the following peer-reviewed journal articles:

- **F.A. Lovell-Read**, S. Funk, U. Obolski, C.A. Donnelly, and R.N. Thompson, 'Interventions targeting nonsymptomatic cases can be important to prevent local outbreaks: SARS-CoV-2 as a case-study'. *J. R. Soc. Interface, 2021, 18(178), 20201014. DOI: 10.1098/rsif.2020.1014.*

  CRediT author contributions for this work are as follows. Conceptualisation: All authors. Methodology: FALR, UO, RNT. Investigation: FALR. Writing – original draft: FALR, RNT. Writing – review and editing: All authors. Supervision: RNT.

  Material published in this article comprises the majority of Chapter 3 of this thesis, as well as part of Chapter 2.

- **F.A. Lovell-Read**, S. Shen and R.N. Thompson, 'Estimating local outbreak risks and the effects of non-pharmaceutical interventions in age-structured populations: SARS-CoV-2 as a case study'. *J. Theor. Biol., 2022, 535, 110983. DOI: 10.1016/j.jtbi.2021.110983.*

  CRediT author contributions for this work are as follows. Conceptualisation: All authors. Methodology: FALR, RNT. Investigation: FALR, SS. Writing – original draft: FALR, RNT. Writing – review and editing: All authors. Supervision: RNT.

  Material published in this article comprises Chapter 4 of this thesis, as well as part of Chapter 2.

Thank you to each of the individuals with whom we collaborated on the above studies for their valuable input to the work presented in this thesis.

Chapter 2

Introduction to Part I

When an infectious pathogen arrives in a susceptible population, if transmission between individuals is viable (either directly or through an appropriate vector), there is a risk that it may lead to a major disease outbreak with severe public health consequences. However, this outcome is by no means certain – it is also possible that the outbreak will fade out, with only a small number of cases ever observed. Indeed, both of these outcomes may be observed for different introductions of the same pathogen into a susceptible population. For example, amongst the recorded outbreaks of Ebola Virus Disease (Ebola) in Africa between 1976 and the present day, the total number of recorded cases has varied substantially (Fig 2.1). In several instances, the 'outbreak' was limited to ten cases or fewer, and sometimes did not spread beyond the index case; at the other extreme, during the 2014-2016 outbreak – the largest outbreak of Ebola on record – over $28,000$ cases were recorded [62]. Whether an outbreak of a given disease takes off or fades out after the initial cases arrive in the population depends on the epidemiological characteristics of the pathogen, the environmental suitability for transmission, the social structure and cultural practices of the population in question, and any control measures that may be in place [40, 63]. Additionally, there is an element of chance that arises due to the unpredictable characteristics and behaviour of individual potential hosts, as well as some randomness in whether pathogen transmission occurs between infectious and susceptible individuals who come into contact with one another [40]. For this reason, it makes sense to consider a major outbreak as an event that, following pathogen introduction, may occur with a certain probability.

Understanding the probability that pathogen introduction to a population will lead to a major disease outbreak is crucial for health policy makers to determine the most appropriate response to a new disease threat [42, 63]. If there is a risk that a major outbreak will occur, a policy maker must decide whether to introduce control interventions that aim to contain the potential outbreak at a low prevalence, and consider which interventions are likely to be the most effective. If the outbreak risk is high, more stringent control measures may be required, but these may not be necessary or justifiable if the outbreak risk is low.

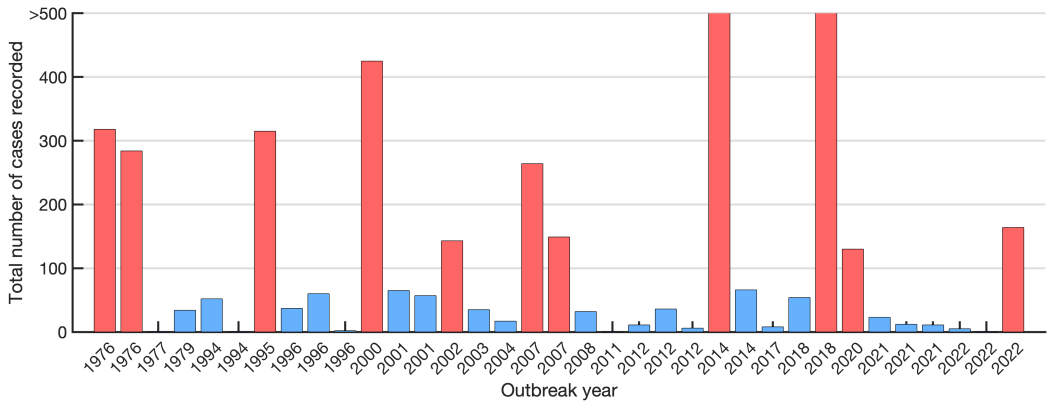


Figure 2.1: Total numbers of recorded cases in Ebola outbreaks in Africa from 1976 to 2022 [62]. Blue bars indicate outbreaks in which fewer than 100 cases were recorded; red bars indicate outbreaks in which more than 100 cases were recorded. Outbreaks spanning multiple years are displayed according to the first year in which cases were detected. The y-axis is cut off at 500 so that smaller outbreaks can be seen more easily; in the large outbreaks beginning in 2014 and 2018, 28,646 and 3,470 cases were recorded, respectively.

Quantitative approaches for estimating the probability that a major outbreak occurs when a pathogen arrives in a population are therefore essential tools for policy makers, since they can be used to assess and compare the effectiveness of potential control interventions. By computing the outbreak probability under different possible control scenarios, policy makers can determine which control interventions are likely to be the most effective at reducing the outbreak risk, and design a package of interventions that, when used in combination with each other, will reduce the probability of an outbreak below an acceptable threshold level [64–66]. Correspondingly, such methods could also be used to determine how and when existing control measures can be safely relaxed – for example, to assess whether previously enforced control measures are still required to suppress the outbreak probability following the rollout of a vaccination programme [67].

Methods for estimating the probability that a major outbreak occurs are particularly useful when there is heterogeneity in the population, and policy makers must determine how to divide resources for surveillance and control between groups with different characteristics. For example, for many infectious pathogens the age of a potential host strongly influences their susceptibility to infection, as well as the severity of the symptoms they express once infected and the likelihood that they transmit the pathogen to others [68, 69]. If the approach that is used to calculate the outbreak probability is adapted to consider different groups within the population that exhibit heterogeneous transmission characteristics, policy makers could identify the specific groups that have a higher risk of triggering an outbreak, and focus surveillance and control interventions accordingly. Targeting high risk groups within the population is particularly important when resources for control are limited, to ensure that

the maximum risk reduction is achieved with those that are available.

Developing flexible methods for estimating the probability of a major outbreak for a particular pathogen in a given population is therefore an important area of public health and epidemiology research, and is the focus of Part I of this thesis. However, before we can begin to consider how to quantify the probability that a pathogen initiates a major outbreak when it arrives in a susceptible population, it is first necessary to clarify what we mean by the term 'major outbreak'.

2.1 What do we mean by a 'major outbreak' of a disease?

Various terms are commonly used by epidemiologists and public health organisations to describe large outbreaks of infectious disease, including 'epidemic', 'pandemic', and 'Public Health Emergency of International Concern'. However, the definitions of such terms are typically imprecise, and their interpretations may vary depending on the specific situation being considered. For example, the term 'epidemic' has been used to describe *"the occurrence in a community or region of cases of an illness ... clearly in excess of normal expectancy"* [70]. However, the 'normal expectancy' of an illness in a population is highly dependent on the illness in question, and the number of cases required for an outbreak to be declared an epidemic is thus highly subjective. Likewise, although a 'pandemic' is commonly defined to be *"an epidemic occurring over a very wide area, crossing international boundaries, and usually affecting a large number of people"* [70], it is unclear how many people or countries must be affected before this term is deemed appropriate. For example, the 2014-2016 Ebola outbreak in West Africa is not usually referred to as a pandemic, despite large numbers of infections occurring across Guinea, Liberia, and Sierra Leone, and further cases being observed in seven other countries [71]. Similarly, following the initial cases observed in December 2019, the World Health Organisation did not declare COVID-19 to be a pandemic until 11^{th} March 2020, although by that time over $118,000$ cases had already been recorded in 114 countries [72].

What constitutes a 'major outbreak' of an infectious disease is similarly highly subjective, and must be defined on a case-by-case basis according to the severity of disease caused by the pathogen in question, and the perceived risk to public health. Thompson *et al.* [42] describe the three following metrics that a policy maker could use to classify an outbreak as 'major' or 'minor', based on important practical considerations.

1. **The total size.** A metric that is commonly used to distinguish between 'major' and 'minor' outbreaks is the total number of infected cases recorded during the course of the outbreak. According to this metric, a outbreak is classified as 'major' if the total number of observed cases exceeds a specified threshold. For a policy maker, a practical

choice for such a threshold could represent, for example, the maximum amount of doses of a vaccine or drug that are available. Then, a 'major' outbreak would be one in which the available resources for treatment are insufficient.

2. **The concurrent size.** A related metric that could be used to classify an outbreak as 'major' or 'minor' is the concurrent size of the outbreak; that is, the total number of individuals who are ever infected simultaneously. This is of practical relevance, since if the concurrent size exceeds the total capacity for treatment at any given time, healthcare providers may be overwhelmed. In this case, a 'major' outbreak would be one in which the concurrent size exceeds a specified threshold, selected to reflect practical constraints on the healthcare capacity (such as the total number of hospital beds available, for example).

3. **The duration.** An outbreak could also be described as 'major' if it is of particularly long duration, necessitating the lengthy imposition of costly control interventions. Additionally, the longer an outbreak goes on, the greater the risk that the causative pathogen may mutate into a more transmissible or severe strain. According to this metric, a 'major' outbreak is one that persists beyond a threshold length of time.

Although each of these definitions may be appropriate in a given context, they still require a decision to be made regarding the threshold number of cases, or outbreak duration, beyond which an outbreak is considered to be 'major'. However, an alternative definition that does not require such a decision can often be obtained through stochastic simulations of the relevant compartmental transmission model, in which there is typically a clear natural distinction between outbreaks that fade out rapidly and those that take off.

2.2 Major outbreaks in stochastic compartmental models

In stochastic simulations of compartmental epidemiological models, there is often a natural distinction between 'major' and 'minor' outbreaks. If R_0 is larger than but not close to one, when repeated stochastic simulations of the transmission model are run beginning from a small number of initial infected hosts, two qualitatively different types of behaviour are typically observed: either the pathogen fades out rapidly, or case numbers begin to increase exponentially (until the transmission rate slows due to depletion of susceptible individuals, or a public health measure is introduced). Therefore, the final epidemic size is typically distributed bimodally, with the total number of individuals ever infected falling into one of two distinct ranges [42, 73, 74]. In that scenario, a natural definition for the probability of a major outbreak is therefore the proportion of outbreak simulations for which the total number of infected individuals falls into the higher of these two ranges.

To illustrate this concept, we used the Gillespie SSA (Algorithm 1.1) to perform $100,000$ simulations of a stochastic SIR model with basic reproduction number $R_0 = 2$, beginning from a single infectious host each time. Each simulation was run until the total number of infected hosts (I) reached zero, at which point the total number of individuals ever infected (the final epidemic size) was recorded (Fig 2.2A). In 50% of the simulations, fewer than 20 infections occurred in total (left hand peak); the initial cases did not lead to sustained transmission in the population. In the remaining 50% of the simulations (major outbreaks), between 620 and 900 individuals were infected in total each time (right hand peak). There is thus a natural partitioning between simulations that faded out quickly and those that went on to become major outbreaks. Correspondingly, a natural definition for the probability of a major outbreak when a single infectious host is introduced into the population is the proportion of simulations for which the final epidemic size lies within this natural upper range (in this case equal to 0.5).

If we are only considering the initial phase of potential major outbreaks, and do not require precise knowledge of the final size distribution, an alternative to performing stochastic simulations of the full compartmental model is to instead consider the analogous branching process model. In a branching process model of pathogen transmission, depletion of the susceptible population is not accounted for, and therefore every infected individual has the potential to generate an infection lineage independently of other previous and current infected hosts [40]. Computationally speaking, this requires a simple adjustment to the Gillespie SSA (Algorithm 1.1) so that the number of susceptible individuals does not decrease when a new infection occurs. If repeated simulations of branching process models are run, then in each simulation the pathogen either fades out with few infections, or case numbers generally increase indefinitely (until hitting a pre-defined limit at which the simulation is stopped). The probability of a major outbreak starting from a small number of infected hosts then corresponds to the proportion of simulations in which the pathogen does not fade out quickly and case numbers increase indefinitely instead. This again provides a natural definition of a major outbreak, since simulations can be partitioned into two distinct sets.

To demonstrate the equivalence of these two approaches, we performed another set of $100,000$ simulations, this time using the branching process formulation of the SIR model with $R_0 = 2$. Again beginning from a single infectious host each time, simulations were run until either the total number of infectious hosts (I) reached zero, or $1,000$ infections had occurred. Once again, the total number of recorded infections was plotted (Fig 2.2B). There is again a natural partitioning between simulations that fade out quickly and those that go on to become major outbreaks, with the proportion of simulations in which a large number of infections occured again equal to 0.5.

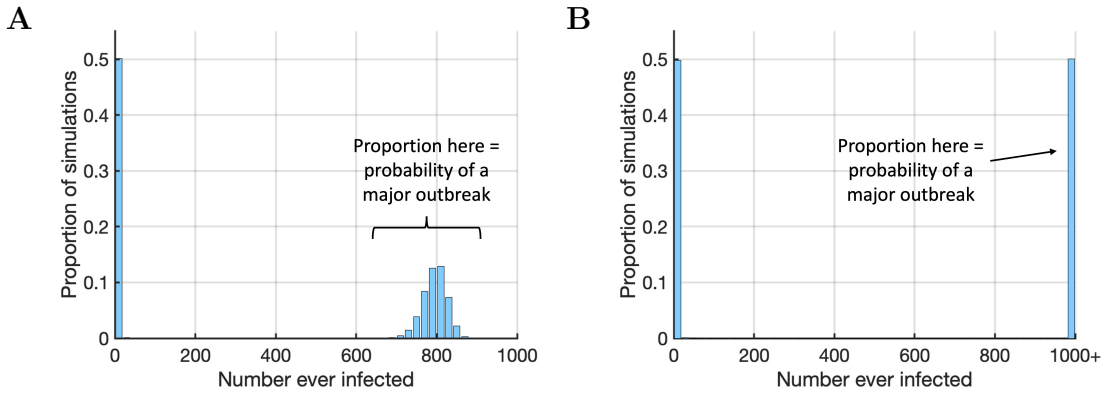


Figure 2.2: Illustrating how the probability of a major outbreak in the SIR model can be obtained through repeated stochastic simulations. A. The total number of individuals ever infected (final epidemic size) in each of $100,000$ simulations of a stochastic SIR model in a population of size $1,000$, beginning from a single infectious host each time. Model parameters were $\beta = 0.4$, $\mu = 0.2$, giving a basic reproduction number of $R_0 = 2$. In 50% of simulations, fewer than 20 infections occurred in total (left hand peak); initial cases did not lead to sustained transmission in the population. In the remaining 50% of simulations (major outbreaks), between 620 and 900 individuals were infected in total each time. The probability of a major outbreak is then defined as the proportion of simulations for which the final epidemic size lies within this natural upper range, here equal to 0.5. **B.** The analogous figure to panel A, but for the branching process version of the SIR model in which depletion of susceptibles is not accounted for (i.e. $S = N$ throughout each simulation). Simulations were run until $1,000$ infections had occurred or until the number of infected individuals in the population reached zero. The rightmost bar corresponds to simulations in which $1,000$ infections occurred. There is again a natural partitioning between simulations that fade out quickly and those that go on to become major outbreaks, with the probability of a major outbreak matching the equivalent value for the stochastic SIR model (panel A).

2.3 First-step analysis method for estimating outbreak risks

As we have seen, a natural definition for the probability of a major outbreak for a given transmission model can be obtained by running large numbers of stochastic simulations and recording the proportion of those simulations in which the outbreak does not fade out quickly, but instead goes on to generate a large number of infections. However, although this approach is straightforward, obtaining outbreak probabilities through repeated stochastic simulations can be computationally expensive, particularly for more complex transmission models. Where possible, it is therefore often desirable to use an analytic approach instead.

Two analytic methods for calculating the probability of a major outbreak are commonly used in the literature. One of these methods involves calculating the probability generating function $G_X(s)$ of the 'offspring distribution' – the distribution that describes the number of secondary cases X that are caused by a single infected individual. Then, the probability that a major outbreak does not occur is the minimal non-negative solution s of the fixed point equation $s = G_X(s)$ (the interested reader is directed to the review paper by Miller for

details [75]). An advantage of the generating function based approach is that it can also be used to calculate other quantities of interest related to the outbreak, such as the probability that extinction occurs by a given time, and the probability distribution for the total number of cases ever observed [75]. However, it can be challenging to apply this approach to models with substantial epidemiological complexity. Therefore, we do not make use of the generating function approach in this thesis, but instead focus on the second of the two analytic techniques, which we call the 'first-step analysis' method. The two approaches are fundamentally equivalent, but the first-step analysis method has the advantage that it is more intuitive to use when considering compartmental outbreak models, and is straightforward to apply to more complicated models. In this section, we will demonstrate the use of the first-step analysis technique by applying it to the SIR model, and showing that the results match those obtained through repeated stochastic simulations. During Chapters 3 and 4 of this thesis, we will then show how the principles described here can be adapted and extended to estimate the major outbreak probability for transmission models that include greater epidemiological complexity, using SARS-CoV-2 as a case study.

As the name suggests, the first-step analysis approach for estimating outbreak probabilities works by considering the 'first event' to occur in the model after the primary case arrives in the population. In a stochastic formulation of the SIR model, at any point in time t there are two possibilities for the next event that occurs:

1. A susceptible individual becomes infected, resulting in the change of state $S \to S - 1$, $I \to I + 1$.
2. An infected individual is removed from the system, resulting in the change of state $I \to I - 1$, $R \to R + 1$.

In the frequency-dependent SIR formulation, the rates at which each of these events occur are assumed to be exponentially distributed, with the following rate parameters:

$$\text{Rate at which new infections occur} = \beta I(t)S(t)/N,$$
$$\text{Rate at which infected hosts are removed} = \mu I(t).$$

Therefore, at time t the probability that the next event to occur in the model is an infection is

$$\mathbb{P}(\text{next event is an infection}) = \frac{\beta I(t)S(t)/N}{\beta I(t)S(t)/N + \mu I(t)}. \tag{2.1}$$

Similarly,

$$\mathbb{P}(\text{next event is a removal}) = \frac{\mu I(t)}{\beta I(t)S(t)/N + \mu I(t)}. \tag{2.2}$$

At the beginning of an outbreak, the number of infected individuals is small, so that $S(t) \approx N$. Therefore, when the initial infected host arrives in the population we have that approximately

$$\mathbb{P}(\text{next event is an infection}) \approx \frac{\beta}{\beta+\mu}, \quad \mathbb{P}(\text{next event is a removal}) \approx \frac{\mu}{\beta+\mu}. \tag{2.3}$$

We denote the probability that a major outbreak does not occur starting from i infected individuals by q_i. Then, the probability that a major outbreak does occur starting from i infected individuals is given by $p_i = 1 - q_i$. Beginning from a single infected individual, we compute the probability that a major outbreak does not occur (q_1) by invoking the law of total probability to condition on the next event to occur in the model, as follows:

$$\begin{aligned}
&\mathbb{P}(\text{no major outbreak starting from one infected host}) \\
&\quad = \mathbb{P}(\text{no major outbreak} \,|\, \text{first event is an infection})\, \mathbb{P}(\text{first event is an infection}) \\
&\qquad + \mathbb{P}(\text{no major outbreak} \,|\, \text{first event is a removal})\, \mathbb{P}(\text{first event is a removal}).
\end{aligned}$$

That is,

$$q_1 = \frac{\beta}{\beta+\mu}\, q_2 + \frac{\mu}{\beta+\mu}\, q_0. \tag{2.4}$$

Starting from two infected individuals, a major outbreak does not occur if and only if neither of those individuals initiates a chain of sustained transmission.

To simplify Equation (2.4), we note that a major outbreak certainly will not occur if no infected hosts are present, and therefore $q_0 = 1$. Additionally, we make the standard branching process assumption that separate infection lineages can be treated as independent, so that $q_2 = q_1^2$. This assumption naturally does not account for shared contact networks within the population – for example, if the two infected hosts are in the same household, their infection lineages are unlikely to be independent – but is nevertheless a common simplification in mathematical epidemiology, particularly when the population can be assumed reasonably well-mixed [40, 42]. With these simplifications, Equation (2.4) becomes

$$q_1 = \frac{\beta}{\beta+\mu}\, q_1^2 + \frac{\mu}{\beta+\mu}. \tag{2.5}$$

This quadratic equation in q_1 can be solved straightforwardly to obtain the two solutions $q_1 = \mu/\beta$ and $q_1 = 1$. As is standard when calculating extinction probabilities using branching process models, we take the minimal non-negative solution[1]; that is, $q_1 = \mu/\beta$ whenever

[1] This formulation is exactly equivalent to the approach used to find hitting probabilities in a Markov chain. If we consider the temporal evolution of i to be a Markov process on the state space $M = Z_{\geqslant 0}$, then q_i is the hitting probability of the state 0 starting from the state i. The vector of hitting probabilities $q^* = \{q_i^* | i \in M\}$ is therefore the minimal non-negative (real) solution to the following system of equations:

$$q_0 = 1, \quad q_i = \sum_{j \in M} p_{i,j}\, q_j \quad \text{for} \quad i \neq 0,$$

where $p_{i,j}$ is the transition probability from state i to state j [76]. Here, minimality means that if $\hat{q} = \{\hat{q}_i | i \in M\}$ is another non-negative real solution, then $q_i^* \leqslant \hat{q}_i$ for all $i \in M$.

$\mu/\beta < 1$, and $q_1 = 1$ otherwise.

Recalling that in the SIR model the basic reproduction number is given by $R_0 = \beta/\mu$, the probability of a major outbreak is therefore given by

$$p_1 = 1 - q_1 = \begin{cases} 1 - 1/R_0 & \text{when} \quad R_0 > 1, \\ 0 & \text{otherwise.} \end{cases} \tag{2.6}$$

This analytic solution is plotted for various values of R_0 in Fig 2.3 (red line), with the outbreak probabilities obtained through repeated stochastic simulations of the SIR model (run using the Gillespie SSA) overlaid for comparison (blue circles). We can see that the analytic results match well with the simulated results.

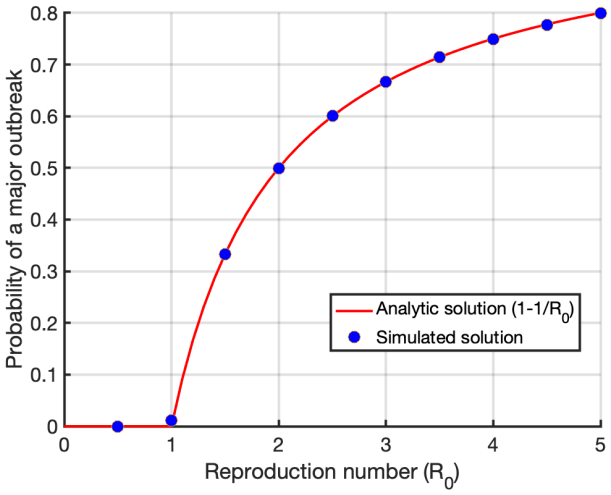


Figure 2.3: The probability of a major outbreak occurring beginning from a single infected individual, as the basic reproduction number (R_0) varies. Red line indicates the analytic outbreak probabilities derived using the first-step analysis method (Equation (2.6)). Blue circles indicate the outbreak probabilities obtained through repeated stochastic simulations, as outlined in Section 2.2.

As described earlier in this section, one of the key strengths of the first-step analysis method for estimating outbreak risks is that it is straightforward to extend this methodology to apply to more complex compartmental models. This approach has therefore been used in several previous studies that aim to calculate the probability of a major outbreak for more realistic models of pathogen transmission [40–42, 64, 67, 77, 78]. For example, Thompson *et al.* [42] demonstrated how a first-step analysis could be used to calculate the outbreak probability in a host-vector model of pathogen transmission, using Zika virus (a mosquito-borne pathogen typically found in tropical and subtropical regions) as a case study. Noting that the transmission rate of a vector-borne pathogen varies throughout the year as a result of seasonal fluctuations in the vector population, Kaye *et al.* [77] subsequently extended

this approach to investigate how the probability of a major outbreak of Zika virus disease depends on the time of year that the pathogen enters the population. As well as calculating the instantaneous outbreak risk based on the vector population size at the time the pathogen is introduced, they demonstrated how future changes in the vector population occurring after the time of introduction could be incorporated into the risk calculation.

The first-step analysis approach has also been used to explore how the probability that a pathogen triggers a major outbreak is affected by existing immunity in the population (which may arise from prior exposure to a related pathogen) [41, 67, 78]. For example, Hartfield and Alizon [78] considered a scenario in which a pathogen that is circulating in a population at low prevalence mutates, giving rise to a new strain that is more strongly adapted for infection (i.e. has a larger reproduction number). Imposing the condition that prior infection with the original pathogen confers immunity to the newly emerged strain, they used a time-dependent version of the first-step analysis technique to calculate the probability of the new strain triggering a major outbreak.

Further applications of the first-step analysis technique have included extensions to models that account for different types of infected host. For example, Southall *et al.* [40] considered an age-structured model in which the population was divided into children and adults, accounting for heterogeneties in the rate at which transmission occurs within and between these groups in their outbreak risk calculation. Similarly, Thompson [64] considered a scenario in which infected hosts are split into two categories based on how rapidly they report symptoms (with the assumption that they are subsequently isolated and generate no further infections). The author used a first-step analysis approach to calculate the probability that a major outbreak is observed given that the initial infected host is a 'fast reporter' (i.e. is likely to be isolated rapidly after developing symptoms) or a 'slow reporter' (i.e. is likely not to be isolated for some time after symptoms develop).

Of particular recent note, the first-step analysis technique has also been used to calculate outbreak probabilities during the COVID-19 pandemic, as outlined below.

2.4 Case study: the COVID-19 pandemic

In December 2019, multiple cases of an unidentified pneumonia-like illness were observed in the city of Wuhan, China [79]. The causative agent of these infections was subsequently identified as a novel coronavirus, which is believed to have originated in bats and been introduced to the human population as a result of zoonotic spillover [59, 60]. This novel virus was later named SARS-CoV-2, whilst the associated disease, symptoms of which typically include coughing, shortness of breath, muscle pain, and fever, became known as COVID-19

[80]. Although many cases of COVID-19 proved mild, in other cases the infection resulted in severe disease or death, particularly amongst the elderly and those with existing underlying health conditions [81, 82]. Therefore, the novel pathogen was considered to pose a serious risk to global public health. At the beginning of the outbreak, cases were limited to China, and it was hoped that the strict containment measures enforced by the Chinese government, coupled with the vigilance of other countries in identifying and isolating suspected imported cases, would be sufficient to prevent widespread global transmission [83]. Unfortunately, this proved not to be the case. Local transmission of SARS-CoV-2 was soon observed in other countries, and the numbers of observed cases and affected countries began to increase rapidly [13, 50]. The outbreak was declared a pandemic by the World Health Organisation (WHO) in March 2020, and went on to become one of the biggest public health crises in recent history, resulting in over 750 million cases and almost seven million deaths worldwide [72, 84]. As the pathogen spread globally, the governments of many countries introduced a variety of stringent control interventions that aimed to reduce transmission and relieve the pressure on healthcare organisations. At various stages throughout the pandemic, these measures included social distancing, the mandatory wearing of face coverings in public places, restrictions on the size of events and gatherings, restricted inter- and intranational travel, school closures, and full 'stay at home' lockdowns [13, 14]. Many of these interventions were of long duration, and only began to be relaxed following the widespread uptake of COVID-19 vaccines that were developed at record speed in response to the global crisis.

Throughout the pandemic, mathematical modelling techniques proved to be vital tools for predicting the course of the outbreak and guiding the introduction and removal of the control interventions described above [85, 86]. Of particular relevance to this thesis, branching process models and the first-step analysis technique were used during the pandemic in a variety of contexts, to understand the risk of major outbreaks occurring in different populations and under different control scenarios [41, 64, 67, 87]. The COVID-19 pandemic therefore provides an excellent example of a real-world scenario in which outbreak risk estimates were useful to inform control interventions.

For example, early in the COVID-19 pandemic, approaches for control in regions of low prevalence were focused on the rapid isolation of identified infected hosts, and the tracing and isolation of individuals they had come into contact with [50]. To investigate whether this approach was likely to be effective, in February 2020 Thompson [64] used a branching process model to estimate the probability that an imported case would lead to a major outbreak of COVID-19, using observed data describing the expected time between symptom onset and hospitalisation as a proxy for the duration of time in which that individual could infect others. Then, the author used the model to investigate how the outbreak risk would be affected if this time could be reduced as a result of intensified control measures. They

found that if the time from symptom onset to isolation could be halved, the likelihood of an outbreak would be substantially reduced, even if there were multiple initial infected hosts. Around the same time, Hellewell *et al.* [87] used simulations of a branching process model of SARS-CoV-2 transmission to predict whether or not new outbreaks would fade out using different contact tracing strategies, in which infected contacts of a previous case were identified with a given probability and subsequently isolated. They demonstrated that the probability that an outbreak was controlled and faded out as a result of contact tracing depended strongly on the reproductive number of the pathogen, the number of initial cases, and the delay from symptom onset to isolation. For a range of plausible values for each of these factors, their results showed that contact tracing alone was unlikely to be sufficient to contain an outbreak.

As the pandemic became established, an ongoing challenge for policy makers was the emergence of new variants of SARS-CoV-2 that were more transmissible than the wild type virus, caused more severe disease, or were able to evade existing immunity in the population [88]. A notable example was the 'Omicron' variant (B.1.1.529) first identified in Botswana in November 2021, which proved to be not only much more transmissible than the wild type virus and previously observed variants, but also able to evade vaccine-induced immunity [89, 90]. Unsurprisingly, Omicron subsequently became the dominant circulating variant of SARS-CoV-2, although it fortunately proved less virulent than previous variants [89, 91]. Whenever a new variant of concern was identified, a key question for policy makers was whether that strain would spread widely in the population, particularly if it was known to cause more severe disease (as was the case with the 'Delta' variant, B.1.617.2, for example [91]). Estimating the probability of an outbreak therefore became relevant once again in the context of predicting the emergence of novel variants. With this in mind, Thompson *et al.* [41] used the first step analysis technique to explore how the risk of a major outbreak of a novel COVID-19 variant would be affected by existing cross-protective immunity in the population arising due to a previously circulating antigenically related strain. They found that the level of cross-protective immunity conferred by prior infection with the related strain substantially affected the likelihood of a major outbreak of the new variant. If prior infection with the related strain confers total protection from the novel variant, the ability of the new variant to take over is compromised, and in order to invade the population it must be more transmissible than the previously circulated strain. However, if the cross-protective immunity conferred by the previous strain is only partial, a new variant may invade even if it is less transmissible than the previous strain. The results of this study emphasised that a rapid assessment of the ability of a novel strain to evade existing immunity is critical in order to assess the likelihood of widespread transmission of novel SARS-CoV-2 variants.

Following the authorisation of the first COVID-19 vaccines for human use in December 2020

[92], large scale vaccination programmes were rolled out in many countries worldwide. As more people were vaccinated and thus protected from the risk of severe disease, with priority given to those most at risk, governments were faced with the challenge of deciding when and how to relax ongoing control interventions whilst avoiding outbreaks amongst the still unvaccinated population. This question was considered by Sachak-Patwa *et al.* [67], who constructed a stochastic branching process model for SARS-CoV-2 transmission and used it to investigate whether vaccinations that lead to a reduction in population susceptibility could eliminate the risk of a major outbreak when control interventions were removed. Using a time-dependent version of the first step analysis technique (similar to that used by Hartfield and Alizon [78], described in the previous section), they assessed the risk of a COVID-19 outbreak occurring when an imported case arrives in a population, accounting for the ongoing changing susceptibility of the population occurring as a result of vaccine uptake. They demonstrated that although vaccination does substantially reduce the risk of sustained transmission occurring, even a high vaccine uptake is unlikely to be enough to completely eliminate the risk of an outbreak when restrictions are removed, given that vaccination does not provide perfect protection against transmission. In addition, new variants may emerge that are able to evade existing immunity. However, their model provides a framework that can be used to identify the dates that restrictions can be lifted based on a specified maximum acceptable outbreak risk.

2.5 The contribution of Part I to the literature

As we have seen, estimating the probability of a major outbreak was highly relevant in the context of the COVID-19 pandemic, and has consequently been the focus of several studies. However, although the studies described above addressed important questions regarding the risk of an outbreak under different control interventions and in scenarios where some level of immunity had been established in the population, they did not specifically account for other important factors that substantially influenced the transmission dynamics of SARS-CoV-2.

For example, early in the COVID-19 pandemic it was recognised that both presymptomatic and asymptomatic transmission of SARS-CoV-2 could occur [93–105]. 'Presymptomatic transmission' refers to transmission by infected individuals who have not yet developed symptoms, but will go on to do so later in their infection, whilst 'asymptomatic transmission' refers to transmission by infected individuals who never go on to develop symptoms. Throughout this thesis, we will use 'nonsymptomatic transmission' as an umbrella term to refer to both pre- and asymptomatic transmission (that is, any transmission that occurs in the absence of clinical symptoms). Nonsymptomatic transmission proved to be a substantial hindrance for containment efforts throughout the pandemic, since nonsymptomatic hosts could transmit the pathogen widely before realising they were infected and self-isolating. For

policy makers, an important question was therefore whether to allocate surveillance resources to individuals who were not showing symptoms, in the hope of identifiying nonsymptomatic infected individuals early, or whether to focus on confirming infection in symptomatic individuals and isolating them. A quantitative assessment of the probability that a major outbreak occurs under different surveillance and control strategies targeting symptomatic and nonsymptomatic hosts would therefore be a valuable tool for determining the optimal division of surveillance resources between these groups.

In addition to nonsymptomatic transmission, another key factor that influenced the transmission of SARS-CoV-2 and the effectiveness of control interventions was the age structure of the population. As the pandemic took off, it became clear that susceptibility to SARS-CoV-2 infection and the severity of the symptoms experienced varied substantially with respect to age [106–115]. Combined with the substantial age-related heterogeneities that are observed in typical daily contact patterns, this suggested that the likelihood of observing an outbreak would be strongly influenced by the ages of the earliest cases in a population. Once again, a clear opportunity to extend the existing methodology for estimating the probability of a major outbreak arose. If the risk estimate could be adapted to account for the age of the index case, it could be used to assess the effectiveness of control interventions that targeted specific age groups within the population, such as school closures.

In Part I of this thesis, we present two additional studies that we conducted during the pandemic, in which we further developed the first-step analysis methodology for estimating COVID-19 outbreak risks to account for these crucial factors. We begin in Chapter 3 by considering the question of presymptomatic and asymptomatic (together referred to as nonsymptomatic) transmission. Using SARS-CoV-2 as a case study, we construct a branching process model of pathogen transmission that includes nonsymptomatic transmission explicitly, and demonstrate how the first-step analysis methodology described in Section 2.3 can be extended to assess the risk of a major outbreak when nonsymptomatic transmission can occur. Then, we explore the effects of interventions targeting nonsymptomatic or symptomatic hosts when surveillance resources are limited. In particular, we consider whether the greatest reductions in outbreak risks are achieved by increasing surveillance and control targeting nonsymptomatic or symptomatic cases, or a combination of both. In Chapter 4 of this thesis, we then further extend the branching process model developed in Chapter 3 to take into account the age-stratification of the population. In particular, we incorporate observed age-related heterogeneities in social contact patterns, susceptibility to SARS-CoV-2 infection and, if infected, the likelihood of experiencing a fully asymptomatic course of infection. We use this extended model to demonstrate how the risk of a major outbreak depends on the age of the index case, and to explore the effects of non-pharmaceutical interventions (NPIs) that target individuals of different ages, including

school closures, workplace closures, and population-wide social distancing policies. Further, we show how multiple contact-reducing NPIs may be used in combination with strategies for surveillance and isolation of infectious hosts to eliminate major outbreak risks almost entirely.

These studies represent novel extensions to the existing methodology for estimating the probability of a major outbreak. The contribution of nonsymptomatic transmission to the outbreak risk and the corresponding optimal division of surveillance resources between symptomatic and nonsymptomatic hosts has not previously been considered. Similarly, the first-step analysis methodology has never been extended to take into account the full range of age groups that we consider here, or to assess how the outbreak risk is affected by interventions that aim to reduce the number of daily contacts between individuals of different ages. As well as providing insights into SARS-CoV-2 transmission, the general modelling frameworks that we present in this thesis can be used to estimate major outbreak risks during future epidemics of a range of pathogens, accounting fully for nonsymptomatic transmission and age-related factors.

Chapter 3

Interventions targeting nonsymptomatic cases can be important to prevent major outbreaks

3.1 Introduction

When an infectious pathogen arrives in a susceptible population, a crucial factor that can influence the likelihood that a major outbreak occurs is whether infected individuals are able to transmit the pathogen when they are not showing clinical symptoms. In this thesis, we use 'nonsymptomatic transmission' as an umbrella term to refer to any transmission of this kind, which can be broken down into two sub-types: 'presymptomatic transmission', which refers to transmission by infected individuals who have not yet developed symptoms, but will go on to do so later in their infection, and 'asymptomatic transmission', which refers to transmission by infected individuals who never go on to develop symptoms.

For some infectious diseases, infected hosts do not become infectious until they have already developed symptoms. For such diseases, the rapid identification and subsequent isolation of symptomatic hosts can be very effective as a means of containment. For example, during the 2002-2004 SARS outbreak, no confirmed cases of transmission from nonsymptomatic hosts were observed [12, 15, 116, 117]. Indeed, infectiousness appeared to be low at the time of symptom development, with peak infectiousness only reached during the second week of illness [15, 117, 118]. This delay between symptom onset and infectiousness provided a crucial window of time in which infected individuals could be isolated before their infectiousness increased, meaning that onward transmission could be effectively interrupted [117]. This feature of SARS-CoV transmission was probably one of the critical factors that allowed public health measures to contain the SARS epidemic with relatively few cases compared to, for example, the COVID-19 pandemic [12, 15].

However, for many infectious pathogens transmission can occur from infected individuals who are not yet showing symptoms [12, 119]. For example, individuals infected with the

measles virus become infectious several days before the characteristic rash is observed [120, 121], and it is estimated that around 46% of new measles infections are generated by nonsymptomatic individuals [119]. Similarly, nonsymptomatic transmission appears to be a key driver of influenza outbreaks [122–126], although estimates of the overall contribution of nonsymptomatic transmission to influenza spread vary [119, 122, 126, 127]. Other examples of pathogens that can be transmitted prior to symptom onset include the viruses that cause mumps, rubella, polio, smallpox, and hepatitis B [12, 119].

In either case, nonsymptomatic transmission can substantially hinder the effectiveness of control interventions that aim to contain pathogen spread. The greater the proportion of transmission that occurs in the absence of clinical symptoms, the less likely it is that the pathogen can be contained through control measures that are triggered by symptom onset, such as isolation and contact tracing [12, 119, 120]. Indeed, according to Fraser *et al.* [12], *"the proportion of transmission that occurs before the onset of clinical symptoms or in asymptomatic infection may be equally as important for determining the ease of control of a novel outbreak as the intrinsic transmissibility, R_0"*. Accounting for the possibility of nonsymptomatic transmission in epidemic models is therefore critical for accurately determining the probability of a major outbreak and the likely effects of control interventions.

At the beginning of the COVID-19 pandemic, it was not known whether nonsymptomatic transmission of SARS-CoV-2 could occur. However, as the outbreak progressed, multiple reports were published that described likely instances of presymptomatic transmission based on analyses of contact tracing data for small clusters of infection [93–100]. Of further concern, in March 2020 Zou *et al.* [101] reported that the viral load present in asymptomatic individuals was comparable with that in symptomatic individuals, suggesting that infected hosts who never go on to develop clinical symptoms may also be able to contribute to transmission. As time passed, it became clear that this was indeed the case [102–105]. Throughout the pandemic, presymptomatic and asymptomatic transmission presented a substantial challenge to control interventions for COVID-19. In particular, control measures that relied upon isolation of symptomatic hosts and tracing of their contacts proved insufficient to contain the spread of the pathogen and prevent COVID-19 outbreaks.

In the first year of the pandemic, rapid identification of presymptomatic and asymptomatic infected hosts on a large scale was not feasible. Although PCR testing was able to identify the presence of the virus in infected samples at very low quantities, and could therefore theoretically have been used to identify pre- or asymptomatic infected hosts, early in the pandemic testing capacity was severely limited due to lack of laboratory infrastucture, and PCR testing was therefore restricted to patients admitted to hospital with suspected COVID-19 [128]. Following projects that were put in place to expand laboratory testing

capacity, PCR testing in the UK was later rolled out to a wider demographic – first to symptomatic people in high risk groups, and then to anyone showing symptoms of COVID-19 [128] – but not to nonsymptomatic individuals. However, following the development of rapid antigen tests for identifying SARS-CoV-2 infection, in September 2020 the UK government announced their plans for rolling out a large scale mass testing scheme, dubbed 'Operation Moonshot' [129]. In the initial phase of the rollout, priority for rapid antigen testing was given to individuals in high risk settings such as hospitals and care homes [129], followed by some key workers, schools and universities [130]. Finally, from 9th April 2021 everyone in England was offered access to twice weekly rapid antigen testing, regardless of whether they were showing symptoms or not [130].

The UK government's rationale for this expensive and logistically demanding rollout was that large scale asymptomatic testing would lead to the identification and self-isolation of many asymptomatic individuals who would never otherwise be found. Indeed, the government press release announcing that rapid antigen testing was being made available to all highlighted that *"Since rapid testing was introduced, over 120,000 positive cases that would not have been found otherwise have already been identified"* [131]. Additionally, it was hoped that mass testing would result in the early identification of presymptomatic individuals, who may then self-isolate more rapidly than they would have done otherwise. However, the question of whether valuable resources should be used to test nonsymptomatic people was hotly debated. Some argued that focusing on testing symptomatic people and tracing their contacts would be a better use of resources; following the announcement of the 'Operation Moonshot' plan in September 2020, the Scientific Advisory Group for Emergencies (SAGE) stated that *"Prioritising rapid testing of symptomatic people is likely to have a greater impact on identifying positive cases and reducing transmission than frequent testing of asymptomatic people in an outbreak area"* [132]. Additional concerns were raised about the potential for mass testing to cause delays in processing the tests of symptomatic individuals, and the potential number of false positives that could occur due to imperfect test specificity [129, 130].

Motivated by the question of whether large scale nonsymptomatic testing was a sensible use of surveillance resources during the COVID-19 pandemic, in this chapter we use the outbreak risk methodology described in Chapter 2 to explore the relative effects of rapidly identifying and isolating infected symptomatic and nonsymptomatic hosts on the probability of observing a major outbreak. We construct a branching process model for SARS-CoV-2 transmission that accounts for presymptomatic and asymptomatic trasnsmission explicitly, and extend the existing first-step analysis methodology to estimate the probability of a major outbreak occurring starting from a single infected individual in either group. Then, using a function that characterises the efficacy of interventions for different surveillance efforts (denoted $f(\rho, \delta)$ in the model), we explore the effects of interventions that aim to reduce this

probability. Under the assumption that detected infected hosts are isolated effectively, we consider whether it is most effective to dedicate resources to enhancing surveillance targeting symptomatic individuals, to instead focus on increasing surveillance for nonsymptomatic individuals, or to use a combination of these approaches.

We show that, when surveillance resources are limited, the maximum reduction in the outbreak probability almost always corresponds to a mixed strategy involving enhanced surveillance of both symptomatic and nonsymptomatic hosts. This remains the case even if the surveillance effort required to find nonsymptomatic infected individuals is significantly larger than the effort required to find symptomatic individuals, since nonsymptomatic surveillance not only interrupts transmission from those who are not showing symptoms, but can also prevent future symptomatic transmission occurring from individuals who are presymptomatic at the time of isolation. This highlights the benefits of not only seeking to find and isolate symptomatic hosts, but also dedicating resources to detecting nonsymptomatic cases during infectious disease epidemics.

3.2 Methods

3.2.1 Model

In this chapter, we consider a branching process model for SARS-CoV-2 transmission in which infectious individuals are classified as asymptomatic (A), presymptomatic (I_1), or symptomatic (I_2). Hosts in any of these classes may generate new infections. The parameter ξ represents the proportion of new infections that are asymptomatic, so that a new infection either involves increasing A by one (with probability ξ) or increasing I_1 by one (with probability $1-\xi$). Presymptomatic hosts (I_1) may go on to develop symptoms (transition from I_1 to I_2), or be detected and isolated (so that I_1 decreases by one). Symptomatic individuals (I_2) can be isolated (so that I_2 decreases by one), or removed due to recovery or death (so that again I_2 decreases by one). Similarly, asymptomatic hosts (A) may be detected and isolated, or recover (so that A decreases by one in either case). A schematic showing the different possible events in the model is shown in Fig 3.1A. The analogous compartmental differential equation model to the branching process model that we consider is given by

$$\begin{aligned}
\frac{\mathrm{d}A}{\mathrm{dt}} &= \xi\left(\eta\beta A + \alpha\beta I_1 + \beta I_2\right) - \frac{\varepsilon\gamma}{1-f(\rho_1,\delta)}A - \nu A,\\
\frac{\mathrm{d}I_1}{\mathrm{dt}} &= (1-\xi)\left(\eta\beta A + \alpha\beta I_1 + \beta I_2\right) - \frac{\varepsilon\gamma}{1-f(\rho_1,\delta)}I_1 - \lambda I_1,\\
\text{and}\quad \frac{\mathrm{d}I_2}{\mathrm{dt}} &= \lambda I_1 - \frac{\gamma}{1-f(\rho_2,\delta)}I_2 - \mu I_2.
\end{aligned}$$

The parameters of the model, and the form of the function $f(\rho,\delta)$ that describes how the

expected time to isolation is reduced for a given surveillance effort, are outlined below.

In our model, the parameter β and its scaled counterparts $\alpha\beta$ and $\eta\beta$ represent the rates at which symptomatic, presymptomatic, and asymptomatic hosts generate new infections, respectively. Since we are modelling the beginning of a potential major outbreak, we assume that the size of the susceptible population remains approximately constant and do not track the depletion of this population. The parameter λ governs the rate at which presymptomatic individuals develop symptoms, so that the expected duration of the presymptomatic period is $1/\lambda$ days in the absence of interventions. Similarly, without interventions, the expected durations of the symptomatic and asymptomatic infectious periods are $1/\mu$ days and $1/\nu$ days, respectively.

The baseline rate at which symptomatic individuals are detected and isolated is determined by the parameter γ. Assuming that nonsymptomatic individuals are more difficult to detect than symptomatic individuals, we take the analogous quantity for nonsymptomatic hosts to be $\varepsilon\gamma$, where the scaling factor $\varepsilon < 1$ reflects the fact that interventions targeting nonsymptomatic hosts are likely to be less effective for the same surveillance effort. We assume that the sensitivity of surveillance is identical for presymptomatic and asymptomatic individuals, and therefore use the same isolation rate for both of these groups.

The parameters ρ_1 and ρ_2 represent the surveillance intensification effort targeted at nonsymptomatic and symptomatic hosts, respectively. The function $f(\rho, \delta) = \delta\rho/(1+\rho)$ governs the proportional reduction in the expected time to isolation for a given surveillance effort, ρ (for a similar approach in which the proportion of infectious cases prevented is assumed to be a function of control effort, see Matthews *et al.* [133]). The functional form of $f(\rho, \delta)$ is chosen for three main reasons. First, it generates a reduced expected time to isolation when the surveillance effort increases. Second, since the proportional reduction in the expected time to isolation is bounded above by the parameter $\delta \in (0, 1)$, the isolation rate saturates and cannot increase indefinitely. Third, the gradient $\delta f/\delta\rho$ decreases with the surveillance effort ρ, meaning that an increase in the surveillance effort has a greater impact at low surveillance efforts compared to when this effort is already large [133]. The function $f(\rho, \delta)$ is shown in Fig 3.1B for different values of the parameter δ.

3.2.2 Reproduction number

The basic reproduction number, R_0, represents the expected number of secondary infections generated by a single infected individual, introduced at the start of their infection into a fully susceptible population, in the absence of intensified surveillance:

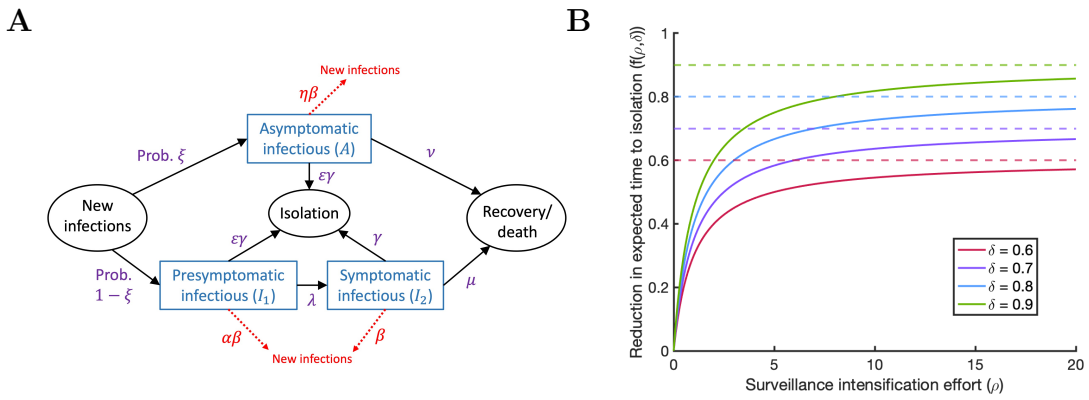


Figure 3.1: The branching process model used in our analyses. A. Schematic showing the different event types in the branching process model, and the rates at which those events occur. The parameters of the model are described in the text and in Table 3.1. **B.** The relationship between the surveillance intensification effort (ρ) and the proportional reduction in the expected time to isolation ($f(\rho, \delta)$), shown for different values of the parameter δ (solid lines). The parameter $\delta \in (0, 1)$ represents the upper bound of $f(\rho, \delta)$ (dotted lines). This general functional relationship between surveillance effort and isolation effectiveness is assumed to hold for surveillance of both nonsymptomatic and symptomatic individuals, although nonsymptomatic hosts are more challenging to detect than symptomatic hosts ($\varepsilon < 1$).

$$R_0 = \frac{\xi\eta\beta}{\nu + \varepsilon\gamma} + (1-\xi)\left[\frac{\alpha\beta}{\lambda + \varepsilon\gamma} + \frac{\lambda}{\lambda + \varepsilon\gamma}\frac{\beta}{\gamma + \mu}\right]. \tag{3.1}$$

This expression is the sum of the expected number of transmissions from a host who begins in the asymptomatic class and from a host who begins in the presymptomatic infectious class, weighted by the respective probabilities ξ and $1 - \xi$ that determine the chance that the host experiences a fully asymptomatic course of infection, or goes on to develop clinical symptoms. The expected number of transmissions from a host who begins in the presymptomatic infectious class comprises transmissions occurring during the incubation period and transmissions occurring during the symptomatic period, accounting for the possibility that the host is isolated prior to developing symptoms.

The proportion of infections arising from presymptomatic hosts in the absence of intensified surveillance is then given by

$$K_p = \frac{\dfrac{(1-\xi)\alpha}{\lambda + \varepsilon\gamma}}{\dfrac{\xi\eta}{\nu + \varepsilon\gamma} + (1-\xi)\left[\dfrac{\alpha}{\lambda + \varepsilon\gamma} + \dfrac{\lambda}{\lambda + \varepsilon\gamma}\dfrac{1}{\gamma + \mu}\right]}, \tag{3.2}$$

and the equivalent quantity for asymptomatic hosts is given by

$$K_a = \frac{\dfrac{\xi\eta}{\nu+\varepsilon\gamma}}{\dfrac{\xi\eta}{\nu+\varepsilon\gamma} + (1-\xi)\left[\dfrac{\alpha}{\lambda+\varepsilon\gamma} + \dfrac{\lambda}{\lambda+\varepsilon\gamma}\dfrac{1}{\gamma+\mu}\right]}. \tag{3.3}$$

3.2.3 Baseline values of model parameters

Since this research was motivated by the need to estimate outbreak risks outside China in the initial stages of the COVID-19 pandemic, we used a baseline set of parameter values in our analyses that was informed by studies conducted during that pandemic (Table 3.1). Where possible, these parameter values were obtained from the existing literature. However, we also performed sensitivity analyses to determine how our results varied when the parameter values were changed (see Supplementary Text A.1 and Supplementary Figs A.1-A.10). In Table 3.1, and throughout, rounded values are given to three significant figures.

The value of the parameter governing the baseline rate at which symptomatic individuals are isolated, γ, was chosen to match empirical observations which indicate that individuals who seek medical care prior to recovery or death do so around four to six days after symptom onset [142]. Specifically, we assumed that the period of time to the first medical visit could be used as a proxy for the time to isolation, and chose γ so that the expected time period to isolation conditional on isolation occurring during the symptomatic period was given by $1/(\gamma+\mu) = 4.6$ days [142] (see also Supplementary Text A.2). This is different to the time period that we refer to as the expected time to isolation for symptomatic hosts, which is $1/\gamma$ days (see Section 3.2.1).

3.2.4 Probability of a major outbreak

We denote the probability of a major outbreak not occurring, starting from i presymptomatic hosts, j symptomatic hosts, and k asymptomatic hosts, by $q_{i,j,k}$. Starting from one presymptomatic host (so that $i = 1$ and $j = k = 0$), there are four possibilities for the next event to occur in the model. That host could: i) generate a new asymptomatic infection; ii) generate a new presymptomatic infection; iii) develop symptoms; or iv) be isolated.

The probabilities of each of these events being the next to occur are obtained by considering the rates of infection, symptom development, and isolation in the branching process model. Presymptomatic hosts generate new infections at rate $\alpha\beta$, and these new infections occur in asymptomatic and presymptomatic hosts with probabilities ξ and $1-\xi$, respectively. Therefore, starting from a single presymptomatic host, new asymptomatic infections occur at rate $\xi\alpha\beta$, whilst new presymptomatic infections occur at rate $(1-\xi)\alpha\beta$. Additionally, presymptomatic hosts develop symptoms at rate λ, and are isolated at rate $\varepsilon\gamma/(1-f(\rho_1,\delta))$.

Param.	Meaning	Value	Justification
R_0	Expected number of secondary infections caused by a single infected individual (when $\rho_1 = \rho_2 = 0$)	3	Within estimated range for wild-type SARS-CoV-2 [134–137]
ξ	Proportion of infections that are asymptomatic	0.2	[138–140]
β	Rate at which symptomatic individuals generate new infections	0.336 days^{-1}	Chosen so that $R_0 = 3$
α	Relative infectiousness of presymptomatic individuals compared to symptomatic individuals	2.78	Chosen so that 48.9% of transmissions arise from presymptomatic hosts (i.e. $K_p = 0.489$) [141]
η	Relative infectiousness of asymptomatic individuals compared to symptomatic individuals	0.519	Chosen so that 10.6% of transmissions arise from asymptomatic hosts (i.e. $K_a = 0.106$) [141]
γ	Isolation rate of symptomatic individuals without intensified surveillance	0.0924 days^{-1}	Chosen so that $1/(\gamma + \mu) = 4.6$ days [142]
ε	Relative isolation rate of nonsymptomatic individuals without intensified surveillance (compared to symptomatic individuals)	0.1	Assumed; chosen within the range $\varepsilon \in (0, 1)$ (for different values, see Supplementary Fig A.6)
λ	Rate at which presymptomatic individuals develop symptoms	0.5 days^{-1}	[97, 143]
μ	Recovery rate of symptomatic individuals	1/8 days^{-1}	[144–146]
ν	Recovery rate of asymptomatic individuals	0.1 days^{-1}	Chosen so that, in the absence of interventions, the expected duration of infection is identical for all infected hosts ($1/\nu = 1/\lambda + 1/\mu$)
δ	Upper bound on the fractional reduction in the time to isolation	0.8	Assumed; chosen within the natural range $\delta \in (0, 1)$ (for different values, see Supplementary Fig A.10)
ρ_1	Surveillance intensification effort targeted at nonsymptomatic hosts	Allowed to vary in the range $[0, 20]$	N/A – range of values explored
ρ_2	Surveillance intensification effort targeted at symptomatic hosts	Allowed to vary in the range $[0, 20]$	N/A – range of values explored

Table 3.1: Parameters of the model and the values used in the baseline version of our analysis.

The overall rate at which events occur is the sum of these individual event rates:

$$\text{Total event rate} := r_{I_1} = \alpha\beta + \lambda + \frac{\varepsilon\gamma}{1 - f(\rho_1, \delta)}.$$

For each of the four possible events (i-iv, above), the probability that event occurs next in the model is equal to the individual rate at which that event occurs divided by the total event rate. This leads to the following expressions for the probability that each event occurs next in the model:

i) $\mathbb{P}(\text{A new asymptomatic infection occurs next}) = \xi\alpha\beta/r_{I_1}$,

ii) $\mathbb{P}(\text{A new presymptomatic infection occurs next}) = (1-\xi)\,\alpha\beta/r_{I_1}$,

iii) $\mathbb{P}(\text{The infected host develops symptoms next}) = \lambda/r_{I_1}$,

iv) $\mathbb{P}(\text{The infected host is isolated next}) = \varepsilon\gamma/\big[r_{I_1}(1 - f(\rho_1, \delta))\big]$.

We use these probabilities to condition on the event that occurs next in the branching process, following the introduction of a single presymptomatic infectious individual into the population. If that event is the generation of a new asymptomatic infection, which occurs with probability $\xi\alpha\beta/r_{I_1}$, the probability that a major outbreak subsequently does not occur is $q_{1,0,1}$. Applying analogous reasoning to the other possible events, we obtain

$$q_{1,0,0} = \frac{\xi\alpha\beta}{r_{I_1}}\, q_{1,0,1} + \frac{(1-\xi)\alpha\beta}{r_{I_1}}\, q_{2,0,0} + \frac{\lambda}{r_{I_1}}\, q_{0,1,0} + \frac{\varepsilon\gamma}{r_{I_1}[1 - f(\rho_1, \delta)]}\, q_{0,0,0}. \tag{3.4}$$

If there are no infectious hosts present in the population (i.e. $i = j = k = 0$), then a major outbreak will not occur, and so $q_{0,0,0} = 1$. Assuming that transmission chains arising from two infectious individuals are independent gives $q_{1,0,1} = q_{1,0,0}\, q_{0,0,1}$ and $q_{2,0,0} = q_{1,0,0}^2$. Hence,

$$q_{1,0,0} = \frac{\xi\alpha\beta}{r_{I_1}}\, q_{1,0,0}\, q_{0,0,1} + \frac{(1-\xi)\alpha\beta}{r_{I_1}}\, q_{1,0,0}^2 + \frac{\lambda}{r_{I_1}}\, q_{0,1,0} + \frac{\varepsilon\gamma}{r_{I_1}[1 - f(\rho_1, \delta)]}. \tag{3.5}$$

Similarly, considering the probability of a major outbreak failing to occur starting from a single symptomatic host gives

$$q_{0,1,0} = \frac{\xi\beta}{r_{I_2}}\, q_{0,1,1} + \frac{(1-\xi)\beta}{r_{I_2}}\, q_{1,1,0} + \frac{1}{r_{I_2}}\left[\mu + \frac{\gamma}{1 - f(\rho_2, \delta)}\right] q_{0,0,0}, \tag{3.6}$$

where

$$r_{I_2} := \beta + \mu + \frac{\gamma}{1 - f(\rho_2, \delta)}.$$

As before, noting that $q_{0,0,0} = 1$ and assuming that different infection lineages are independent leads to

$$q_{0,1,0} = \frac{\xi\beta}{r_{I_2}}\, q_{0,1,0}\, q_{0,0,1} + \frac{(1-\xi)\beta}{r_{I_2}}\, q_{1,0,0}\, q_{0,1,0} + \frac{1}{r_{I_2}}\left[\mu + \frac{\gamma}{1-f(\rho_2,\delta)}\right]. \tag{3.7}$$

Finally, considering the probability of a major outbreak failing to occur starting from a single asymptomatic host gives

$$q_{0,0,1} = \frac{\xi\eta\beta}{r_A}\, q_{0,0,1}^2 + \frac{(1-\xi)\eta\beta}{r_A}\, q_{1,0,0}\, q_{0,0,1} + \frac{1}{r_A}\left[\nu + \frac{\varepsilon\gamma}{1-f(\rho_1,\delta)}\right], \tag{3.8}$$

where

$$r_A := \eta\beta + \nu + \frac{\varepsilon\gamma}{1-f(\rho_1,\delta)}.$$

Equations (3.5), (3.7), and (3.8) may be combined to give a single quartic equation for $q_{0,0,1}$, yielding four sets of solutions for $q_{1,0,0}$, $q_{0,1,0}$, and $q_{0,0,1}$ (see Supplementary Text A.3). It is straightforward to verify that $q_{1,0,0} = q_{0,1,0} = q_{0,0,1} = 1$ is always a solution, and further solutions can be found numerically. The appropriate solution to take is the minimal non-negative real solution $q_{1,0,0} = q^*_{1,0,0}$, $q_{0,1,0} = q^*_{0,1,0}$, $q_{0,0,1} = q^*_{0,0,1}$ (see Supplementary Text A.3). Then, the probability of a major outbreak occurring beginning from a single presymptomatic host is given by

$$p_{1,0,0} = 1 - q^*_{1,0,0},$$

with equivalent expressions holding for $p_{0,1,0}$ and $p_{0,0,1}$ (the probability of a major outbreak occurring beginning from a single symptomatic host or a single asymptomatic host, respectively).

Throughout, we consider the probability p of a major outbreak starting from a single nonsymptomatic host entering the population, accounting for the possibility that the nonsymptomatic host is either presymptomatic or asymptomatic:

$$p = (1-\xi)\, p_{1,0,0} + \xi\, p_{0,0,1}.$$

By varying the values of the parameters ρ_1 and ρ_2 used to calculate p, we construct a two-dimensional probability landscape showing how the probability of a major outbreak (p) varies as the surveillance efforts targeted at nonsymptomatic hosts (ρ_1) and at symptomatic hosts (ρ_2) are increased.

3.2.5 Computational approach for optimising surveillance enhancement

In Section 3.3.2, we explore how resources for enhanced surveillance can be optimally divided between symptomatic and nonsymptomatic hosts. Specifically, we consider the effects of increasing the surveillance effort targeted at nonsymptomatic hosts (ρ_1) and at symptomatic hosts (ρ_2) by varying degrees, and investigate how ρ_1 and ρ_2 can be optimally selected under

given constraints to achieve a specified objective related to the outbreak probability. In this subsection, we outline the computational approaches that we used to identify these optimal strategies.

Initially, we computed the steepest descent contours on the outbreak probability landscape using a gradient maximisation approach, in which at each point the contour direction was determined by minimising the major outbreak probability over a fixed search radius (Figs 3.2A, 3.2B). Starting from $\rho_1 = \rho_2 = 0$, at each step we considered increasing ρ_1 by an amount $\Delta\rho_1$ and increasing ρ_2 by an amount $\Delta\rho_2$ subject to the constraint $(\Delta\rho_1)^2 + (\Delta\rho_2)^2 = r^2$, where r is a small pre-specified constant. To achieve this, we scanned over the circular arc $\Delta\rho_1 = r\cos\theta$, $\Delta\rho_2 = r\sin\theta$, for $\theta \in [0, \pi/2]$ (Fig 3.2A). In practice, this range was discretised into 33 search directions evenly spaced between 0 and $\pi/2$. We then selected the pair of $\Delta\rho_1$, $\Delta\rho_2$ values for which the major outbreak probability evaluated at $\rho_1 + \Delta\rho_1$, $\rho_2 + \Delta\rho_2$ was minimised. The process was then repeated beginning from $\rho_{1_{\text{new}}} = \rho_1 + \Delta\rho_1$, $\rho_{2_{\text{new}}} = \rho_2 + \Delta\rho_2$ (Fig 3.2B). This approach was used to generate the contours shown in Fig 3.4A.

We then considered a scenario in which, at any time, it is only possible to direct resources towards enhancing surveillance of either nonsymptomatic individuals or symptomatic individuals, and computed a critical contour that divides the region in which increasing ρ_1 has a greater effect on the major outbreak probability from the region in which increasing ρ_2 has a greater effect on the major outbreak probability. This process was performed in an analogous way to that described above, with the additional restriction that we only considered the search directions $\theta = 0$ and $\theta = \pi/2$ (i.e. intensifying only surveillance of nonsymptomatic or symptomatic hosts; see Fig 3.2C). This approach was used to generate the contour shown in Fig 3.4B.

Finally, we computed the contours of steepest descent under the constraint that the total change in surveillance effort ($\Delta\rho_1 + \Delta\rho_2 = S$) is held constant at each step, rather than fixing the search radius $(\Delta\rho_1)^2 + (\Delta\rho_2)^2 = r^2$ as shown in Figs 3.2A, 3.2B. To do so, instead of scanning over a circular arc, at each step we scanned along the line $\Delta\rho_1 = s$, $\Delta\rho_2 = S - s$, where s varies in the range $[0, S]$ (Fig 3.2D). Otherwise, the process was completely analogous to that described above. This approach was used to generate the white contour shown in Figs 3.4C and 3.4D.

3.2.6 Code availability

All computing code used to implement the above methods was written in MATLAB version R2019a, and is available at https://github.com/francescalovellread/nonsymptomatic-transmission.

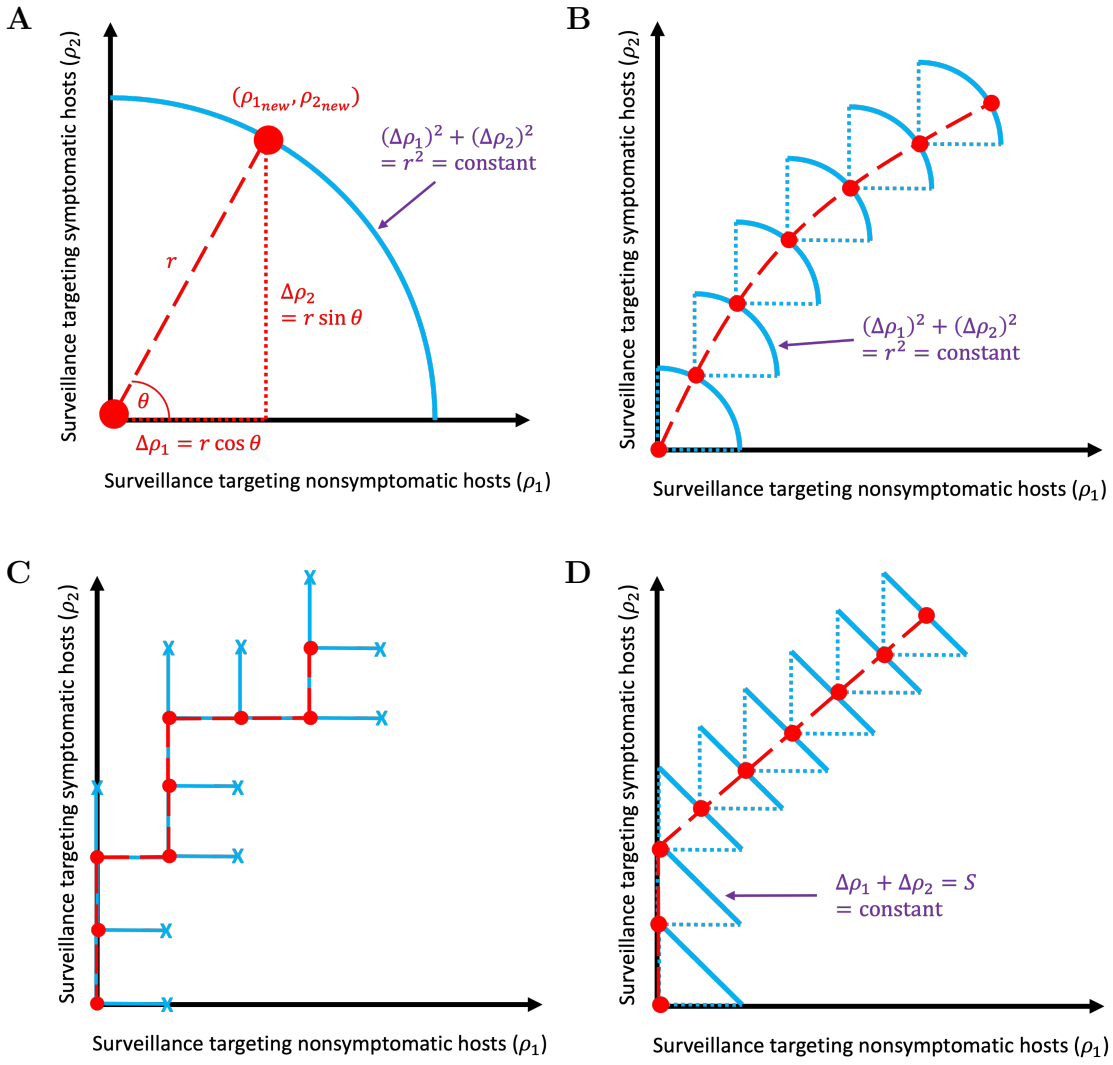


Figure 3.2: Computation of the steepest descent contours used in Section 3.3.2. A. To compute the steepest descent contours shown in Fig 3.4A, we increment ρ_1 and ρ_2 by scanning over a constant search radius $(\Delta\rho_1)^2 + (\Delta\rho_2)^2 = r^2$ (blue arc), and moving to the point $(\rho_{1_{\text{new}}}, \rho_{2_{\text{new}}})$ along that arc at which the major outbreak probability is minimised. **B.** The process shown in A is repeated to generate the complete contour (red dashed line). **C.** The analogous figure to B, in which the search direction is limited to directly to the right ($\theta = 0$) or directly upwards ($\theta = \pi/2$). This procedure is used to generate the contour in Fig 3.4B. **D.** The analogous figure to B, in which the total change in surveillance effort ($\Delta\rho_1 + \Delta\rho_2 = S$) is held constant at each step, rather than the search radius. This procedure is used to generate the contours in Figs 3.4C and 3.4D.

3.3 Results

3.3.1 Probability of a major outbreak

First, we considered the effect of varying R_0 and the duration of the presymptomatic and asymptomatic periods on the probability of a major outbreak when a nonsymptomatic host enters a new population (Fig 3.3). We examined presymptomatic periods of length $1/\lambda = 1$ day, $1/\lambda = 2$ days, and $1/\lambda = 4$ days; in each case, the duration of the asymptomatic period ($1/\nu$ days) was adjusted so that the relative proportion of infections arising from asymptomatic hosts compared to presymptomatic hosts remained fixed ($K_a/K_p = 0.218$, as in the baseline case). If nonsymptomatic infections are not accounted for, the infectious period follows an exponential distribution and the probability of a major outbreak is given by $p = 1 - 1/R_0$ (red dash-dotted line in Fig 3.3A). Compared to that case, including nonsymptomatic infection in the model led to an increased risk of a major outbreak in the absence of surveillance intensification (Fig 3.3A).

We then considered the dependence of the probability of a major outbreak on the intensity of surveillance targeting nonsymptomatic and symptomatic hosts (Figs 3.3B, 3.3C, 3.3D). The maximum value of the surveillance intensification effort that we considered (given by ρ_1 or ρ_2 values of 20) corresponded to a 76% reduction in the expected time to isolation (see blue line in Fig 3.1B), i.e. a 76% reduction in $1/\epsilon\gamma$ or $1/\gamma$.

The length of the presymptomatic and asymptomatic periods significantly affected the dependence of the probability of a major outbreak on the level of surveillance targeted at nonsymptomatic and symptomatic hosts. In Fig 3.3B, in which the duration of the presymptomatic period was one day, increasing surveillance targeted at nonsymptomatic hosts (ρ_1) had a limited effect on the probability of a major outbreak, while increasing surveillance targeted at symptomatic hosts (ρ_2) had a more substantial effect. For example, increasing the surveillance effort targeted at nonsymptomatic hosts to $\rho_1 = 5$ (a 67% reduction in the time to isolation) only reduced the probability of a major outbreak from 0.730 to 0.716, whereas the equivalent effort targeted at symptomatic hosts ($\rho_2 = 5$) reduced the probability to 0.630. As shown in Figs 3.3C and 3.3D, however, when the presymptomatic and asymptomatic periods were longer, the benefit of directing surveillance resources towards detecting nonsymptomatic individuals increased. This was because longer presymptomatic and asymptomatic periods increased the proportion of infections generated by nonsymptomatic individuals ($K_p + K_a$, see Equations (3.2) and (3.3)); a presymptomatic period of one day, two days, and four days corresponded to values of $K_p + K_a$ equal to 0.424, 0.595, and 0.746, respectively.

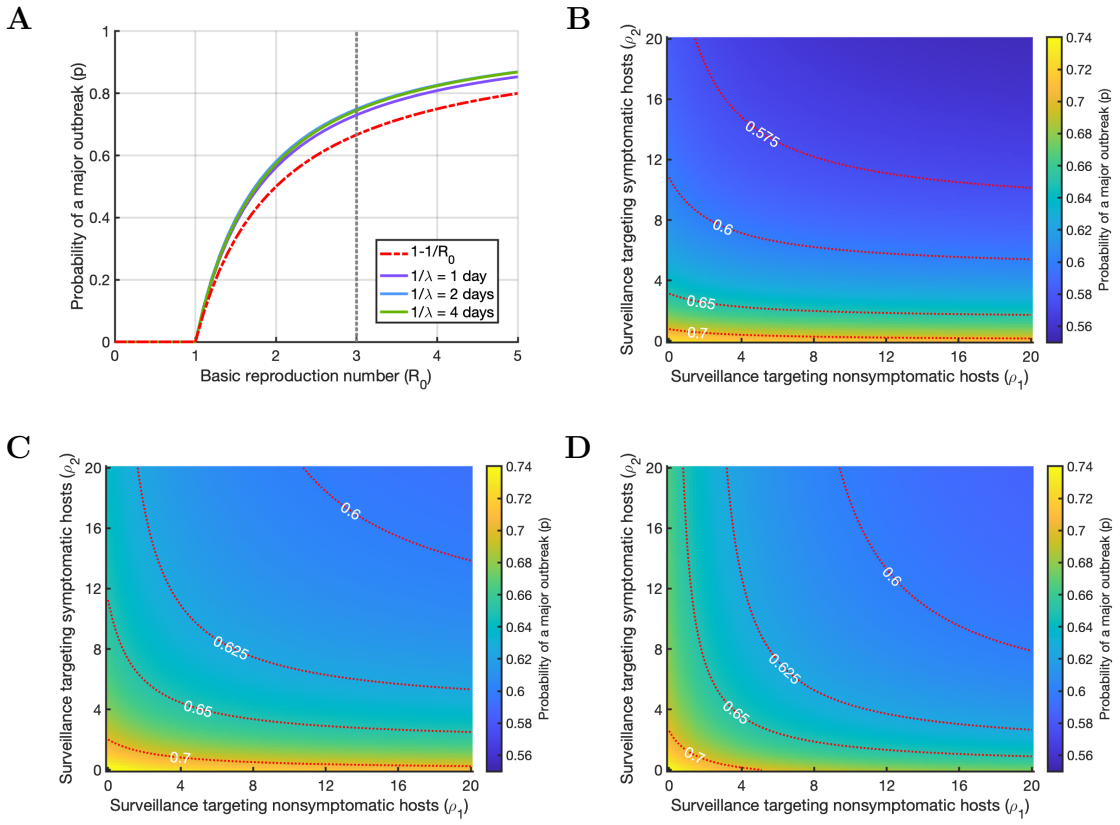


Figure 3.3: The effect of the duration of the presymptomatic and asymptomatic periods on the probability of a major outbreak (p), starting from a single nonsymptomatic host. **A.** The probability of a major outbreak as a function of the basic reproduction number, R_0, for presymptomatic periods of lengths $1/\lambda = 1$ day (purple), $1/\lambda = 2$ days (blue), and $1/\lambda = 4$ days (green) in the absence of enhanced surveillance ($\rho_1 = \rho_2 = 0$). In each case, the duration of the asymptomatic period ($1/\nu$) is adjusted so that the relative proportion of infections arising from asymptomatic hosts compared to presymptomatic hosts remains constant ($K_a/K_p = 0.218$, as in the baseline case). The red dash-dotted line indicates the probability of a major outbreak in the absence of nonsymptomatic transmission. The vertical grey dotted line indicates $R_0 = 3$, the baseline value used throughout. **B.** The probability of a major outbreak as a function of the surveillance intensification efforts ρ_1 and ρ_2, for $1/\lambda = 1$ day. Red dotted lines indicate contours of constant major outbreak probability (i.e. lines on which the probability of a major outbreak takes the values shown). **C.** The analogous figure to B but with $1/\lambda = 2$ days. **D.** The analogous figure to B but with $1/\lambda = 4$ days. In each of panels B,C,D, the value of β is selected to fix $R_0 = 3$. All other parameter values are held fixed at the values in Table 3.1 (except where stated).

3.3.2 Optimising surveillance enhancement

We next considered in more detail the impact of surveillance targeted at nonsymptomatic hosts (ρ_1) relative to the impact of surveillance targeted at symptomatic hosts (ρ_2). For our baseline parameter values, we considered the probability of a major outbreak starting from a single imported nonsymptomatic individual for a range of values of ρ_1 and ρ_2. We calculated the steepest descent contours (white lines in Fig 3.4A) numerically using a gradient maximisation approach, in which at each point the contour direction was determined by

minimising the major outbreak probability over a fixed search radius (see Section 3.2.5 and Fig 3.2). These contours indicate how ρ_1 and ρ_2 should be altered to maximise the reduction in the probability of a major outbreak. In this case, enhancing surveillance targeting both symptomatic and nonsymptomatic hosts is always optimal (the steepest descent contours are neither horizontal nor vertical).

We then considered a scenario in which, at any time, it is only possible to direct resources towards enhancing surveillance of either nonsymptomatic individuals or symptomatic individuals (e.g. antigen testing of nonsymptomatic contacts of known infectious individuals, or screening for symptomatic individuals at public events). In Fig 3.4B, the blue region represents values of ρ_1 and ρ_2 for which enhancing surveillance targeting symptomatic hosts (i.e. increasing ρ_2) leads to a larger reduction in the major outbreak probability than enhancing surveillance targeting nonsymptomatic hosts (i.e. increasing ρ_1). In contrast, in the green region, enhancing surveillance of nonsymptomatic individuals is more effective than enhancing surveillance of symptomatic individuals. The white line represents the steepest descent contour starting from $\rho_1 = \rho_2 = 0$, under the constraint that surveillance can only be enhanced for symptomatic or nonsymptomatic hosts at any time.

Practical deployment of surveillance is often subject to logistical constraints, and policymakers may wish to design surveillance strategies to achieve a specific objective – for example, to maximise the effectiveness of limited resources or to minimise the cost of achieving a desired outcome. We therefore also considered the following two examples of such objectives.

Objective 1: Minimise the probability of a major outbreak for a fixed total surveillance effort.
First, we considered the question: given a fixed maximum surveillance effort ($\rho_1 + \rho_2 = S$), how should surveillance be targeted at nonsymptomatic and symptomatic hosts? This involves setting the values of ρ_1 and ρ_2 to minimise the major outbreak probability. The optimal strategies in this case are shown in Fig 3.4C. The red dotted lines represent contours along which the total surveillance effort $\rho_1 + \rho_2$ is held constant (i.e. different values of S). On each contour, the red circle indicates the point at which the major outbreak probability is minimised. If surveillance resources are increased (i.e. S increases), a further question is how surveillance should then be increased. In Fig 3.4C, the white line represents the contour of steepest descent, under the constraint that the total change in surveillance effort ($\rho_1 + \rho_2$) is held constant at each step (rather than a constant search radius, as in Fig 3.4A – for more details, see Section 3.2.5 and Fig 3.2). This contour coincides exactly with that shown in Fig 3.4B.

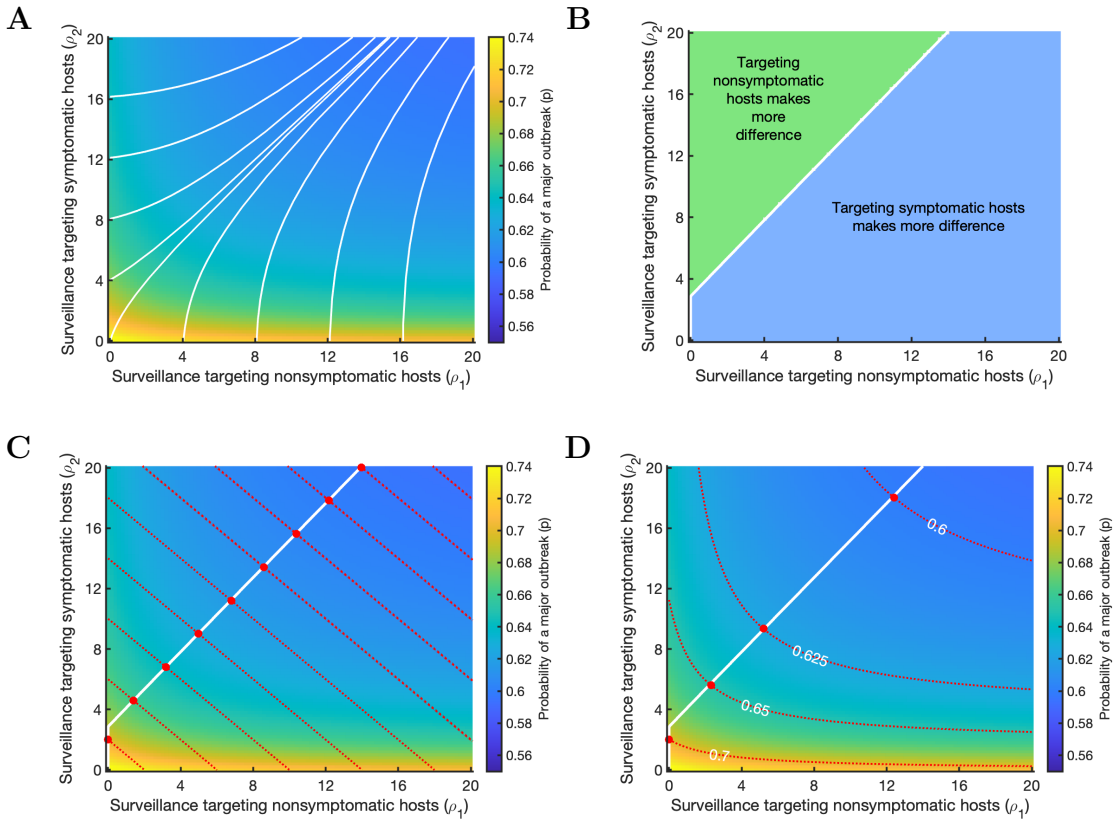


Figure 3.4: Optimal surveillance strategies to reduce the probability of a major outbreak (p) starting from a single nonsymptomatic host. **A.** The major outbreak probability for different values of ρ_1 and ρ_2, with the steepest descent contours overlaid (white lines). For the maximum reduction in the probability of a major outbreak at each point, surveillance must be enhanced for both nonsymptomatic and symptomatic individuals, with different levels of prioritisation depending on the current values of ρ_1 and ρ_2. **B.** Values of ρ_1 and ρ_2 for which increasing surveillance for nonsymptomatic hosts (i.e. increasing ρ_1) is more effective at reducing the major outbreak probability than increasing surveillance for symptomatic hosts (i.e. increasing ρ_2) (green region) and vice versa (blue region). The white line represents the steepest descent contour starting from $\rho_1 = \rho_2 = 0$, under the constraint that surveillance can only be enhanced for either symptomatic or nonsymptomatic hosts at any time. The diagonal section of the steepest descent contour is made up of small horizontal and vertical sections. **C.** Strategies for minimising the major outbreak probability for a given fixed total surveillance effort ($\rho_1 + \rho_2 = S$). Red dotted lines indicate contours on which $\rho_1 + \rho_2$ is constant (i.e. lines on which $\rho_1 + \rho_2$ takes the values shown); red circles indicate the points along these contours at which the major outbreak probability is minimised. The white line indicates the optimal surveillance enhancement strategy if the maximum possible surveillance level (i.e. the maximum value of $\rho_1 + \rho_2 = S$) is increased. **D.** Strategies for minimising the surveillance effort required to achieve a pre-specified risk level (an "acceptable" major outbreak probability). Red dotted lines indicate contours of constant major outbreak probability (i.e. lines on which the probability of a major outbreak takes the values shown); red circles indicate the points along these contours at which the total surveillance effort $\rho_1 + \rho_2$ is minimised. The white line indicates the optimal strategy to follow if the pre-specified risk level is increased or reduced.

These results indicate that, if surveillance resources are such that S is greater than 2.8 (corresponding to a 59% reduction in time to isolation of symptomatic hosts), the optimal surveillance strategy involves both enhanced surveillance of symptomatic individuals and nonsymptomatic individuals (the red circles correspond to strictly positive values of both ρ_1 and ρ_2, unless S is less than 2.8).

Objective 2: Minimise the total surveillance effort to achieve a pre-specified reduction in the probability of a major outbreak.
Second, we considered the question: given a pre-specified acceptable risk level (i.e. probability of a major outbreak), how should the surveillance level targeted at nonsymptomatic and symptomatic hosts be chosen? This involves choosing ρ_1 and ρ_2 to minimise $\rho_1 + \rho_2$ along a given contour corresponding to a fixed major outbreak probability (red dotted lines in Fig 3.4D). On each contour, the red circle indicates the point along that contour at which the total surveillance effort $\rho_1 + \rho_2$ is minimised. These optimal points also lie exactly along the line on which enhancing surveillance targeted at symptomatic hosts is equally effective compared to enhancing surveillance targeted at nonsymptomatic hosts.

As long as the target major outbreak probability is less than 0.69, optimal surveillance involves enhanced surveillance of nonsymptomatic individuals as well as symptomatic individuals. For example, in order to reduce the major outbreak probability to 0.6, the optimal approach is to deploy resources such that $\rho_1 = 12.4$ (a 74% reduction in time to isolation of nonsymptomatic individuals) and $\rho_2 = 18.0$ (a 76% reduction in time to isolation of symptomatic individuals).

Plots analogous to Fig 3.4D in which the parameters were varied from their baseline values are shown in Supplementary Figs A.1-A.10. In each case that we considered, our main finding was unchanged. There always exists a threshold major outbreak probability such that, if the target major outbreak probability is below this threshold, the optimal strategy for further reduction in the major outbreak probability involves enhancing surveillance targeting both nonsymptomatic and symptomatic individuals.

3.4 Discussion

A key component of infectious disease epidemic management is inferring the risk of outbreaks in different locations [42, 147–152]. Surveillance and control strategies can be introduced to reduce the risk that imported cases will lead to major outbreaks [12, 64, 87, 153–155]. However, for a range of pathogens, public health measures are hindered by nonsymptomatic infectious hosts who can transmit the pathogen, yet are challenging to detect [12, 123, 125, 141, 152, 156]. Here, we showed how the probability of a major outbreak

can be estimated using a branching process model that accounts for transmission from nonsymptomatic infected individuals (Fig 3.1A). The model can be used to assess the major outbreak probability for different surveillance strategies that target nonsymptomatic or symptomatic hosts (Fig 3.3). Previous studies have shown that detection of nonsymptomatic infections can be a key component of epidemic forecasting [152] and containment [12], and have demonstrated the benefits of identifying and isolating infectious nonsymptomatic hosts to reduce transmission [141, 157]. We focused instead on investigating how surveillance should be targeted at nonsymptomatic or symptomatic hosts in order to reduce the probability that cases imported to new locations will trigger a major outbreak (Figs 3.4A, 3.4B). We also showed how the optimal surveillance level targeting these two groups can be assessed when surveillance resources are limited and policy makers have specific objectives (Figs 3.4C, 3.4D). In each case, our main conclusion was that surveillance for nonsymptomatic infected hosts ($\rho_1 > 0$) can be an important component of reducing the major outbreak risk during epidemics. This result has broad implications, and our analysis could be extended to assess the potential for containing outbreaks at their source using a range of specific interventions targeting symptomatic and nonsymptomatic hosts.

Our goal here was to use the simplest possible model to explore the effects of surveillance of nonsymptomatic and symptomatic individuals on the risk of major outbreaks. However, this model is not without its limitations. One area of uncertainty is the precise values of the parameters governing pathogen transmission and control. In this chapter, we chose a baseline set of parameter values that is consistent with the findings of studies conducted during the COVID-19 pandemic, although constructing a detailed transmission model for this pandemic was not our main focus. For example, we set the relative rates at which presymptomatic and asymptomatic individuals generate new infections compared to symptomatic individuals so that 48.9% of transmissions arise from presymptomatic infectors, and 10.6% arise from asymptomatic infectors [141]. While this is in line with reported estimates [98, 158], there is substantial variation between studies. Similarly, the proportion of individuals who experience a fully asymptomatic course of infection (denoted by ξ in our model) is subject to a considerable degree of uncertainty. Here, we chose $\xi = 0.2$ as the baseline value [138–140] but estimates in the literature range from 0.04 to over 0.8 [139, 159–161]. We therefore also conducted sensitivity analyses in which we explored a range of different values of model parameters (Supplementary Text A.1 and Supplementary Figs A.1-A.10). In each case that we considered, our main conclusion was unchanged: surveillance of nonsymptomatic individuals can contribute to reducing the risk of major outbreaks. This result is expected to hold for epidemics of any pathogen for which nonsymptomatic individuals contribute significantly to transmission.

For our modelling approach to be used to make precise quantitative predictions during

epidemics, it would be necessary to update the model to include the range of different specific surveillance and control interventions that are in place. For example, detection of nonsymptomatic infected individuals is facilitated by contact tracing and antigen testing, which are carried out routinely during epidemics and can be included in models explicitly [12, 87, 162, 163]. Reductions in contacts due to social distancing strategies and school or workplace closures could also be accounted for [46, 164], although such interventions are often introduced after a major outbreak has begun rather than in the initial phase of a potential major outbreak as considered here. We modelled the level of surveillance targeted at nonsymptomatic and symptomatic hosts in a simple way, using a function describing the relationship between surveillance effort and effectiveness (Fig 3.1B). We assumed that this general functional relationship could be applied to interventions targeting both symptomatic and nonsymptomatic hosts, accounting for logistical differences in the ease of targeting either group by scaling the effectiveness of surveillance for nonsymptomatic hosts using the parameter ε (results are shown for different values of ε in Supplementary Fig A.6). In principle, it would be possible to include entirely different functional forms describing the relationship between surveillance effort and effectiveness for strategies targeting symptomatic and nonsymptomatic individuals, and these could be tailored to the effects of particular interventions. If different public health measures are included in the model explicitly, then it would be possible to increase the accuracy of assessments of the relative public health benefits of specific interventions that only target symptomatic individuals (e.g. screening for passengers with heightened temperatures at airports [55, 165]) compared to interventions that also target nonsymptomatic hosts (e.g. travel bans or quarantine of all inbound passengers [166, 167]). Of course, this would require data from which the relative effectiveness of different measures could be inferred.

The underlying transmission model could also be extended to include additional realism in several ways. Transmission dynamics are influenced by marked heterogeneities in the patterns of contacts between individuals in different age groups [106, 168], and, for COVID-19, susceptibility to infection, the likelihood of developing symptoms, and the average severity of those symptoms increase with age [107, 169]. Age-dependent variation in the proportion of asymptomatic cases in particular implies that the optimal balance of surveillance between symptomatic and nonsymptomatic hosts may differ between age groups. In the next chapter of this thesis, we will therefore further extend the methodology developed in this chapter to account for the age structure of the population. Similarly, for a range of infectious diseases, the distribution characterising the number of secondary infections generated by each infected host (the offspring distribution) exhibits a high degree of overdispersion [170–173]. For a fixed value of R_0, a higher degree of overdispersion increases the likelihood that initial cases will fade out without leading to a major outbreak [116, 174], and suggests that greater reductions in major outbreak risks could theoretically be achieved for the same surveillance

effort, if potential superspreaders or superspreading events can be identified and targeted.

Despite the necessary simplifications, we have shown how the risk of major outbreaks can be estimated during epidemics using a branching process model that includes nonsymptomatic infectious hosts explicitly. Determining the extent to which nonsymptomatic individuals contribute to transmission is essential early in emerging epidemics of a novel pathogen. As we have shown, if transmissions occur from nonsymptomatic infectors, dedicating surveillance resources towards finding nonsymptomatic cases can be an important component of public health measures that aim to prevent major outbreaks.

Appendix A

Supplementary material for Chapter 3

A.1 Robustness of results to parameter values used

We conducted supplementary analyses to investigate how the results presented in Chapter 3 were affected by varying the model parameters from their baseline values given in Table 3.1. Specifically, we performed sensitivity analyses on the values of R_0, ξ, K_p, K_a, $\gamma + \mu$, ε, λ, ν, and δ. For each of these, we present plots analogous to Fig 3.4D for six different values of the relevant parameter (Supplementary Figs A.1-A.10). In each case we considered, our qualitative message was unchanged: whenever the maximum acceptable risk level was below a particular threshold value, the optimal strategy involved surveillance targeting both nonsymptomatic and symptomatic individuals.

A.2 The expected time to isolation, given that isolation occurs before recovery

In this Supplementary Text, we derive the expression $1/(\gamma + \mu)$ for the expected time to isolation of a symptomatic host conditional on the fact that isolation occurs before the individual recovers (see Section 3.2.3).

Let X and Y be the random variables representing the time to isolation and the time to recovery of an infected host respectively. Then, $X \sim \exp(\gamma)$ and $Y \sim \exp(\mu)$, so that the probability density functions of X and Y are given by $f_X(x) = \gamma \exp(-\gamma x)$ and $f_Y(y) = \mu \exp(-\mu y)$ respectively, and the corresponding cumulative distribution functions of X and Y are $F_X(x) = 1 - \exp(-\gamma x)$, $F_Y(y) = 1 - \exp(-\mu y)$. To obtain the expected time to isolation, given that isolation occurs before recovery, we must calculate the value of $\mathbb{E}[X|X < Y]$.

To begin, we define the random variable

$$X_{\mathbb{I}_{X<Y}} := \begin{cases} X & \text{if } X < Y, \\ 0 & \text{if } X > Y, \end{cases}$$

and calculate two useful expectations that involve this variable. Firstly, we note that

$$\mathbb{E}[X_{\mathbb{I}_{X<Y}}|X] = X\mathbb{P}(X < Y|X) = X(1 - F_Y(X)) = X\exp(-\mu X). \tag{A.1}$$

Secondly, we note that

$$\begin{aligned} \mathbb{E}[X_{\mathbb{I}_{X<Y}}] &= \mathbb{E}[X_{\mathbb{I}_{X<Y}}|X < Y]\,\mathbb{P}(X < Y) + \mathbb{E}[X_{\mathbb{I}_{X<Y}}|X > Y]\,\mathbb{P}(X > Y) \\ &= \mathbb{E}[X|X < Y]\,\mathbb{P}(X < Y). \end{aligned} \tag{A.2}$$

Rearranging Equation (A.2), applying the tower property of expectations, and using Equation (A.1), we therefore have that

$$\mathbb{E}[X|X < Y] = \frac{\mathbb{E}[X_{\mathbb{I}_{X<Y}}]}{\mathbb{P}(X < Y)} = \frac{\mathbb{E}\Big[\mathbb{E}[X_{\mathbb{I}_{X<Y}}|X]\Big]}{\mathbb{P}(X < Y)} = \frac{\mathbb{E}\Big[X\exp(-\mu X)\Big]}{\mathbb{P}(X < Y)}. \tag{A.3}$$

The numerator of this expression can be calculated as follows:

$$\mathbb{E}\Big[X\exp(-\mu X)\Big] = \int_{x=0}^{\infty} x\mathrm{e}^{-\mu x} f_X(x)\,\mathrm{d}x = \int_{x=0}^{\infty} \gamma x\mathrm{e}^{-(\gamma+\mu)x}\,\mathrm{d}x.$$

A straightforward integration by parts yields the solution $\mathbb{E}[X\exp(-\mu X)] = \gamma/(\gamma+\mu)^2$, and noting that $\mathbb{P}(X < Y) = \gamma/(\gamma+\mu)$, from Equation (A.3) we obtain $\mathbb{E}[X|X < Y] = 1/(\gamma+\mu)$.

A.3 Probability of a major outbreak

In the Methods section of Chapter 3, we outlined an approach for deriving the probability of a major outbreak starting from a single infectious host in either the presymptomatic, symptomatic, or asymptomatic classes. Here, we provide more details about that derivation. The probability of a major outbreak not occurring, starting from i presymptomatic hosts, j symptomatic hosts, and k asymptomatic hosts, is denoted by $q_{i,j,k}$. If we consider the temporal evolution of (i, j, k) to be a Markov process on the state space $M = Z_{\geqslant 0} \times Z_{\geqslant 0} \times Z_{\geqslant 0}$, then $q_{i,j,k}$ is the hitting probability of the state $(0, 0, 0)$ starting from the state (i, j, k). The vector of hitting probabilities $q^* = \{q^*_{i,j,k}|(i, j, k) \in M\}$ is therefore the minimal non-negative (real) solution to the following system of equations:

$$q_{0,0,0} = 1,$$

$$q_{i,j,k} = \sum_{(l,m,n)\in M} p_{(i,j,k),(l,m,n)}\, q_{(l,m,n)} \quad \text{for} \quad (i,j,k) \neq (0,0,0),$$

where $p_{(i,j,k),(l,m,n)}$ is the transition probability from state (i,j,k) to state (l,m,n) [76]. Here, minimality means that if $\hat{q} = \{\hat{q}_{i,j,k} | (i,j,k) \in M\}$ is another non-negative real solution, then $q^*_{i,j,k} \leqslant \hat{q}_{i,j,k}$ for all $(i,j,k) \in M$.

From this, Equations (3.5), (3.7), and (3.8) in the main text are obtained. These equations may be reduced to the following quartic equation for $q_{0,0,1}$:

$$\omega_4\, q^4_{0,0,1} + \omega_3\, q^3_{0,0,1} + \omega_2\, q^2_{0,0,1} + \omega_1\, q_{0,0,1} + \omega_0 = 0,$$

where

$\omega_4 = d(a-d)(d-c)\xi;$
$\omega_3 = cd(a-d)(1-d)\xi - bd^3(1-c)(1-\xi) + (d-c)[d-a-ad\xi + d^2(a-1+b(1-\xi)+\xi)];$
$\omega_2 = c(1-d)[d-a-ad\xi + d^2(a-1+b(1-\xi)+\xi)] + (d-c)[d(d-2a-1)+2a];$
$\omega_1 = c(1-d)[d(d-2a-1)+2a] - a(1-d)^2(d-c);$
$\omega_0 = -ac(1-d)^3;$

and the parameters a, b, c, and d are defined as follows:

$$a = \frac{\alpha\beta}{r_{I_1}}, \quad b = \frac{\lambda}{r_{I_1}}, \quad c = \frac{\beta}{r_{I_2}}, \quad d = \frac{\eta\beta}{r_A}.$$

Expressions for r_{I_1}, r_{I_2}, and r_A are as given in the main text.

This yields four solutions for $q_{0,0,1}$ and four corresponding solutions for each of $q_{1,0,0}$ and $q_{0,1,0}$. One solution is always given by $q_{1,0,0} = q_{0,1,0} = q_{0,0,1} = 1$; the other solutions can be found numerically. As described above, we take the minimal non-negative real solution $q_{1,0,0} = q^*_{1,0,0}$, $q_{0,1,0} = q^*_{0,1,0}$, $q_{0,0,1} = q^*_{0,0,1}$, and observe that the probability of a major outbreak occurring starting from i presymptomatic hosts, j symptomatic hosts, and k asymptomatic hosts is simply $1 - q_{i,j,k}$, giving the result stated in the main text.

If required, this result can be confirmed for specific model parameter values via repeated stochastic simulation of the branching process model, starting from i presymptomatic hosts, j symptomatic hosts, and k asymptomatic hosts.

A.4 Supplementary Figures

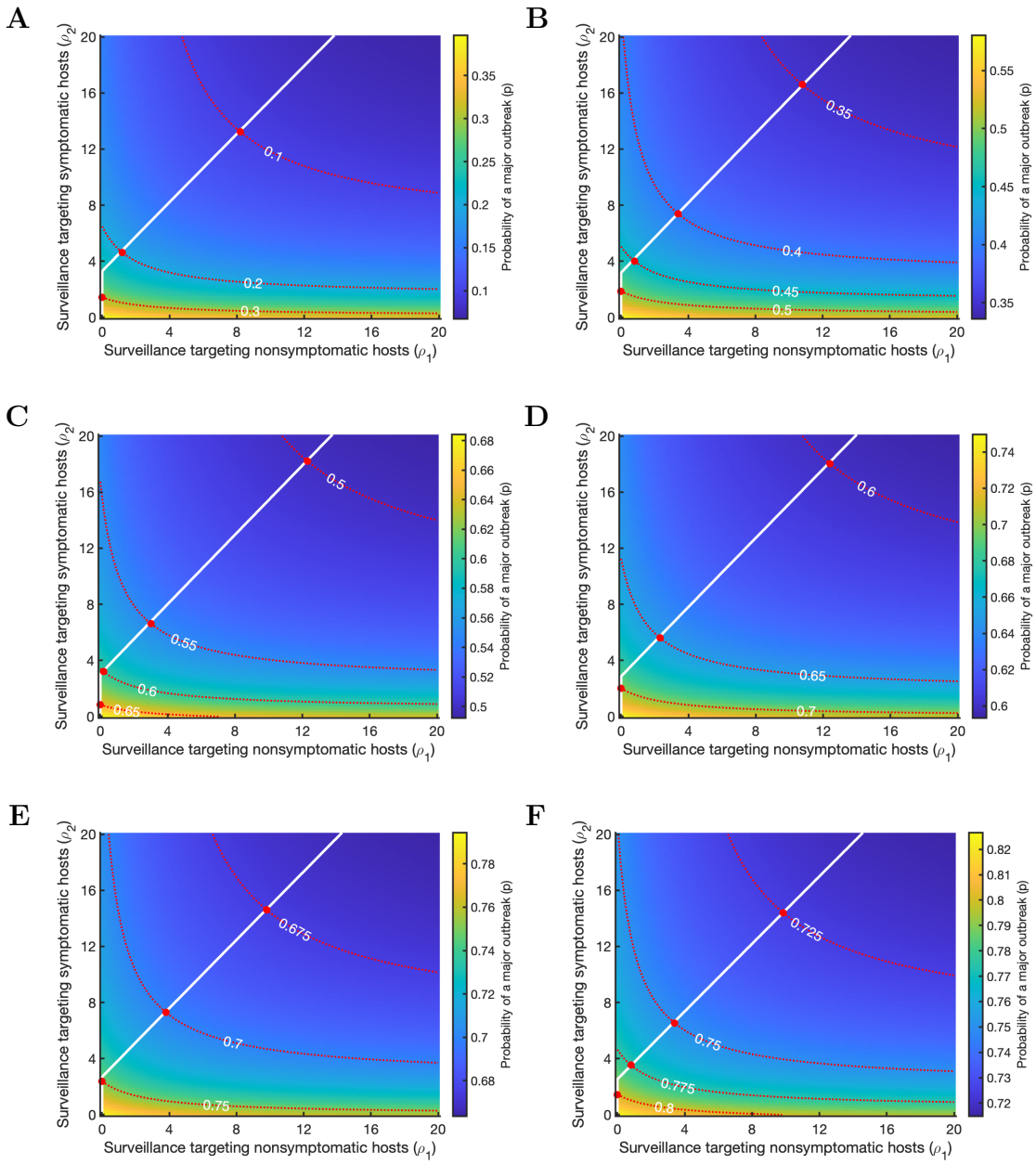


Figure A.1: Varying the basic reproduction number R_0 from its baseline value ($R_0 = 3$). Plots are analogous to Fig 3.4D in the main text, showing strategies for minimising the surveillance effort required to achieve a pre-specified risk level (an "acceptable" major outbreak probability). Red dotted lines represent contours along which the probability of a major outbreak is constant, as labelled; red circles indicate the points along these contours at which the total surveillance effort $\rho_1 + \rho_2$ is minimised. The white line indicates the optimal strategy to follow if the pre-specified risk level is reduced. Apart from R_0 and β (which is changed in each panel to set the value of R_0), all parameters are held fixed at their baseline values given in Table 3.1. **A.** $R_0 = 1.5$. **B.** $R_0 = 2$. **C.** $R_0 = 2.5$. **D.** $R_0 = 3$ (baseline). **E.** $R_0 = 3.5$. **F.** $R_0 = 4$.

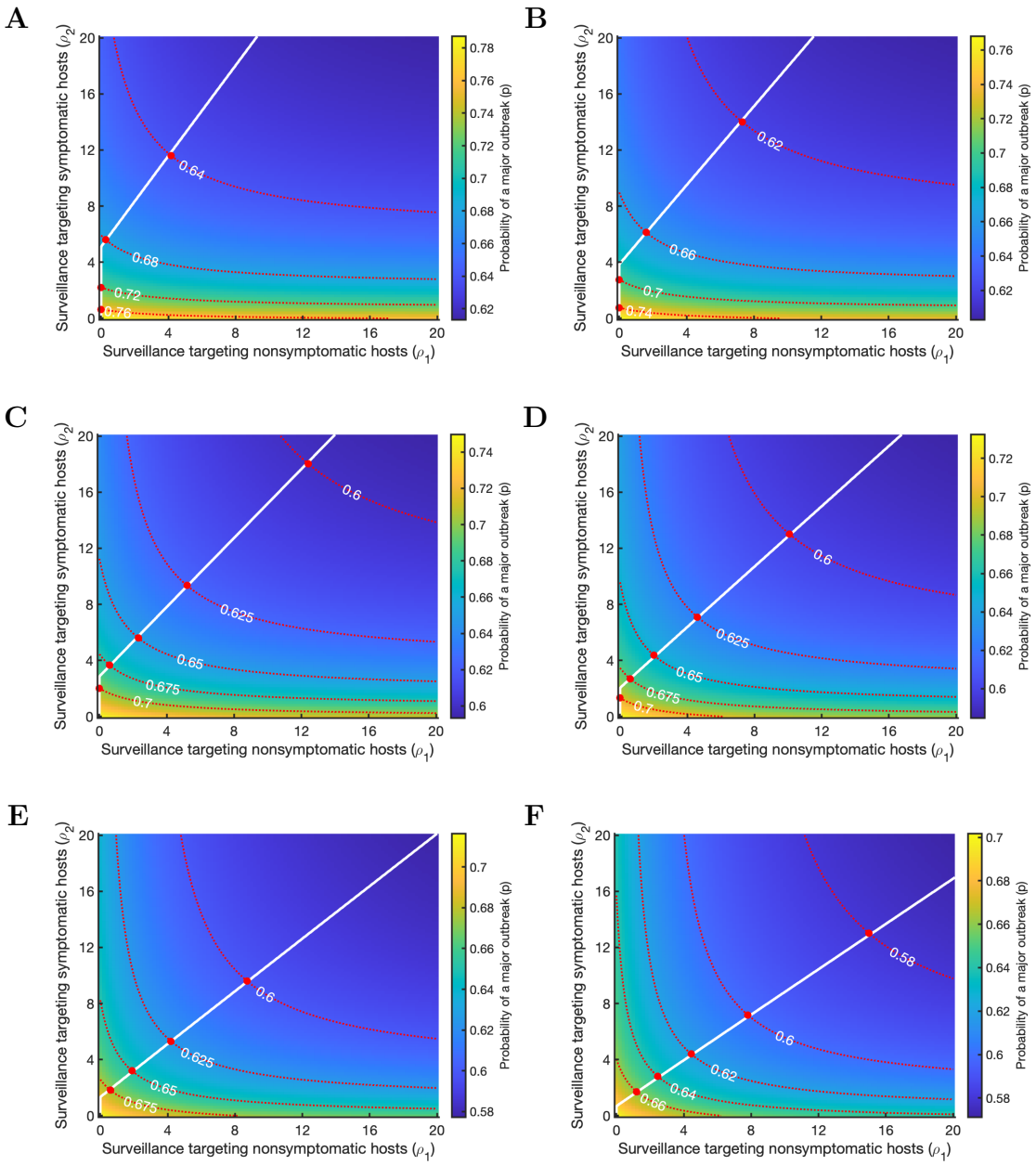


Figure A.2: Varying the proportion of hosts who are asymptomatic, ξ, from its baseline value ($\xi = 0.2$). Plots are analogous to Fig 3.4D in the main text, showing strategies for minimising the surveillance effort required to achieve a pre-specified risk level (an "acceptable" major outbreak probability). Red dotted lines represent contours along which the probability of a major outbreak is constant, as labelled; red circles indicate the points along these contours at which the total surveillance effort $\rho_1 + \rho_2$ is minimised. The white line indicates the optimal strategy to follow if the pre-specified risk level is reduced. Apart from ξ and β (which is changed in each panel to set $R_0 = 3$), all parameters are held fixed at their baseline values given in Table 3.1. **A.** $\xi = 0$. **B.** $\xi = 0.1$. **C.** $\xi = 0.2$ (baseline). **D.** $\xi = 0.3$. **E.** $\xi = 0.4$. **F.** $\xi = 0.5$.

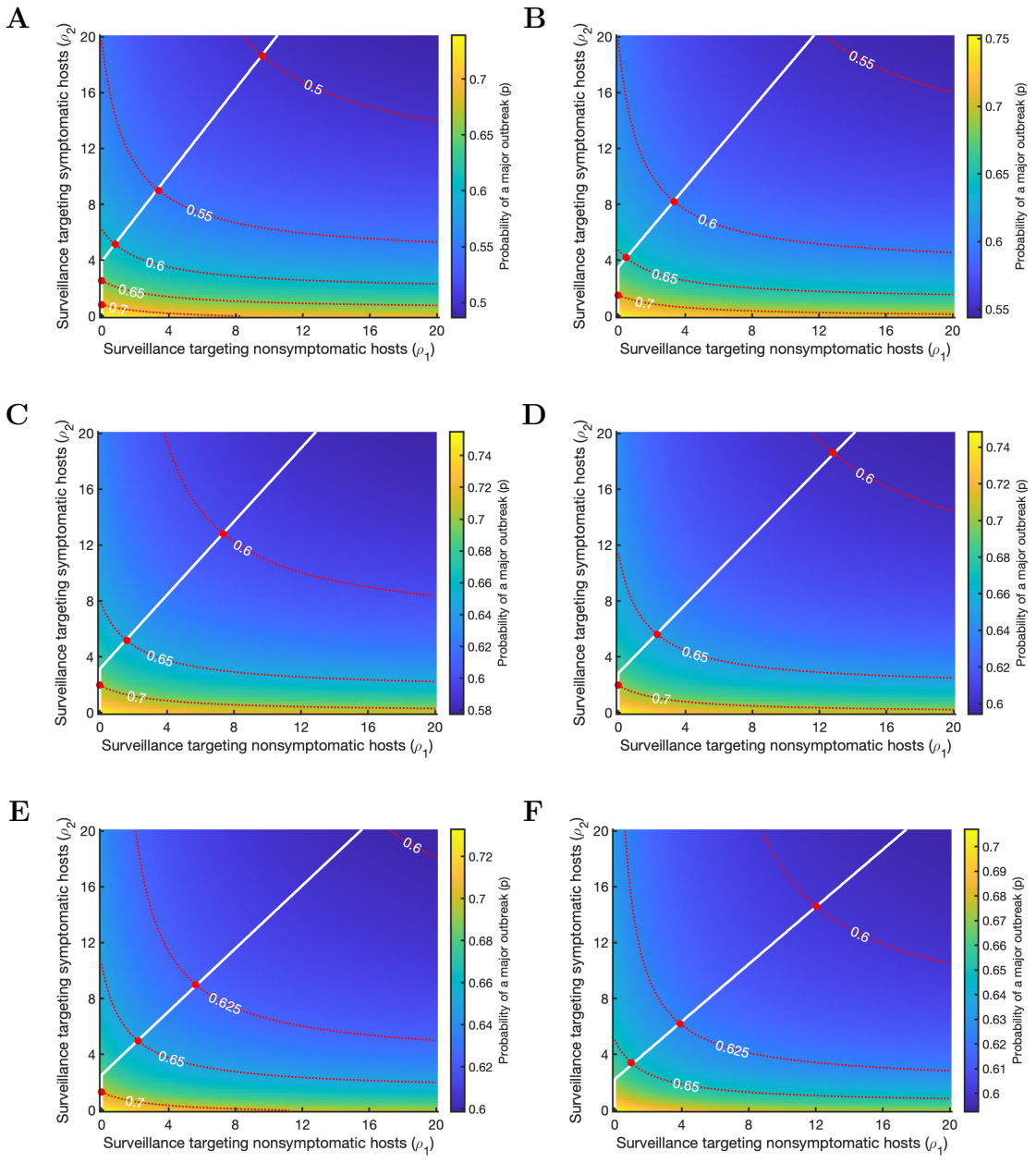


Figure A.3: Varying the proportion of infections arising from presymptomatic hosts in the absence of intensified surveillance (K_p, given by expression 3.2 in the main text) from its baseline value ($K_p = 0.489$). In each case, the proportions of infections arising from symptomatic and asymptomatic hosts are adjusted so that they remain in the same ratio as in the baseline case. Plots are analogous to Fig 3.4D in the main text, showing strategies for minimising the surveillance effort required to achieve a pre-specified risk level (an "acceptable" major outbreak probability). Red dotted lines represent contours along which the probability of a major outbreak is constant, as labelled; red circles indicate the points along these contours at which the total surveillance effort $\rho_1 + \rho_2$ is minimised. The white line indicates the optimal strategy to follow if the pre-specified risk level is reduced. Apart from K_p and K_a, as well as α and η (which are changed in each panel to set the values of K_p and K_a), and β (which is changed in each panel to set $R_0 = 3$), all parameters are held fixed at their baseline values given in Table 3.1. **A.** $K_p = 0.2$. **B.** $K_p = 0.3$. **C.** $K_p = 0.4$. **D.** $K_p = 0.5$. **E.** $K_p = 0.6$. **F.** $K_p = 0.7$.

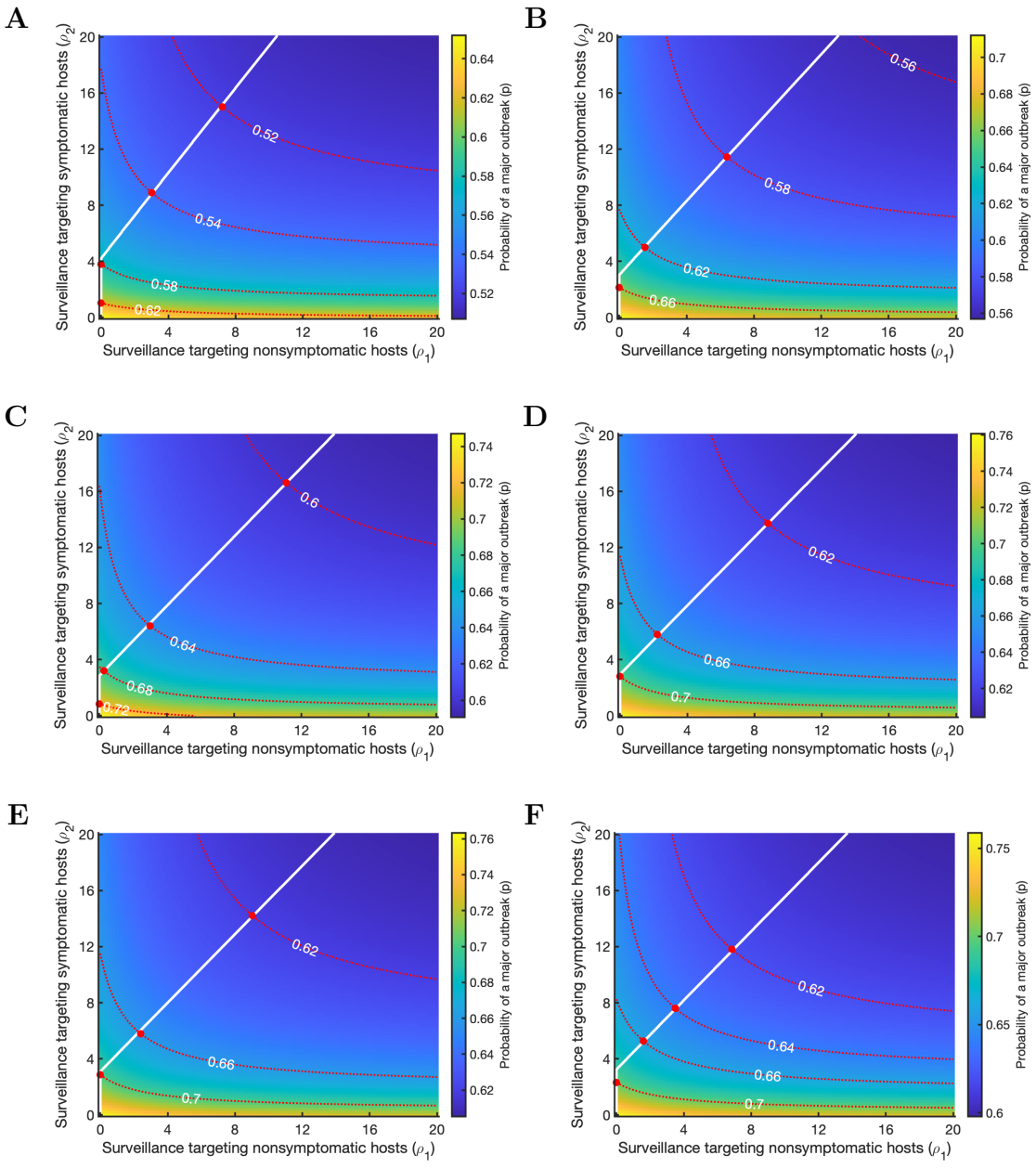


Figure A.4: Varying the proportion of infections arising from asymptomatic hosts in the absence of intensified surveillance (K_a, given by expression 3.3 in the main text) from its baseline value ($K_a = 0.106$). In each case, the proportions of infections arising from presymptomatic and symptomatic hosts are adjusted so that they remain in the same ratio as in the baseline case. Plots are analogous to Fig 3.4D in the main text, showing strategies for minimising the surveillance effort required to achieve a pre-specified risk level (an "acceptable" major outbreak probability). Red dotted lines represent contours along which the probability of a major outbreak is constant, as labelled; red circles indicate the points along these contours at which the total surveillance effort $\rho_1 + \rho_2$ is minimised. The white line indicates the optimal strategy to follow if the pre-specified risk level is reduced. Apart from K_p and K_a, as well as α and η (which are changed in each panel to set the values of K_p and K_a), and β (which is changed in each panel to set $R_0 = 3$), all parameters are held fixed at their baseline values given in Table 3.1. **A.** $K_a = 0.01$. **B.** $K_a = 0.05$. **C.** $K_a = 0.1$. **D.** $K_a = 0.15$. **E.** $K_a = 0.2$. **F.** $K_a = 0.25$.

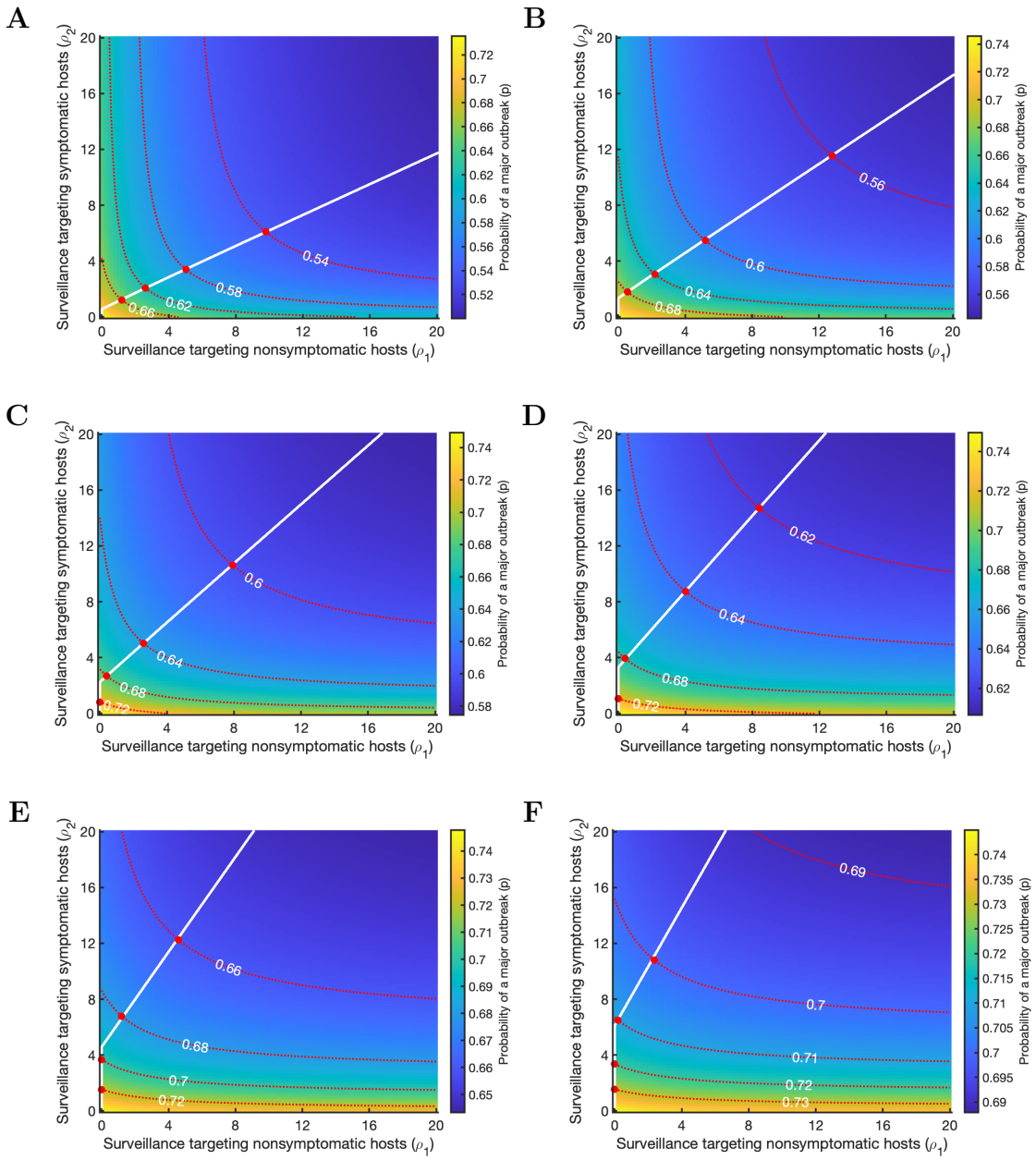


Figure A.5: Varying the expected time period to isolation conditional on isolation occurring during the symptomatic period, $1/(\gamma+\mu)$, from its baseline value ($1/(\gamma+\mu) = 4.6$ days). This is achieved by varying the parameter γ, whilst holding the recovery rate of symptomatic individuals μ equal to its baseline value ($\mu = 1/8$ days^{-1}). Plots are analogous to Fig 3.4D in the main text, showing strategies for minimising the surveillance effort required to achieve a pre-specified risk level (an "acceptable" major outbreak probability). Red dotted lines represent contours along which the probability of a major outbreak is constant, as labelled; red circles indicate the points along these contours at which the total surveillance effort $\rho_1 + \rho_2$ is minimised. The white line indicates the optimal strategy to follow if the pre-specified risk level is reduced. Apart from γ and β (which is changed in each panel to set $R_0 = 3$), all parameters are held fixed at their baseline values given in Table 3.1. **A.** $1/(\gamma + \mu) = 2$ days. **B.** $1/(\gamma + \mu) = 3$ days. **C.** $1/(\gamma + \mu) = 4$ days. **D.** $1/(\gamma + \mu) = 5$ days. **E.** $1/(\gamma + \mu) = 6$ days. **F.** $1/(\gamma + \mu) = 7$ days.

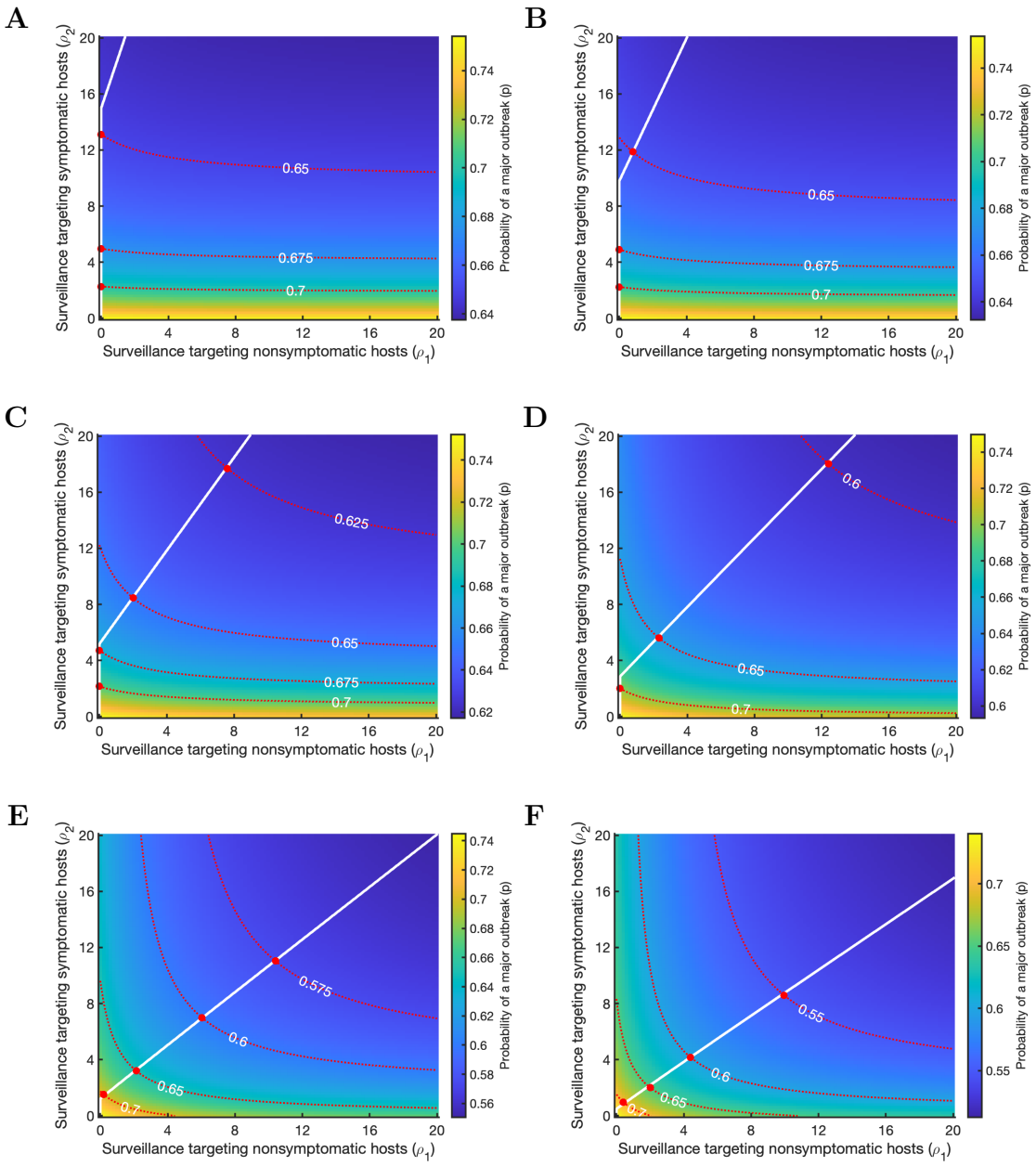


Figure A.6: Varying ε, the relative isolation rate of nonsymptomatic individuals without intensified surveillance (compared to symptomatic individuals), from its baseline value ($\varepsilon = 0.1$). Plots are analogous to Fig 3.4D in the main text, showing strategies for minimising the surveillance effort required to achieve a pre-specified risk level (an "acceptable" major outbreak probability). Red dotted lines represent contours along which the probability of a major outbreak is constant, as labelled; red circles indicate the points along these contours at which the total surveillance effort $\rho_1 + \rho_2$ is minimised. The white line indicates the optimal strategy to follow if the pre-specified risk level is reduced. Apart from ε and β (which is changed in each panel to set $R_0 = 3$), all parameters are held fixed at their baseline values given in Table 3.1. **A.** $\varepsilon = 0.01$. **B.** $\varepsilon = 0.02$. **C.** $\varepsilon = 0.05$. **D.** $\varepsilon = 0.1$ (baseline). **E.** $\varepsilon = 0.2$. **F.** $\varepsilon = 0.3$.

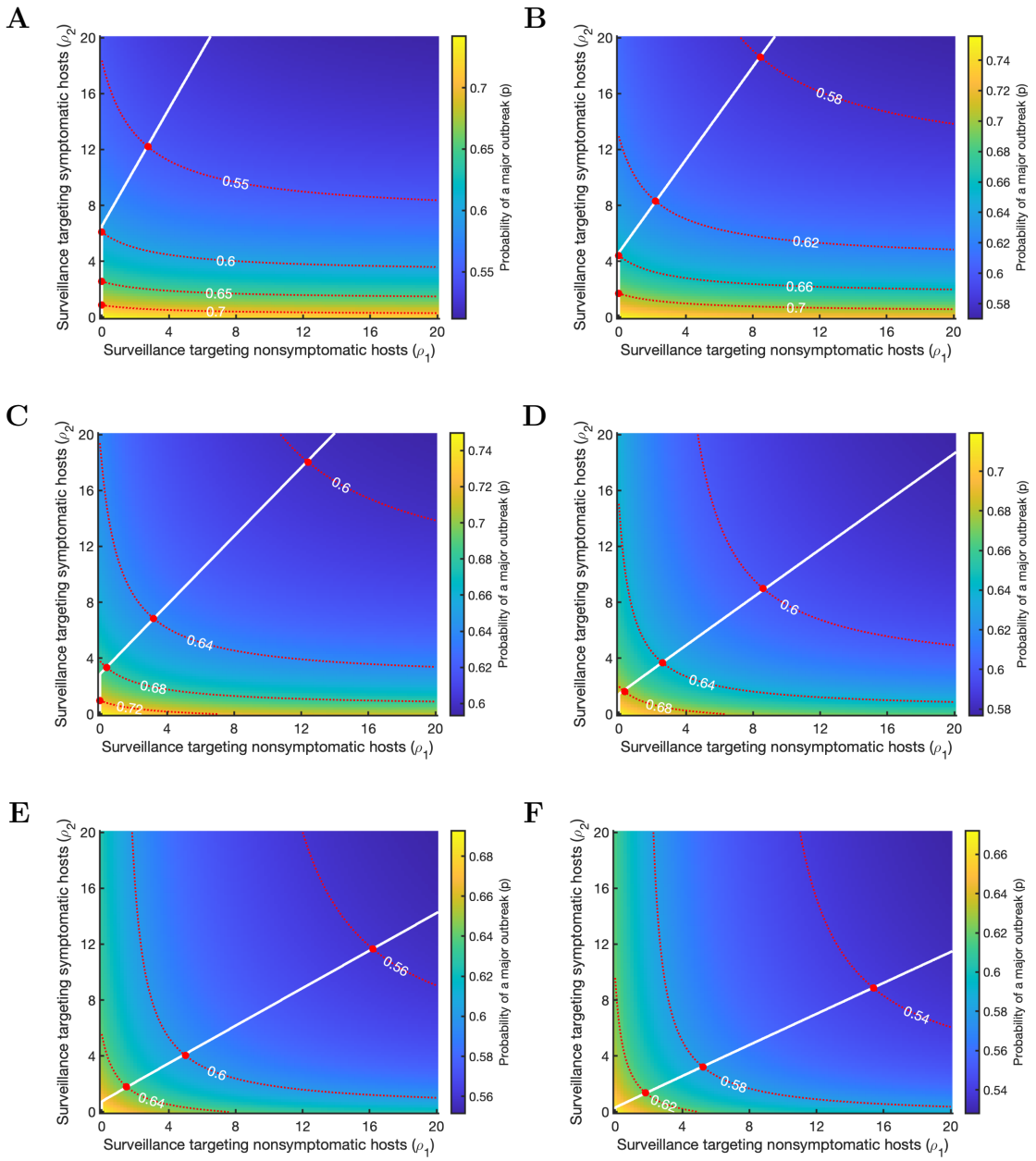


Figure A.7: Varying the duration of the presymptomatic period, $1/\lambda$, from its baseline value ($1/\lambda = 0.5$ days). Plots are analogous to Fig 3.4D in the main text, showing strategies for minimising the surveillance effort required to achieve a pre-specified risk level (an "acceptable" major outbreak probability). Red dotted lines represent contours along which the probability of a major outbreak is constant, as labelled; red circles indicate the points along these contours at which the total surveillance effort $\rho_1 + \rho_2$ is minimised. The white line indicates the optimal strategy to follow if the pre-specified risk level is reduced. Apart from λ and β (which is changed in each panel to set $R_0 = 3$), all parameters are held fixed at their baseline values given in Table 3.1. **A.** $1/\lambda = 0.5$ days. **B.** $1/\lambda = 1$ day. **C.** $1/\lambda = 2$ days (baseline). **D.** $1/\lambda = 4$ days. **E.** $1/\lambda = 6$ days. **F.** $1/\lambda = 8$ days.

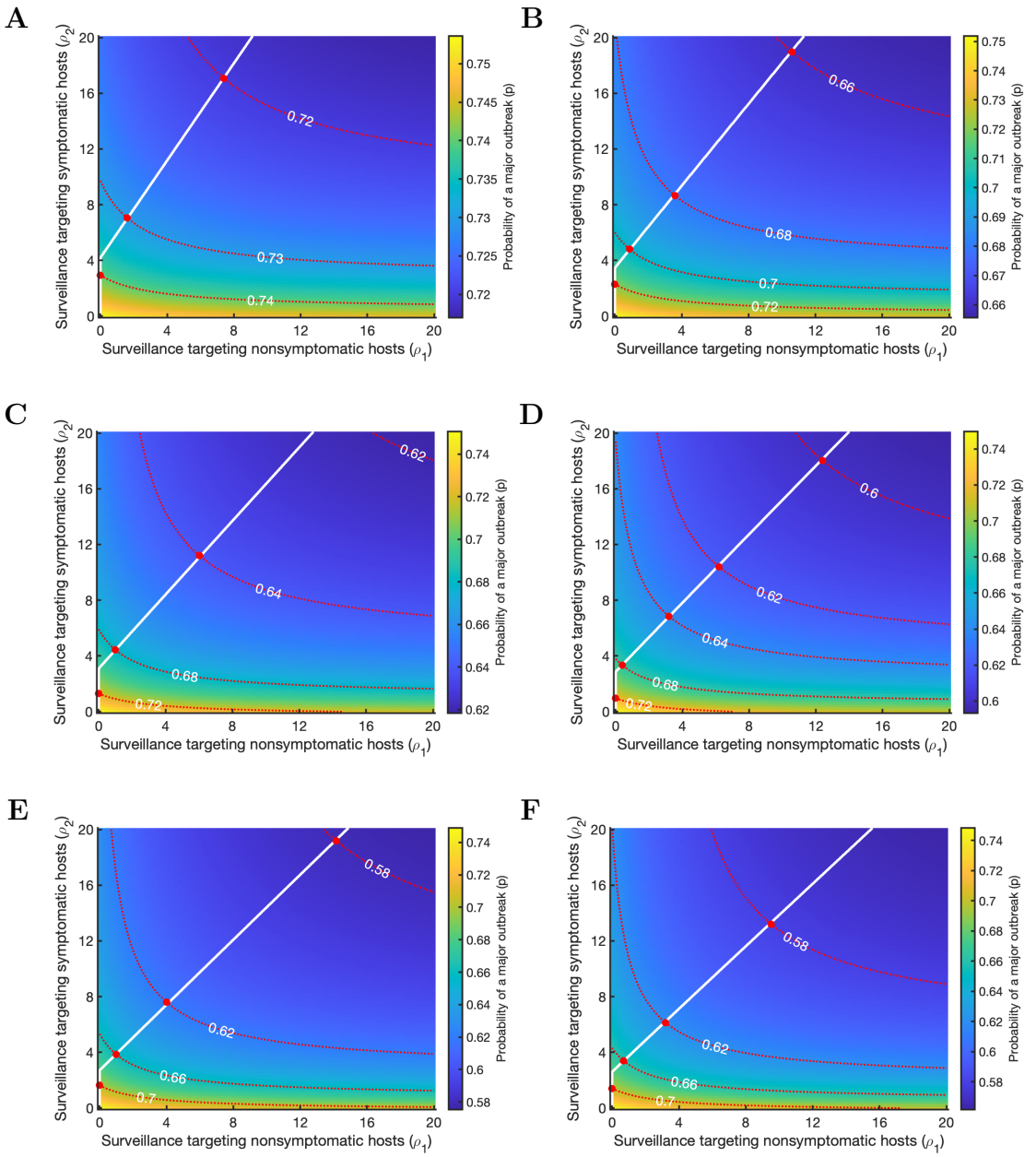


Figure A.8: Varying the duration of the symptomatic period, $1/\mu$, from its baseline value ($1/\mu = 8$ days). The parameter γ is varied simultaneously such that $1/(\gamma + \mu)$, the expected time period to isolation conditional on isolation occurring during the symptomatic period, remains at its baseline value ($1/(\gamma + \mu) = 4.6$ days). Plots are analogous to Fig 3.4D in the main text, showing strategies for minimising the surveillance effort required to achieve a pre-specified risk level (an "acceptable" major outbreak probability). Red dotted lines represent contours along which the probability of a major outbreak is constant, as labelled; red circles indicate the points along these contours at which the total surveillance effort $\rho_1 + \rho_2$ is minimised. The white line indicates the optimal strategy to follow if the pre-specified risk level is reduced. Apart from μ and γ, and β (which is changed in each panel to set $R_0 = 3$), all parameters are held fixed at their baseline values given in Table 3.1. **A.** $1/\mu = 5$ days. **B.** $1/\mu = 6$ days. **C.** $1/\mu = 7$ days. **D.** $1/\mu = 8$ days (baseline). **E.** $1/\mu = 9$ days. **F.** $1/\mu = 10$ days.

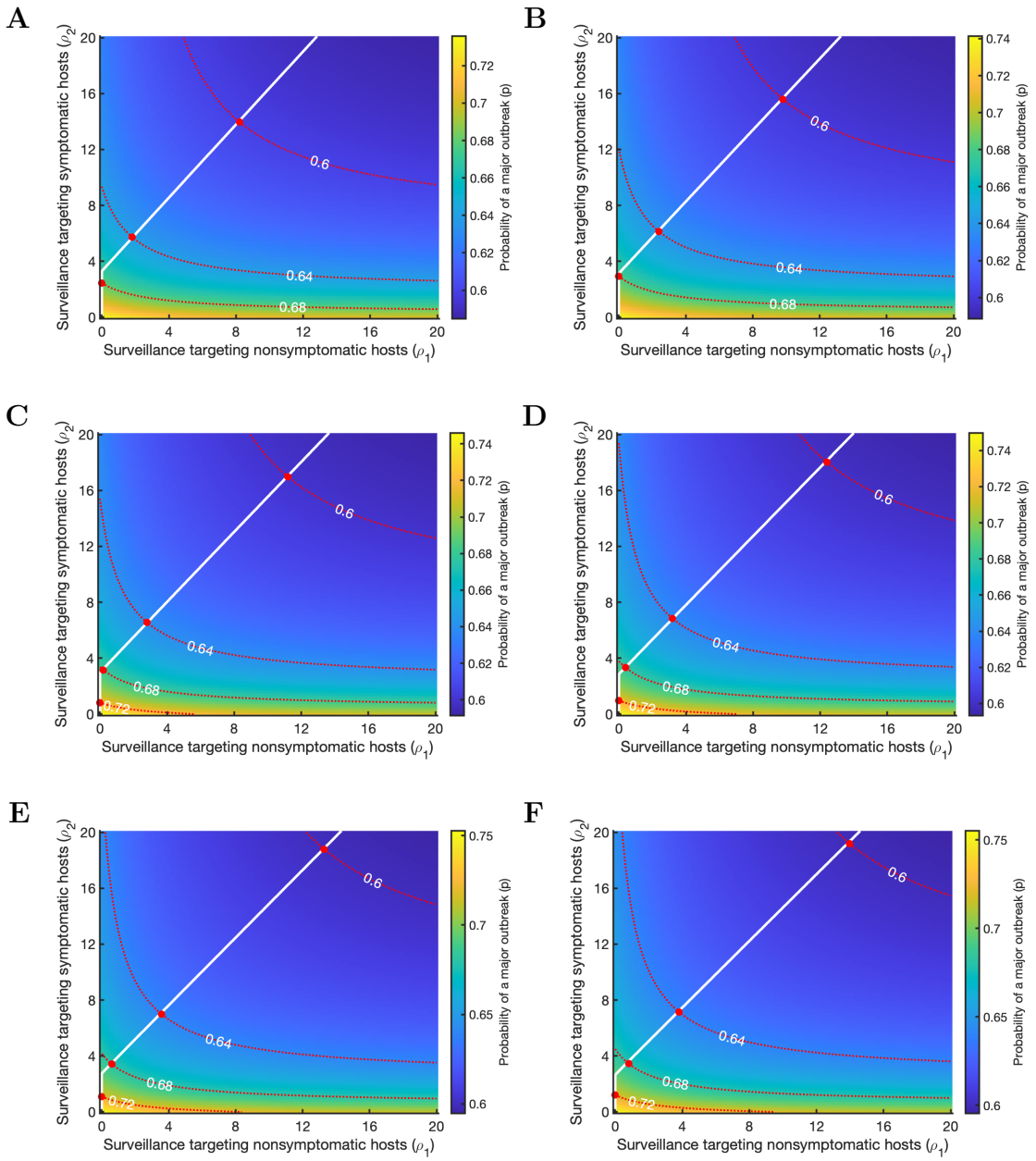


Figure A.9: Varying $1/\nu$, the duration of the asymptomatic period, from its baseline value ($1/\nu = 10$ days). Plots are analogous to Fig 3.4D in the main text, showing strategies for minimising the surveillance effort required to achieve a pre-specified risk level (an "acceptable" major outbreak probability). Red dotted lines represent contours along which the probability of a major outbreak is constant, as labelled; red circles indicate the points along these contours at which the total surveillance effort $\rho_1 + \rho_2$ is minimised. The white line indicates the optimal strategy to follow if the pre-specified risk level is reduced. Apart from ν and β (which is changed in each panel to set $R_0 = 3$), all parameters are held fixed at their baseline values given in Table 3.1. **A.** $1/\nu = 7$ days. **B.** $1/\nu = 8$ days. **C.** $1/\nu = 9$ days. **D.** $1/\nu = 10$ days (baseline). **E.** $1/\nu = 11$ days. **F.** $1/\nu = 12$ days.

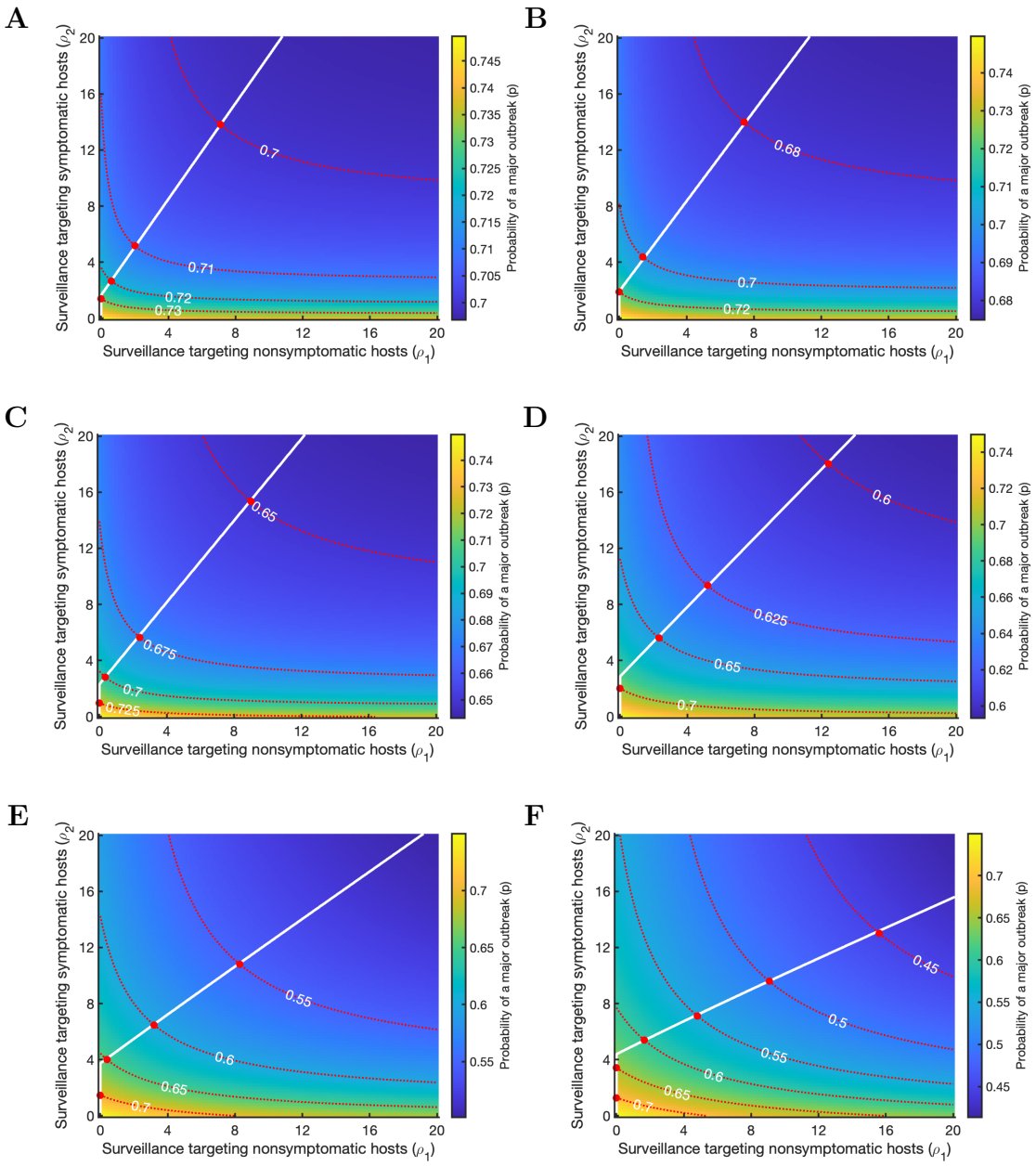


Figure A.10: Varying the upper bound on the fractional reduction in the time to isolation (if no other event occurs), δ, from its baseline value ($\delta = 0.8$). Plots are analogous to Fig 3.4D in the main text, showing strategies for minimising the surveillance effort required to achieve a pre-specified risk level (an "acceptable" major outbreak probability). Red dotted lines represent contours along which the probability of a major outbreak is constant, as labelled; red circles indicate the points along these contours at which the total surveillance effort $\rho_1 + \rho_2$ is minimised. The white line indicates the optimal strategy to follow if the pre-specified risk level is reduced. Apart from δ and β (which is changed in each panel to set $R_0 = 3$), all parameters are held fixed at their baseline values given in Table 3.1. **A.** $\delta = 0.5$. **B.** $\delta = 0.6$. **C.** $\delta = 0.7$. **D.** $\delta = 0.8$ (baseline). **E.** $\delta = 0.9$. **F.** $\delta = 0.95$.

Chapter 4

Estimating outbreak risks and the effects of non-pharmaceutical interventions in age-structured populations

4.1 Introduction

In addition to surveillance for infected hosts and their subsequent isolation, throughout the COVID-19 pandemic policy makers worldwide have relied on a range of additional non-pharmaceutical interventions (NPIs) to limit the spread of SARS-CoV-2. Commonly introduced NPIs have included school closures, workplace closures, and population-wide social distancing policies, all of which aim to reduce the numbers of contacts between individuals and disrupt potential chains of transmission [173, 175–177]. Similar measures have previously been adopted for countering other infectious diseases such as Ebola and pandemic influenza [178–180], and are likely to remain a key line of defence against emerging pathogens that are directly transmitted between individuals. NPIs are particularly important when no effective treatment or vaccine is available, and they are also beneficial when vaccination programmes are being rolled out [181–183]. If vaccines do not prevent transmission completely, then NPIs may be important even when vaccination is widespread [67]. However, the negative economic, social, and non-disease health consequences of NPIs have been widely discussed, with the impact of school closures on the academic progress and wellbeing of school-aged individuals a particular concern [180, 184–188]. Therefore, assessing the effectiveness of different NPIs at reducing transmission is critical for determining whether or not they should be used.

Since NPIs such as school and workplace closures affect distinct age groups within the population, when evaluating their effectiveness it is important to account for age-dependent factors that influence transmission. Multiple studies have documented marked heterogeneities in the patterns of contacts between individuals in different age groups, with school-aged indi-

viduals tending to have more contacts each day than older individuals [168, 189–194]. Since close contact between individuals is a key driver of transmission for respiratory pathogens such as influenza viruses and SARS-CoV-2, these contact patterns influence transmission dynamics and consequently the effects of interventions that target different age groups [106, 189, 190, 195–197]. Additionally, many diseases are characterised by significant age-related variations in pathophysiology. For example, for SARS-CoV-2, children may be less susceptible to infection than adults [106–110], and more likely to experience asymptomatic or subclinical courses of infection [107, 108, 111–115]. Since the secondary attack rate (the proportion of close contacts that lead to new infections) from asymptomatic or subclinical hosts is lower than from hosts with clinical symptoms [138, 139, 198–201], children are likely to be less infectious on average than older individuals who are at increased risk of developing symptoms [202–204].

Previous studies have used age-stratified deterministic transmission models to investigate the effects of NPIs on COVID-19 epidemic peak incidence and timing. Prem *et al.* [37] projected the outbreak in Wuhan, China, over a one year period under different control scenarios, and demonstrated that a period of intense control measures including school closures, a 90% reduction in the workforce, and a significant reduction in other social mixing could delay the epidemic peak by several months. Zhang *et al.* [106] predicted that eliminating all school contacts during the outbreak period would lead to a noticeable decrease in the peak incidence and a later peak; however, they did not take differences between symptomatic and asymptomatic cases into account explicitly. In contrast, Davies *et al.* [107] used estimates of age-dependent susceptibility and clinical fraction fitted to the observed age distribution of cases in six countries to demonstrate that school closures alone were unlikely to reduce SARS-CoV-2 transmission substantially. Davies *et al.* [46] subsequently concluded that a combination of several strongly enforced NPIs would be necessary to avoid COVID-19 cases exceeding available healthcare capacity in the UK.

Rather than considering the entire epidemic curve, here we focus on estimating the probability that cases introduced to a new location trigger a major outbreak as opposed to fading out with few cases. Localised clusters of transmission have been a feature of the COVID-19 pandemic [205–207], and assessing the risk that such major outbreaks occur requires a stochastic model in which the pathogen can either invade or fade out. As discussed in Chapter 2, stochastic branching process models have been applied previously to assess outbreak risks for many pathogens without considering different age groups explicitly [64, 65, 87, 147, 208], and extended to consider adults and children as two distinct groups [40, 209]. However, the significant heterogeneities in contact patterns and pathophysiology between individuals across the full range of ages have never previously been considered in estimates of major outbreak risks. Here, we develop an age-structured branching process

model that can be used to estimate the probability of a major outbreak occurring for index cases of different ages, and demonstrate how the age-dependent risk profile changes when susceptibility to infection and clinical fraction vary with age.

We use the model to investigate the effects on the major outbreak probability of NPIs that reduce the numbers of contacts between individuals. Specifically, we use location-specific contact data for the UK detailing the average numbers of daily contacts occurring in school, in the workplace, and elsewhere [168] to model the impacts of school closures, workplace closures, and broader social distancing policies. We demonstrate that, for SARS-CoV-2, contacts occurring outside schools and the workplace are a key driver of sustained transmission. Thus, population-wide social distancing policies that affect individuals of all ages lead to a substantial reduction in the risk of major outbreaks. In contrast, since school-aged individuals only make up around one quarter of the UK population and tend to have large numbers of contacts outside school, school closures are predicted to have only a limited effect when applied as the sole NPI.

We then go on to consider the impacts of mixed strategies made up of multiple NPIs, as well as additional NPIs that do not only reduce numbers of contacts. Specifically, we show that rigorous surveillance and effective isolation of infected hosts can reduce the level of contact-reducing NPIs required to achieve substantial reductions in the risk of a major outbreak. Although we use SARS-CoV-2 as a case study, our approach can be applied more generally to explore the effects of NPIs on the risk of outbreaks of any pathogen for which age-related heterogeneities play a significant role in transmission dynamics.

4.2 Methods

4.2.1 Mathematical model

In this chapter, we consider a branching process model in which the population is divided into 16 age groups, denoted $G_1, G_2, \ldots, G_{16}$. The first 15 groups represent individuals aged $0-74$, divided into five-year intervals ($0-4$, $5-9$, $10-14$, etc.). The final group represents individuals aged 75 and over. The total number of individuals in age group G_k is denoted N_k. Infected individuals in each age group G_k are classified into compartments representing asymptomatic (A_k), presymptomatic (P_k), or symptomatic (S_k) hosts, where an individual in the A_k compartment does not develop symptoms at any time during their course of infection.

An infected individual of any type in group G_k may generate new infections in any age group. In our model, the rate at which a single infected symptomatic individual in group

G_k generates infections in group G_j is given by

$$\beta_{kj} = B\,\tau_k\,C_{kj}\,\omega_j.$$

Here, τ_k represents the infectivity of individuals in group G_k, C_{kj} represents the daily number of unique contacts a single individual in group G_k has with individuals in group G_j, ω_j represents the susceptibility to infection of individuals in group G_j, and B is a scaling factor that can be used to set the reproduction number of the pathogen being considered (see Section 4.2.2). Since we focus here on the initial phase of potential major outbreaks, we do not account for depletion of the susceptible population explicitly. The relative transmission rates from presymptomatic and asymptomatic individuals compared to symptomatic individuals are given by the scaled quantities $\eta\beta_{kj}$ and $\theta\beta_{kj}$, respectively, where η and θ were chosen so that the proportions of transmissions generated by presymptomatic and asymptomatic hosts were in line with literature estimates [141]. The parameter ξ_k represents the proportion of asymptomatic infections in group G_k, so that a new infection in group G_k either increases A_k by one (with probability ξ_k) or increases P_k by one (with probability $1-\xi_k$).

A presymptomatic individual in group G_k may go on to develop symptoms (transition from P_k to S_k) or be detected and isolated (so that P_k decreases by one). A symptomatic individual in group G_k may be detected and isolated as a result of successful surveillance, or may be removed due to self-isolation, recovery or death (so that S_k decreases by one in either case). Similarly, an asymptomatic individual in group G_k may be detected and isolated or recover (so that A_k decreases by one). A schematic of the different possible events in the model is shown in Fig 4.1.

As in Chapter 3, the parameter λ represents the rate at which presymptomatic individuals develop symptoms, so that the expected duration of the presymptomatic infectious period is $1/\lambda$ days in the absence of surveillance of nonsymptomatic infected individuals. Similarly, the expected duration of the asymptomatic infectious period in the absence of surveillance is $1/\nu$ days. The parameter μ represents the rate at which symptomatic individuals are removed as a result of self-isolation, recovery, or death, so that the duration of time for which they are able to infect others is $1/\mu$ days.

For each group G_k, the rate at which symptomatic individuals are detected and isolated as a result of enhanced surveillance is determined by the parameter ρ_k. Analogously, the parameter σ_k governs the rate at which presymptomatic and asymptomatic individuals in G_k are detected and isolated. We assumed that surveillance measures targeting nonsymptomatic hosts are equally effective for those who are presymptomatic and those who are asymptomatic, and therefore used the same rate of isolation due to surveillance for both of these groups.

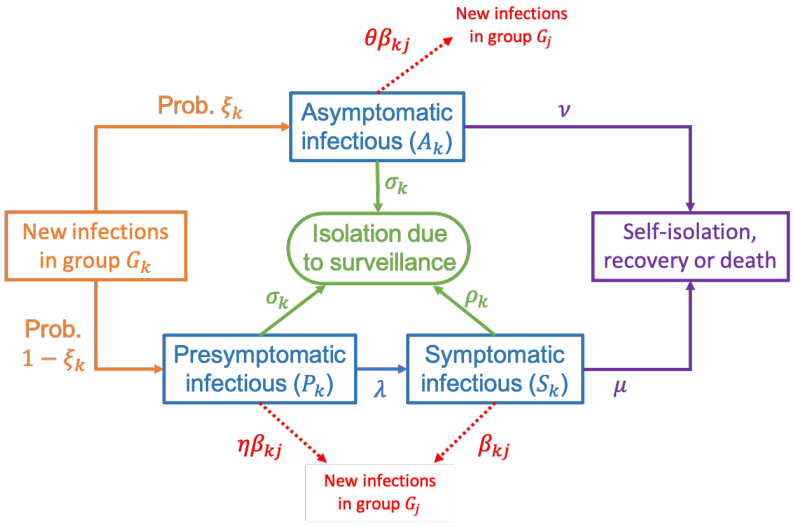


Figure 4.1: The branching process model used in our analyses. Schematic showing the different possible events in the branching process model and the rates at which they occur. The parameters of the model are described in the text and in Tables 4.1, 4.2, and 4.3.

4.2.2 Reproduction number

The effective reproduction number, R, represents the expected number of secondary infections generated by a single infected individual during their entire course of infection, accounting for interventions that are in place. Here, we take a heuristic approach to derive the following expression for R:

$$R = \sum_{k=1}^{16} \frac{N_k}{N} \sum_{j=1}^{16} \left[\frac{\xi_k \theta \beta_{kj}}{\nu + \sigma_k} + (1 - \xi_k) \left(\frac{\eta \beta_{kj}}{\lambda + \sigma_k} + \frac{\lambda}{\lambda + \sigma_k} \frac{\beta_{kj}}{\mu + \rho_k} \right) \right], \tag{4.1}$$

where $N = N_1 + \cdots + N_{16}$ is the total population size. To obtain this expression, we first consider the expected number of secondary infections an infected individual in age group G_k will generate in age group G_j. If an individual in age group G_k experiences a fully asymptomatic course of infection, which occurs with probability ξ_k, they will generate new infections in age group G_j at rate $\theta\beta_{kj}$ and recover or be isolated at rate $\nu + \sigma_k$. Therefore, the total number of infections they are expected to cause in age group G_j is $\theta\beta_{kj}/(\nu + \sigma_k)$. If instead the individual in age group G_k experiences a symptomatic course of infection, which occurs with probability $1 - \xi_k$, whilst presymptomatic they will generate new infections in age group G_j at rate $\eta\beta_{kj}$ and be isolated or develop symptoms at rate $\lambda + \sigma_k$. Thus, the total number of infections they are expected to cause in age group G_j whilst presymptomatic is $\eta\beta_{kj}/(\lambda + \sigma_k)$. If they go on to develop symptoms before being isolated, which occurs with probability $\lambda/(\lambda + \sigma_k)$, applying similar reasoning they are expected to

cause $\beta_{kj}/(\mu + \rho_k)$ new infections in age group G_j whilst symptomatic. Combining these possibilities leads to the term in square brackets in Equation (4.1), which is then summed over all possible age groups G_j of the infectee. Finally, to obtain the full expression given in Equation (4.1) we take a weighted average across all possible age groups G_k of the infector, where the weights N_k/N represent the proportions of the population belonging to each age group. This corresponds to the assumption that the initial infected host is more likely to belong to an age group containing more individuals than an age group with fewer individuals.

In the absence of interventions, i.e. when $\sigma_k = \rho_k = 0$ (representing no enhanced isolation as a result of surveillance) and β_{kj} is calculated using contact patterns that are characteristic of normal behaviour, the effective reproduction number, R, is equal to the basic reproduction number, R_0.

4.2.3 Model parameterisation

The numbers of individuals in each age group (values of N_k) were chosen according to United Nations age demographic data for the UK [210] (Fig 4.2A). The daily numbers of contacts between individuals in each age group (values of C_{kj}) were set according to the 16×16 'contact matrix' for the UK, in which the $(k, j)^{\text{th}}$ entry represents the expected daily number of unique contacts an individual in age group G_k has with individuals in age group G_j [168]. In addition to matrices representing 'all' contacts (Fig 4.2B), we also considered matrices detailing a breakdown into 'school', 'work', 'home', and 'other' contacts (Figs 4.2C-4.2F), allowing us to investigate the effects of control interventions that reduce contacts in each of these settings.

Since we considered SARS-CoV-2 as a case study, we once again used studies conducted during the COVID-19 pandemic to inform the epidemiological parameters of our model. Despite previous research assessing the relationships between age and factors such as susceptibility to SARS-CoV-2 infection or the propensity to develop symptoms, there is some variation in estimated parameters between different studies. To test the robustness of our results to this uncertainty, we conducted our analyses under three different scenarios (A, B, and C). In scenario A, we assumed that susceptibility to infection (values of ω_k) and the proportion of hosts who experience a fully asymptomatic course of infection (values of ξ_k) were independent of age. In scenario B, susceptibility was assumed to vary with age, but the proportion of asymptomatic infections was assumed independent of age. In scenario C, we allowed both susceptibility to infection and the asymptomatic proportion to vary with age. The values used for the parameters ω_k and ξ_k in each of these three scenarios are shown in Table 4.1 (see also [107]).

In all scenarios considered, the inherent infectivity was not assumed to be age-dependent

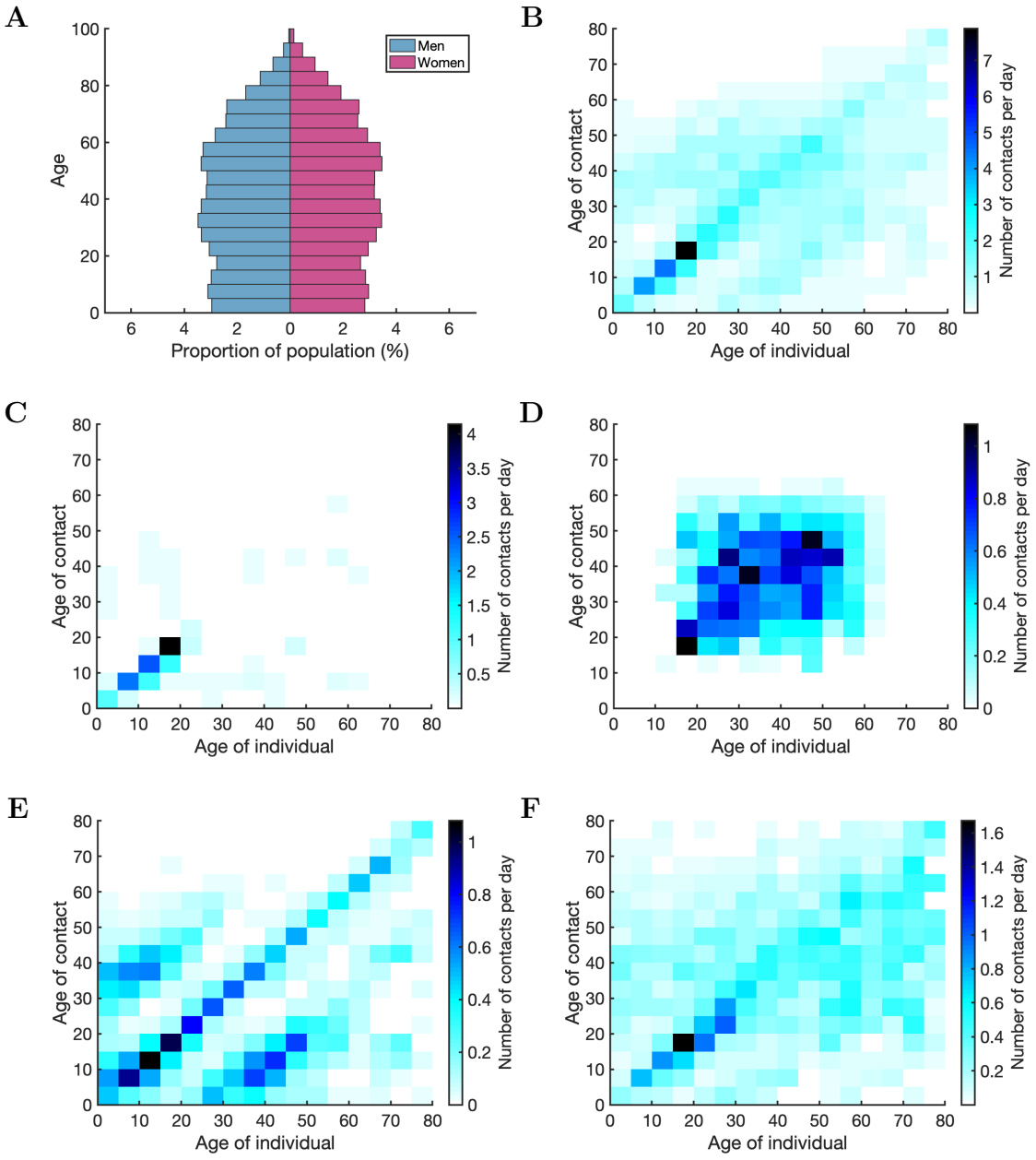


Figure 4.2: Age demographic and age-structured contact patterns for the United Kingdom. **A.** United Nations age demographic data for the UK in 2020, split into five-year age groups (where the final age group contains all ages 95+) [210]. **B.** Heat map of UK 'all contacts' matrix, representing the expected daily number of unique contacts that an individual in each age group G_k has with individuals in each other age group G_j [168]. **C.** The analogous figure to B, but showing only the subset of 'all' contacts that occur in schools ('school' contacts). **D.** The analogous figure to C, but showing only 'workplace' contacts. **E.** The analogous figure to C, but showing only 'home' contacts. **F.** The analogous figure to C, but showing only 'other' contacts (i.e. all contacts outside schools, workplaces, or homes).

(i.e. $\tau_k = 1$ for all values of k). In other words, the expected infectiousness of infected hosts in different age groups was governed solely by the proportion of asymptomatic infections

in that age group. We chose the scaling factors η and θ for the relative transmission rates from presymptomatic and asymptomatic individuals compared to symptomatic individuals so that the proportions of infections arising from each of these groups were in line with literature estimates (see Table 4.2 and [141]).

In the absence of enhanced isolation, we set the expected duration of the presymptomatic infectious period and the time for which symptomatic individuals are able to infect others to be $1/\lambda = 2$ days and $1/\mu = 8$ days, respectively [97, 144–146]. The asymptomatic infectious period was then chosen so that all infected individuals are expected to be infectious for the same period (i.e. $1/\nu = 10$ days). In our initial analysis, we set the isolation rates ρ_k and σ_k equal to zero for all k; later, we considered the effects of increasing these rates.

Initially, we fixed $R_0 = 3$ (in line with initial estimates of SARS-CoV-2 transmissibility [134–137], before the emergence of more transmissible variants) and used Equation (4.1) to determine the appropriate corresponding value of the scaling factor B. Later, when considering the impact of NPIs on the probability of a major outbreak, we retained this value of B, allowing the reproduction number to change as a result of the control implemented.

Age group (G_k)	**Relative susceptibility (ω_k) in each scenario**		**Asymptomatic proportion (ξ_k) in each scenario**	
	A	**B & C**	**A & B**	**C**
G_1, G_2 (0-9)	1.0	0.4	0.584 (weighted average of age-dependent values of ξ_k in scenario C)	0.71
G_3, G_4 (10-19)		0.38		0.79
G_5, G_6 (20-29)		0.79		0.73
G_7, G_8 (30-39)		0.86		0.67
G_9, G_{10} (40-49)		0.8		0.6
G_{11}, G_{12} (50-59)		0.82		0.51
G_{13}, G_{14} (60-69)		0.88		0.37
G_{15}, G_{16} (70+)		0.74		0.31

Table 4.1: Baseline values of age-dependent parameters. Values used for the age-dependent relative susceptibility to infection (ω_k) and the proportion of infections that are asymptomatic (ξ_k) for each of the scenarios A, B, and C [107].

4.2.4 Probability of a major outbreak

The probability that an infected individual in a particular age group initiates a major outbreak when they are introduced to the population was calculated using the branching process model. We derived a nonlinear system of simultaneous equations that determined the age-dependent outbreak probabilities, as described below. The major outbreak probabilities were then obtained by solving these equations numerically.

Param.	Meaning	Value in each scenario			Justification
		A	B	C	
B	Transmission rate scaling factor	0.0386	0.0559	0.0607	Chosen so that $R_0 = 3$ [134–137]
η	Relative transmission rate of presymptomatic hosts compared to symptomatic hosts	4.83	4.83	4.83	Chosen so that the proportion of all infections arising from presymptomatic hosts is 0.489 [141]
θ	Relative transmission rate of asymptomatic hosts compared to symptomatic hosts	0.149	0.149	0.130	Chosen so that the proportion of all infections arising from entirely asymptomatic hosts is 0.106 [141]

Table 4.2: Baseline values of scenario-dependent scaling parameters. Values used for the scaling parameters B, η, and θ for each of the scenarios A, B, and C.

Firstly, we considered a scenario in which the initial infected host in the population is a symptomatic individual in age group G_k. In that case, the possibilities for the next event to occur in the branching process model are as follows:

1. The infected individual in G_k infects a susceptible individual in G_j, so that either A_j increases by one (with probability ξ_j) or P_j increases by one (with probability $(1-\xi_j)$. This occurs with probability

$$\alpha_{kj} = \frac{\beta_{kj}}{\mu + \rho_k + \sum_{l=1}^{16} \beta_{kl}}.$$

2. The infected individual in G_k recovers, dies, or is isolated before infecting anyone else, so that S_k decreases to zero (and there are no infected individuals left in the population). This occurs with probability

$$\gamma_k = \frac{\mu + \rho_k}{\mu + \rho_k + \sum_{l=1}^{16} \beta_{kl}}.$$

We denote the probability of a major outbreak not occurring starting from a single symptomatic, presymptomatic, or asymptomatic infectious individual in age group G_k by x_k, y_k, and z_k respectively. Then, assuming that chains of transmission arising from infectious individuals are independent, and noting that if there are no infectious hosts present in the population then a major outbreak will certainly not occur, the probability x_k that no major outbreak occurs beginning from a single symptomatic individual in G_k satisfies the equation

Param.	Meaning	Baseline value	Justification
R_0	Expected number of secondary infections generated by a single infectious host in the absence of interventions	$R_0 = 3$	Within estimated range for wild-type SARS-CoV-2 [134–137]
λ	Rate at which presymptomatic hosts develop symptoms	$\lambda = 1/2$ days^{-1}	[97]
μ	Rate at which symptomatic hosts are removed due to self-isolation, recovery or death	$\mu = 1/8$ days^{-1}	[144–146]
ν	Rate at which asymptomatic hosts are removed due to recovery or death	$\nu = 1/10$ days^{-1}	Chosen so that, in the absence of interventions, the expected duration of infection is identical for all infected hosts ($1/\nu = 1/\lambda + 1/\mu$)
τ_k	(Relative) infectivity parameter	$\tau_k = 1$ for $k = 1, \ldots, 16$	Assumed
ρ_k	Isolation rate due to surveillance of symptomatic individuals	$\rho_k = 0$ for $k = 1, \ldots, 16$	N/A
σ_k	Isolation rate due to surveillance of nonsymptomatic individuals	$\sigma_k = 0$ for $k = 1, \ldots, 16$	N/A

Table 4.3: Baseline values of scenario-independent parameters. Values used for the parameters that were assumed not to vary between scenarios A, B, and C. We also considered strategies involving enhanced surveillance ($\rho_k, \sigma_k > 0$) – see Fig 4.6.

$$x_k = x_k\left[\sum_{j=1}^{16} \alpha_{kj}\left(\xi_j z_j + (1-\xi_j)y_j\right)\right] + \gamma_k, \tag{4.2}$$

where α_{kj} and γ_k are as above.

Similarly, beginning instead with a single presymptomatic individual in G_k, the possibilities for the next event are:

1. The presymptomatic infected individual in G_k infects a susceptible individual in G_j,

so that (as before) either A_j increases by one (with probability ξ_j) or P_j increases by one (with probability $(1-\xi_j)$). This occurs with probability

$$\tilde{\alpha}_{kj} = \frac{\eta\beta_{kj}}{\lambda + \sigma_k + \eta\sum_{l=1}^{16}\beta_{kl}}.$$

2. The infected individual in G_k develops symptoms (transitions from P_k to S_k). This occurs with probability

$$\delta_k = \frac{\lambda}{\lambda + \sigma_k + \eta\sum_{l=1}^{16}\beta_{kl}}.$$

3. The infected individual in G_k is isolated before infecting anyone else, so that S_k decreases by one. This occurs with probability

$$\phi_k = \frac{\sigma_k}{\lambda + \sigma_k + \eta\sum_{l=1}^{16}\beta_{kl}}.$$

Therefore, the probability y_k that no major outbreak occurs beginning from a single presymptomatic individual in G_k satisfies the equation

$$y_k = y_k\left[\sum_{j=1}^{16}\tilde{\alpha}_{kj}\left(\xi_j z_j + (1-\xi_j)y_j\right)\right] + \delta_k x_k + \phi_k. \tag{4.3}$$

Similarly, the probability z_k that a major outbreak does not occur starting from a single asymptomatic individual in G_k satisfies the equation

$$z_k = z_k\left[\sum_{j=1}^{16}\hat{\alpha}_{kj}\left(\xi_j z_j + (1-\xi_j)y_j\right)\right] + \varepsilon_k, \tag{4.4}$$

where

$$\hat{\alpha}_{kj} = \frac{\theta\beta_{kj}}{\nu + \sigma_k + \theta\sum_{l=1}^{16}\beta_{kl}} \quad \text{and} \quad \varepsilon_k = \frac{\nu + \sigma_k}{\nu + \sigma_k + \theta\sum_{l=1}^{16}\beta_{kl}}.$$

The system of simultaneous equations (4.2), (4.3), and (4.4) can be solved numerically to obtain x_k, y_k, and z_k (here, we did this using the MATLAB nonlinear system solver '*fsolve*'). Specifically, we take the minimal non-negative solution, as is standard when calculating extinction probabilities using branching process models [65, 76]. Then, for each k, the probability of a major outbreak occurring beginning from a single symptomatic, presymptomatic, or asymptomatic individual in group G_k is given by $1-x_k$, $1-y_k$, and $1-z_k$, respectively.

Throughout, we consider the probability p_k of a major outbreak occurring beginning from a single nonsymptomatic individual in group G_k arriving in the population at the beginning of their infection:

$$p_k = \xi_k(1-z_k) + (1-\xi_k)(1-y_k).$$

The average major outbreak probability, P, which is defined as the probability of a major outbreak when the index case is chosen randomly from the population, is also considered. The value of P is a weighted average of the p_k values, where the weights correspond to the proportion of the population represented by each group:

$$P = \frac{1}{N} \sum_{k=1}^{16} N_k p_k.$$

This reflects an assumption that the index case is more likely to come from an age group with more individuals than an age group with fewer individuals.

4.2.5 Code availability

All computing code used to implement the above methods was written in MATLAB version R2019a, and is available at https://github.com/francescalovellread/age-dependent-outbreak-risks.

4.3 Results

4.3.1 Effect of the age of the index case on the risk of a major outbreak

We first considered the probability that a single infected individual in a particular age group G_k initiates a major outbreak when introduced to a new population. This quantity was calculated for each of the three scenarios A, B, and C (Fig 4.3).

In scenario A, the variation in the major outbreak risk for introduced cases of different ages is driven solely by the numbers of contacts between individuals. As a result, due to their higher numbers of daily contacts, school- and working-age individuals are more likely to trigger a major outbreak than children under five or adults over 60, with index cases aged 15-19 posing the highest risk (Fig 4.3A). These findings do not change substantially when susceptibility is allowed to vary with age in scenario B (Fig 4.3B). However, in scenario C, assuming that the clinical fraction also varies between age groups alters the age-dependent risk profile substantially. This is because asymptomatic individuals are assumed to be less infectious than symptomatic individuals, and therefore an index case in an age group with a high proportion of asymptomatic infections is less likely to initiate a major outbreak. In this scenario, index cases aged 40 or over had a disproportionately high probability of generating a major outbreak, with individuals aged 70-74 presenting the highest risk (Fig 4.3C). These individuals are more likely to develop symptoms than younger individuals (Table 4.1), leading to a higher expected infectiousness. In contrast, individuals under the age of 40 had a below average probability of generating a major outbreak, with individuals aged 10-14 presenting the lowest risk. Noticeably, individuals aged 5-19 presented relatively

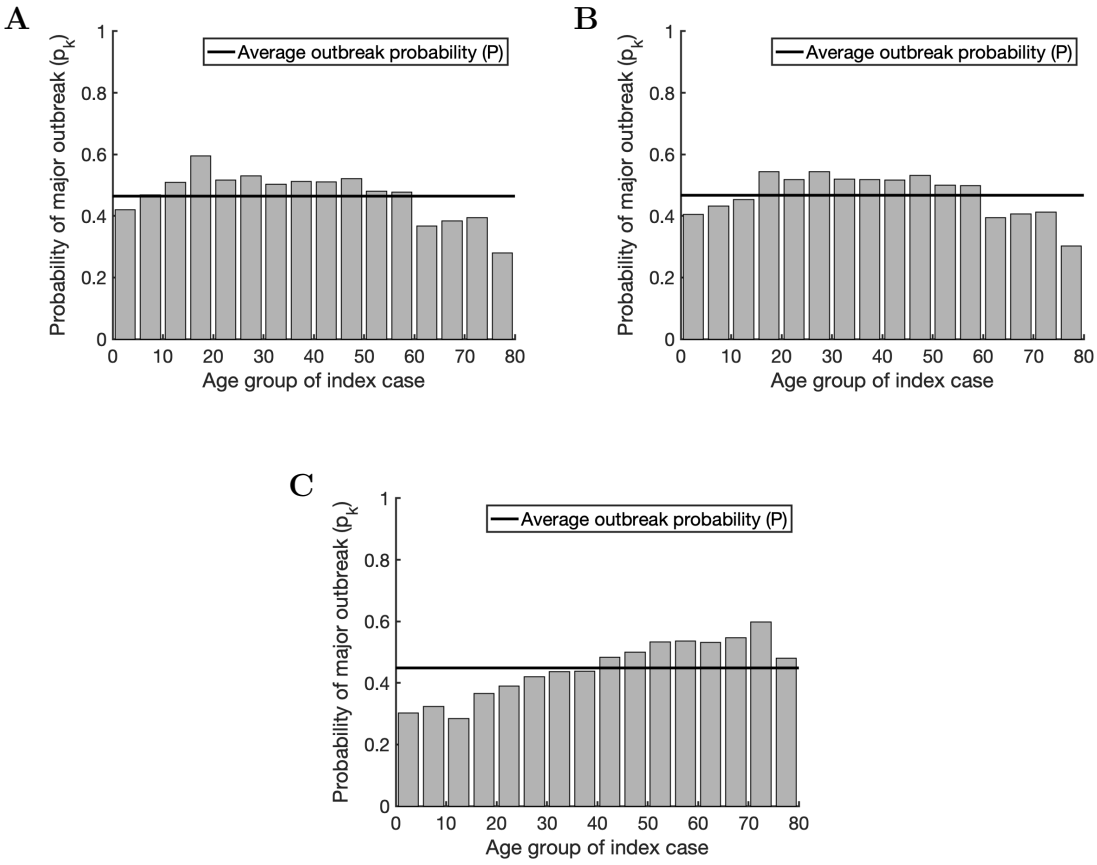


Figure 4.3: The probability of a major outbreak depends on the age of the index case. **A.** The probability that a single infected individual in any given age group triggers a major outbreak (grey bars) for scenario A, in which susceptibility and clinical fraction are assumed constant across all age groups. The weighted average major outbreak probability P is shown by the black horizontal line. **B.** The analogous figure to A but for scenario B, in which clinical fraction is assumed constant across all age groups but susceptibility varies with age (see Table 4.1). **C.** The analogous figure to A but for scenario C, in which both susceptibility and clinical fraction vary with age (see Table 4.1).

low risks, despite the high numbers of contacts occurring among these age groups (Fig 4.2B). In this scenario, the large number of contacts was offset by the fact that individuals in these age groups are more likely to be asymptomatic and consequently less infectious than older individuals (Table 4.1). Therefore, an index case in one of these age groups is likely to lead to fewer secondary transmissions. Furthermore, the contact patterns between individuals in these age groups are highly assortative with respect to age (Fig 4.2B). Therefore, in addition to the index case being less infectious, a high proportion of the contacts they make are with individuals who are also likely to be less infectious, as well as being less susceptible to infection in the first place.

We performed our subsequent analyses for each of the three scenarios A, B, and C, with qualitatively similar results. The figures shown in the main text are for scenario C, since we

deem this scenario to be the most realistic for SARS-CoV-2 transmission, but the analogous results for scenarios A and B are presented in Supplementary Figs B.1-B.6.

4.3.2 Effect of the target age group on NPI effectiveness

We next considered the effects of NPIs that reduce the number of contacts between individuals on the probability that an introduced case will lead to a major outbreak. To approximate the relative effects of school closures, workplace closures, and population-wide social distancing policies, we calculated the age-dependent risk profiles when each of these types of contact were excluded from the overall contact matrix.

First, we removed all 'school' contacts from the total contact matrix (Fig 4.4A). For scenario C, removing 'school' contacts led to a 4.2% reduction in the average probability of a major outbreak (from 0.449 to 0.430). This small reduction is unsurprising for scenario C, since in that scenario school-aged infected individuals are assumed more likely to be asymptomatic than other infected individuals, and therefore their expected infectiousness is lower. However, even for scenarios A and B, in which school-aged individuals present the greatest risk of triggering a major outbreak, the effectiveness of removing 'school' contacts alone at reducing the major outbreak probability was limited (reductions of 7.2% and 4.75%, respectively; see Supplementary Figs B.1A, B.2A). In each scenario, the reduction in risk was predominantly for school-aged index cases, with the risk from index cases of other ages only slightly reduced. Second, we considered the effects of removing 'work' contacts from the total contact matrix (Fig 4.4B). This led to a more substantial 25.4% reduction in the average probability of a major outbreak for scenario C (with corresponding reductions of 19.0% and 24.0% for scenarios A and B, respectively; see Supplementary Figs B.1B, B.2B). As well as reducing the risk of a major outbreak from an index case of working age, removing 'work' contacts also reduced the probability of a major outbreak occurring starting from a school-aged individual. This is because closing workplaces helps to block chains of transmission that begin with an infected child. For example, a transmission chain involving a child transmitting to an adult at home, followed by subsequent spread around the adult's workplace, will be less likely to occur. Third, we investigated the effect of removing all 'other' contacts, reflecting perfect social distancing being observed outside of the home, school, or workplace (Fig 4.4C). This had the most significant effect of the three types of contact-reducing intervention considered, reducing the probability of a major outbreak by 41.7% for scenario C (and 30.7% or 33.2% for scenarios A and B, respectively; see Supplementary Figs B.1C, B.2C).

In the three cases described above, we considered complete reductions in 'school', 'work', and 'other' contacts, respectively. In practice, such complete elimination of contacts is unlikely. We therefore also considered partial reductions in 'school', 'work', and 'other' contacts, and compared the resulting reductions in the major outbreak probability (Fig 4.4D). For any

given percentage reduction in contacts, reducing 'other' contacts always led to the largest reduction in the major outbreak probability (see also Supplementary Figs B.1D, B.2D). This suggests that reducing social contacts outside schools and workplaces can be an important component of strategies to reduce the risk of major outbreaks of SARS-CoV-2. However, this alone is not enough to eliminate the risk of major outbreaks entirely. For greater risk reductions using contact-reducing NPIs, a mixed approach involving combinations of reductions in 'school', 'workplace', and 'other' contacts is needed.

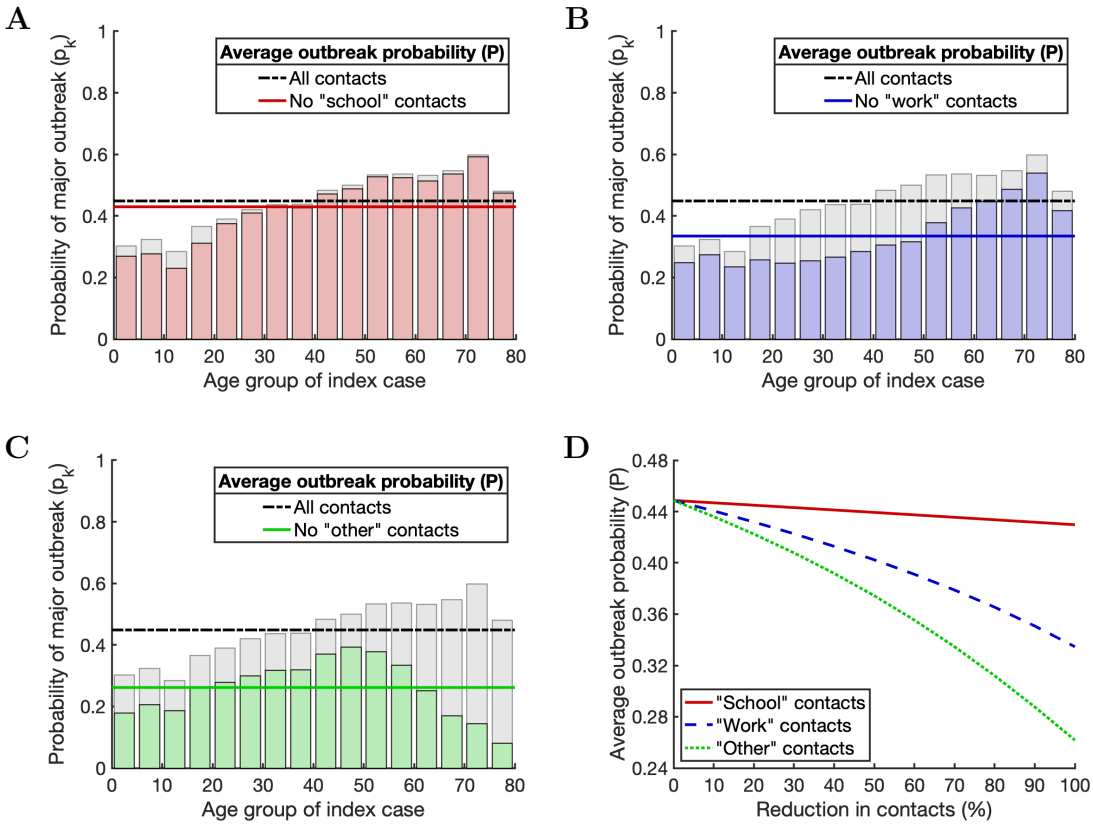


Figure 4.4: The effects of interventions that reduce contacts between individuals on the probability of a major outbreak. A. The effect of removing all 'school' contacts on the probability of a major outbreak. Pale grey bars and black dash-dotted line represent the major outbreak probabilities without any contacts removed (as in Fig 4.3C). Red bars and the solid red line represent the major outbreak probabilities and their weighted average when 'school' contacts are removed. **B.** The analogous figure to A, but with all 'work' contacts removed. **C.** The analogous figure to A, but with all 'other' contacts removed. **D.** Partial reductions in 'school', 'work', and 'other' contacts, and the resulting reductions in the average major outbreak probability (solid red, dashed blue, and dotted green lines, respectively).

4.3.3 Mixed strategies for reducing the major outbreak risk

Next, we considered the effects of combining reductions in 'school', 'work', and 'other' contacts on the major outbreak probability (Fig 4.5; analogous results for scenarios A and B are shown in Supplementary Figs B.3 and B.4). We allowed reductions in 'school' and 'work' contacts to vary between 0% and 100% whilst 'other' contacts were reduced by 25%, 50%, or 75% (Figs 4.5A, 4.5B, and 4.5C, respectively).

Since NPIs have negative economic, social, and non-disease health consequences, policy makers may choose to implement public health measures in which the risk of major outbreaks is not eliminated completely. These results provide contact reduction targets for mixed strategies in which the major outbreak probability is reduced to a pre-specified 'acceptable' level. For example, to reduce the major outbreak probability to 0.25, 'other' contacts could be reduced by 25% from the baseline level, and 'school' and 'home' contacts reduced as indicated by the red dotted contour marked '0.25' in Fig 4.5A. Alternatively, to achieve the same major outbreak risk, 'other' contacts could instead be reduced by 50% or by 75%, with the degree of 'school' and 'work' reductions chosen according to the contours marked '0.25' in Figs 4.5B and 4.5C, respectively.

If a policy maker wishes to eliminate the major outbreak risk entirely using contact-reducing NPIs, for the model parameterisation considered very significant reductions in multiple types of contacts are needed in combination. For example, even if all 'school' and 'work' contacts are removed, 'other' contacts must be reduced by 66% for the overall average major outbreak probability to fall below 0.01 (Fig 4.5D). Since such substantial reductions in multiple types of contacts are unlikely to be possible, this suggests that contact-reducing NPIs must be combined with other interventions, such as effective surveillance and isolation strategies, to eliminate major outbreaks.

4.3.4 Effect of surveillance on the contact-reducing NPIs required for major outbreak control

We next considered whether or not low major outbreak probabilities can be achieved using limited contact-reducing NPIs in combination with other interventions. Specifically, the effects of surveillance and isolation of infected individuals (through e.g. contact tracing) as well as reducing contacts in schools, workplaces, and other locations were assessed. While results are shown for scenario C in Fig 4.6, analogous results for scenarios A and B are presented in Supplementary Figs B.5 and B.6.

Initially, we considered the effect of increasing the rate at which symptomatic and/or non-symptomatic infected individuals are detected and isolated as a result of surveillance, in

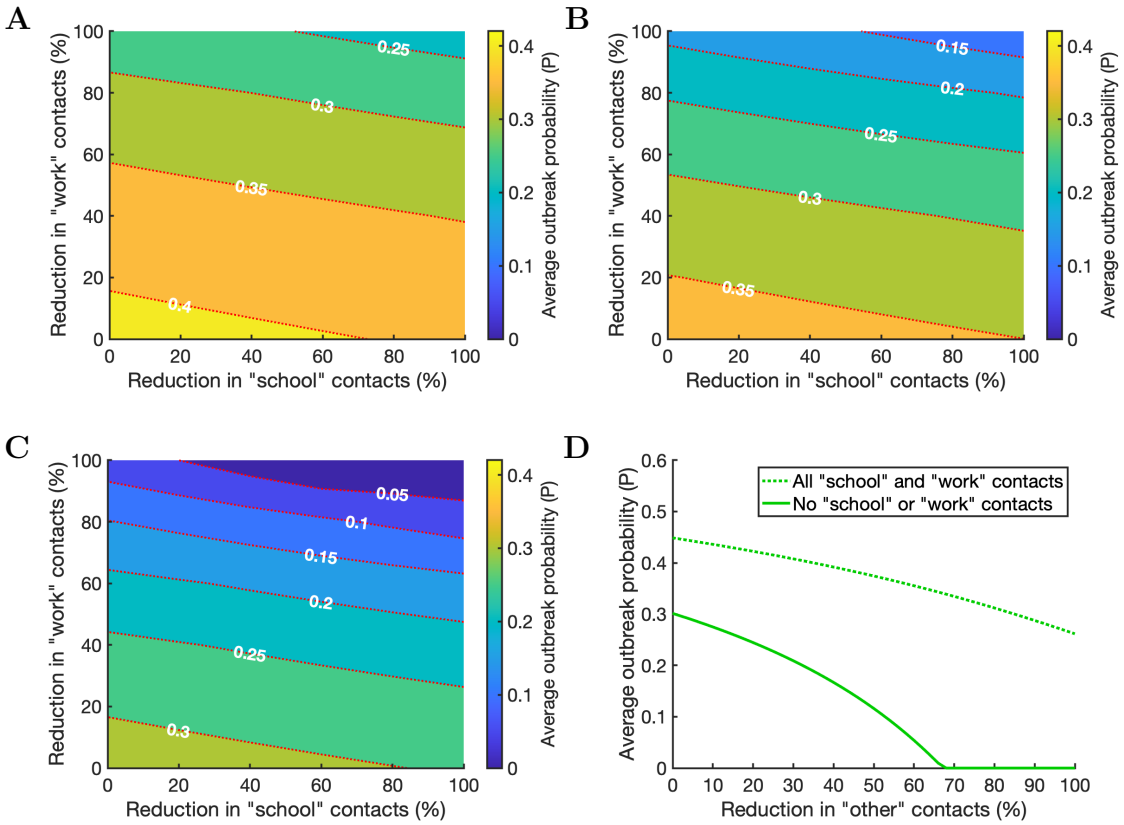


Figure 4.5: The effects of intervention strategies that combine reductions in 'school', 'work', and 'other' contacts. A. The effect of reducing 'school' and 'work' contacts on the weighted average probability of a major outbreak (P), when 'other' contacts are reduced by 25% across all age groups. Red dotted lines indicate contours along which the major outbreak probability is constant. **B.** The analogous figure to A, but with a 50% reduction in 'other' contacts. **C.** The analogous figure to A, but with a 75% reduction in 'other' contacts. **D.** The effect of reducing 'other' contacts on the average major outbreak probability when 'school' and 'work' contacts are not reduced at all (dotted line), and when 'school' and 'work' contacts are reduced by 100% (solid line).

the absence of contact-reducing NPIs (i.e. with no reduction in the number of contacts between individuals compared to the baseline case in Fig 4.2B) (Figs 4.6A, 4.6B). For symptomatic hosts, this represents an enhanced rate of isolation compared to the baseline rate of self-isolation already present in the model. Isolation of nonsymptomatic hosts was more effective at reducing the major outbreak probability than isolation of symptomatic hosts (Figs 4.6A, 4.6B), although of course this is more challenging to achieve [65]. However, if fast isolation of nonsymptomatic hosts could be achieved through efficient large-scale testing (potentially in combination with contact tracing [211]), the probability of major outbreaks could be reduced substantially through this measure alone.

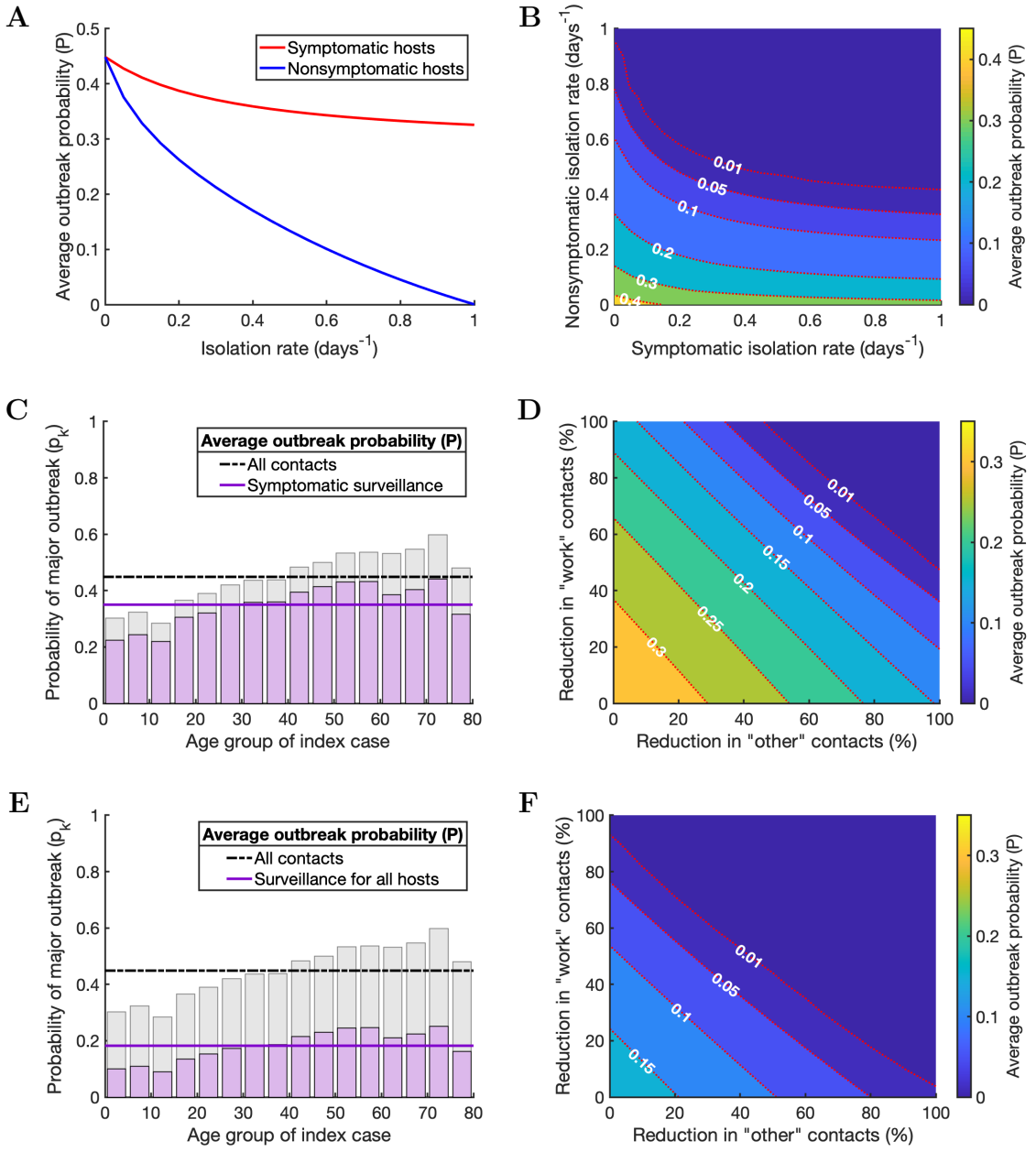


Figure 4.6: Surveillance as part of a mixed strategy to reduce the major outbreak probability. **A.** The effect of increasing the isolation rate of symptomatic (red line) or nonsymptomatic (blue line) infected hosts on the average probability of a major outbreak (P), in the absence of contact-reducing NPIs. The isolation rates ρ_k and σ_k are varied in turn between 0 days^{-1} and 1 days^{-1}. **B.** The effect of simultaneously varying the isolation rate of symptomatic and nonsymptomatic hosts on the average probability of a major outbreak (P), again without contact-reducing NPIs. **C.** The age-dependent probability of a major outbreak when the isolation rate of symptomatic individuals is $\rho_k = 1/2$ days^{-1}, without contact-reducing NPIs or surveillance of nonsymptomatic infected individuals (purple bars and solid line). Pale grey bars and black dash-dotted line represent the original major outbreak probabilities without any contact-reducing NPIs or enhanced surveillance (as in Fig 4.3C). **D.** The effect of reducing 'work' and 'other' contacts when the isolation rate of symptomatic infected individuals is $\rho_k = 1/2$ days^{-1}, as in panel C, without surveillance of nonsymptomatic infected individuals. **E,F.** The analogous figures to panels C and D, with enhanced surveillance of both symptomatic and nonsymptomatic infected hosts ($\rho_k = 1/2$ days^{-1} and $\sigma_k = 1/7$ days^{-1}).

We then demonstrated the effects of combining contact-reducing NPIs with enhanced isolation of infected hosts due to surveillance. First, we increased the enhanced isolation rate of symptomatic individuals to $\rho_k = 1/2$ days^{-1}. In the absence of other interventions, this reduced the major outbreak probability by 22.0% (Fig 4.6C). With this level of surveillance, the major outbreak risk could be reduced below 0.01 with a reduction in 'work' and 'other' contacts of around 73% each, for example (Fig 4.6D).

Finally, keeping the enhanced isolation rate of symptomatic individuals equal to $\rho_k = 1/2$ days^{-1}, we increased the isolation rate of nonsymptomatic individuals to $\sigma_k = 1/7$ days^{-1}. In this case, the major outbreak probability without contact-reducing NPIs fell by 59.4% compared to a situation without enhanced surveillance (Fig 4.6E), and the reductions in 'work' and 'other' contacts needed to bring the major outbreak probability below 0.01 were significantly smaller (Fig 4.6F). For example, if 'work' contacts can be reduced by 50%, then 'other' contacts only need to be reduced by 43%. This indicates that effective surveillance of both symptomatic and nonsymptomatic individuals can substantially lower the extent of contact-reducing NPIs that are required to achieve substantial reductions in major outbreak risks.

4.4 Discussion

During the COVID-19 pandemic, public health measures that reduce the numbers of contacts between individuals have been implemented in countries globally. These measures include school closures, workplace closures, and population-wide social distancing policies. Contact-reducing NPIs have been shown to be effective at reducing SARS-CoV-2 transmission, and have also been used previously during influenza pandemics [178, 212–214]. However, long-term implementation of these measures has negative social, psychological, and economic consequences [180, 184–188]. It is therefore important to assess the effectiveness of different contact-reducing NPIs at lowering transmission and preventing major outbreaks, in order to design effective targeted control strategies that avoid unnecessarily strict measures.

In this chapter, we constructed a branching process model to estimate the risk of major outbreaks under different contact-reducing NPIs and different levels of surveillance for symptomatic and nonsymptomatic infected individuals. Unlike previous approaches for estimating outbreak risks using branching processes [64, 65, 87, 147, 208, 209], we considered the effects of age-related heterogeneities affecting transmission for infected individuals of a wide range of ages, including age-dependent variations in social mixing patterns, susceptibility to infection, and clinical fraction. Using SARS-CoV-2 as a case study, we demonstrated that the risk that an introduced case initiates a major outbreak depends on these age-related factors and on the age of the introduced case (Fig 4.3), as well as the age structure of the local population.

We used our model to assess the effects of reducing the numbers of contacts that occur in school, in the workplace, and elsewhere. Of the three contact-reducing NPIs considered, removing 'school' contacts had the smallest effect on the probability of observing a major outbreak, even when age-dependent variations in susceptibility and clinical fraction were ignored (Figs 4.4A, 4.4D, and Supplementary Figs B.1A, B.1D, B.2A, B.2D). This can be attributed to the fact that school closures predominantly reduce contacts between individuals aged 5-19, who only account for approximately 23% of the total population [210]. Additionally, these individuals tend to have large numbers of contacts outside of the school environment (Figs 4.2E, 4.2F). Therefore, compared to other measures, interrupting within-school transmission may have only a limited effect on transmission in the wider population, particularly when school-aged individuals are less susceptible to infection and more likely to experience subclinical courses of infection. In contrast, reducing contacts that occur outside schools or workplaces was the most effective intervention, substantially lowering the major outbreak risk across all age groups, and for those aged over 60 in particular (Figs 4.4C, 4.4D, and Supplementary Figs B.1C, B.1D, B.2C, B.2D). This could explain the success of social distancing strategies worldwide for reducing observed COVID-19 cases and deaths.

Mixed strategies combining reductions in 'school', 'work', and 'other' contacts led to greater reductions in the major outbreak probability than individual interventions (Figs 4.5A, 4.5B, 4.5C), but very large reductions in all three types of contact were required to eliminate the risk of major outbreaks entirely (Fig 4.5D). However, implementing effective surveillance to identify infected hosts (followed by isolation) led to substantial reductions in the risk of major outbreaks even in the absence of other control measures (Figs 4.6A, 4.6B). In the scenarios considered here, with an efficient surveillance strategy in place, significantly smaller reductions in 'work' and 'other' contacts were needed to render the major outbreak probability negligible, even when 'school' contacts were not reduced at all (Figs 4.6C, 4.6D, 4.6E, 4.6F). This supports the use of surveillance that targets both symptomatic and non-symptomatic individuals, such as contact tracing and isolation strategies or population-wide diagnostic testing, to help prevent major outbreaks [65].

Our model could also be used to explore the effects of surveillance targeted at specific age groups. Since individuals of different ages contribute to transmission by varying degrees, a surveillance strategy in which resources are preferentially allocated to age groups that have a high risk of triggering an outbreak may be beneficial. Additionally, it would be interesting to consider how the relative effectiveness of surveillance for symptomatic and nonsymptomatic infected hosts varies with age. Nonsymptomatic surveillance may be much more important in age groups for which the likelihood of developing clinical symtoms is low. By allowing the isolation rate parameters ρ_k and σ_k to vary with the age group index k, our

model could be used to explore the effects of varying the resource allocation for symptomatic and nonsymptomatic surveillance between the different age groups G_k in the population.

In this study, we used age demographic and contact data for the UK, but equivalent data for other countries are available and could easily be substituted into our model to estimate outbreak risks elsewhere [168, 210]. An interesting extension to our work would be to explore how our results vary when our model is applied to countries that have very different age demographics or patterns of social contacts compared to the UK.

Although here we used SARS-CoV-2 as a case study, our model provides a framework for estimating the risk of major outbreaks in age-structured populations that can be adapted for other pathogens, provided sufficient data are available to parameterise the model appropriately. The effects of age-structure on major outbreak risks may vary for pathogens with different epidemiological characteristics. For influenza-A viruses, for example, susceptibility to infection tends to decrease with age, whilst the risk of an infection leading to severe symptoms is greater both for the elderly and for the very young [107, 215, 216]. This is in contrast to SARS-CoV-2, for which children are more likely to experience subclinical courses of infection.

One caveat of the results for SARS-CoV-2 presented here is that, although the epidemiological parameters of our model were chosen to be consistent with reported literature estimates, there is considerable variation between studies. In particular, the precise age-dependent variation in susceptibility and clinical fraction remains unclear, and the relative infectiousness of asymptomatic, presymptomatic, and symptomatic hosts has not been determined exactly. Furthermore, the inherent transmissibility of SARS-CoV-2 changed as the pandemic progressed, due to the appearance of novel variants of the virus. To explore ongoing major outbreak risks due to SARS-CoV-2, it would be necessary to update the model to reflect the transmission characteristics of those variants. Due to the uncertainty in model parameter values, we conducted sensitivity analyses to explore the effects of varying the parameters of the model on our results (Supplementary Figs B.1-B.12). In each case that we considered, our main conclusions were unchanged: the probability that an introduced case initiates a major outbreak depends on age-dependent factors affecting pathogen transmission and control, with widespread interventions and combinations of NPIs reducing the risk of major outbreaks most significantly.

An important limitation of our approach to modelling contact-reducing NPIs is that we made a standard assumption in our main analyses that 'school', 'work', and 'other' contacts are independent [46, 106, 107]. In other words, reducing the numbers of contacts in one location did not affect the numbers of contacts occurring in another. In reality, this is unlikely to be

the case. For example, closing schools also affects workplace contacts, as adults may then work from home in order to fulfil childcare requirements. Additionally, the contact data that we used represent the number of unique contacts per day and do not reflect repeated contacts with the same person, which affect the risk of transmission between individuals. These assumptions could in principle be removed, if relevant data become available – for example, data describing the effects of school closures on numbers of contacts in other locations. To demonstrate how changes in multiple types of contact due to NPIs could be implemented in the model, we conducted a supplementary analysis in which we investigated the effects of removing all 'school' contacts and allowing for concurrent changes in 'work', 'home', and 'other' contacts (Supplementary Figs B.13, B.14). These results support our conclusion that school closures are unlikely to have a substantial impact on SARS-CoV-2 transmission when applied as the sole NPI. The benefit of school closures could even potentially be outweighed by the possible secondary effects on other types of contacts. An improved understanding of how NPIs affect different types of contact is important for more accurate assessments of interventions in future.

Despite the simplifications made, our model provides a useful framework for estimating the risk of major outbreaks and the effects of NPIs. Different measures can be considered in combination in the model to develop strategies for lowering major outbreak risks. Our results emphasise the importance of quantifying age-dependent factors that affect transmission dynamics, such as susceptibility to infection and the proportion of hosts who develop clinical symptoms, for individuals of different ages. As we have shown, it is crucial to take age-dependent factors into account when assessing major outbreak risks and designing public health measures.

Appendix B

Supplementary material for Chapter 4

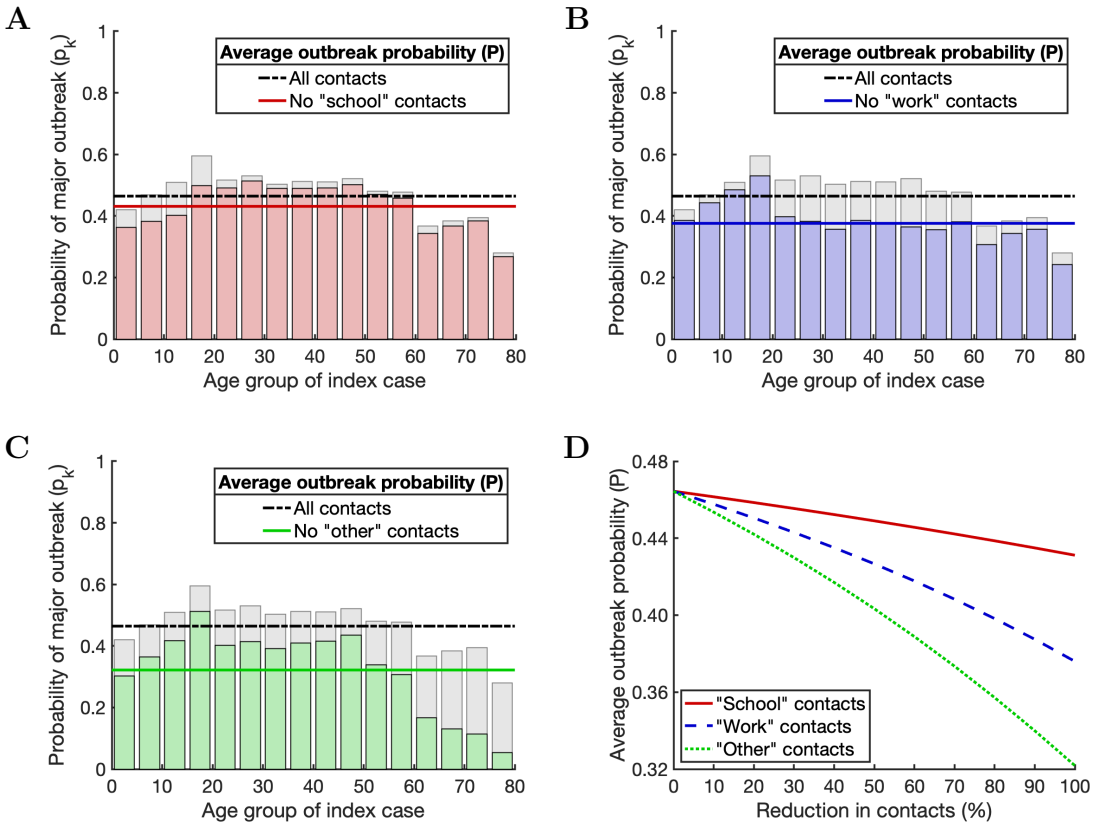


Figure B.1: Scenario A: The effects of interventions that reduce contacts between individuals on the probability of a major outbreak. For scenario A, susceptibility to infection and the proportion of hosts who experience a fully asymptomatic course of infection are independent of age. **A.** The effect of removing all 'school' contacts on the probability of a major outbreak. Pale grey bars and black dash-dotted line represent the major outbreak probabilities without any contacts removed (as in Fig 4.3A). Red bars and the solid red line represent the major outbreak probabilities and their weighted average when 'school' contacts are removed. **B.** The analogous figure to A, but with all 'work' contacts removed. **C.** The analogous figure to A, but with all 'other' contacts removed. **D.** Partial reductions in 'school', 'work', and 'other' contacts, and the resulting reductions in average major outbreak probability (solid red, dashed blue, and dotted green lines respectively).

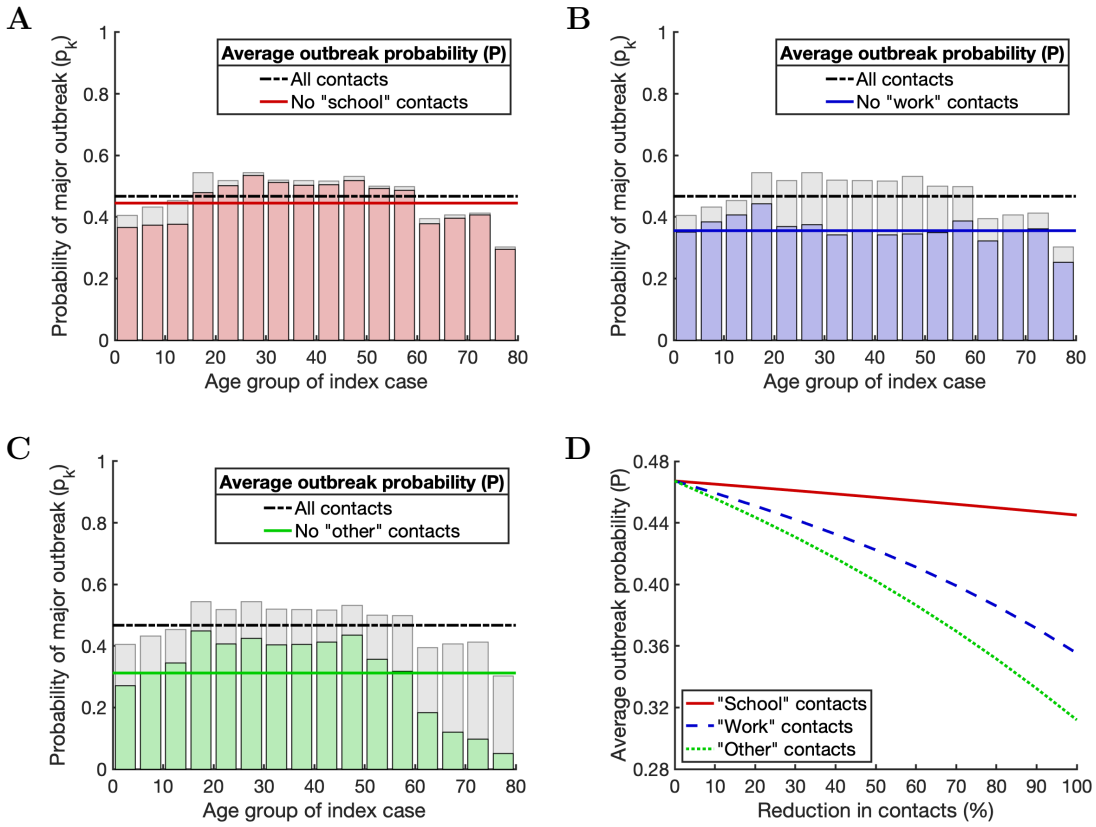


Figure B.2: Scenario B: The effects of interventions that reduce contacts between individuals on the probability of a major outbreak. For scenario B, susceptibility to infection varies with age but the proportion of hosts who experience a fully asymptomatic course of infection is independent of age. **A.** The effect of removing all 'school' contacts on the probability of a major outbreak. Pale grey bars and black dash-dotted line represent the major outbreak probabilities without any contacts removed (as in Fig 4.3B). Red bars and the solid red line represent the major outbreak probabilities and their weighted average when 'school' contacts are removed. **B.** The analogous figure to A, but with all 'work' contacts removed. **C.** The analogous figure to A, but with all 'other' contacts removed. **D.** Partial reductions in 'school', 'work', and 'other' contacts, and the resulting reductions in average major outbreak probability (solid red, dashed blue, and dotted green lines respectively).

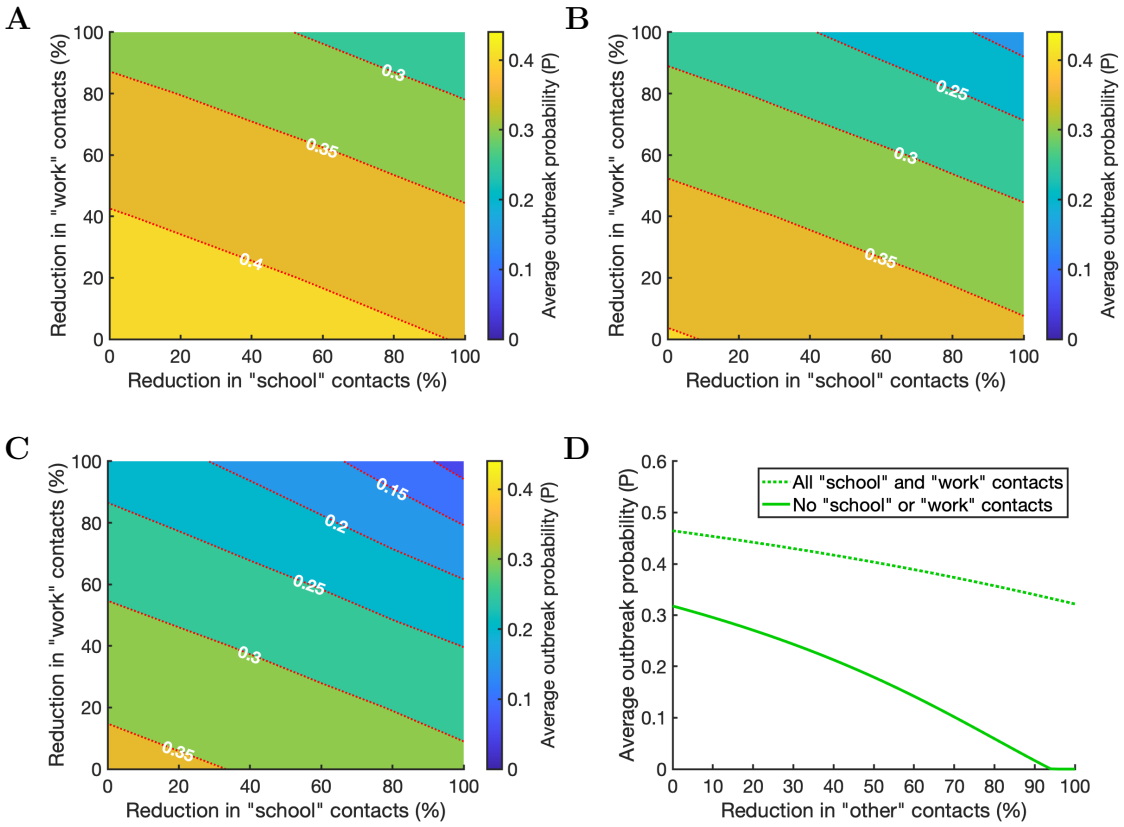


Figure B.3: Scenario A: The effects of intervention strategies that combine reductions in 'school', 'work', and 'other' contacts. For scenario A, susceptibility to infection and the proportion of hosts who experience a fully asymptomatic course of infection are independent of age. **A.** The effect of reducing 'school' and 'work' contacts on the weighted average probability of a major outbreak (P), when 'other' contacts are reduced by 25% across all age groups. Red dotted lines indicate contours along which the major outbreak probability is constant. **B.** The analogous figure to A, but with a 50% reduction in 'other' contacts. **C.** The analogous figure to A, but with a 75% reduction in 'other' contacts. **D.** The effect of reducing 'other' contacts on the average major outbreak probability when 'school' and 'work' contacts are not reduced at all (dotted line) and when 'school' and 'work' contacts are reduced by 100% (solid line).

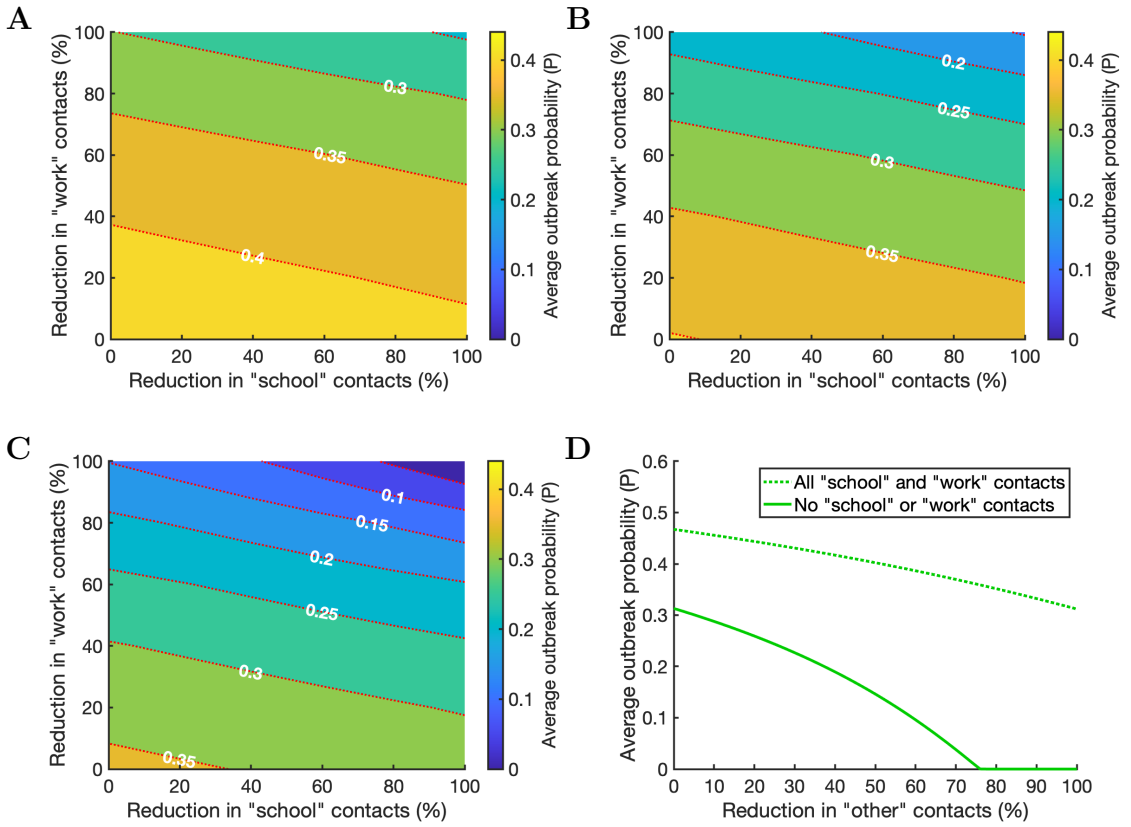


Figure B.4: Scenario B: The effects of intervention strategies that combine reductions in 'school', 'work', and 'other' contacts. For scenario B, susceptibility to infection varies with age but the proportion of hosts who experience a fully asymptomatic course of infection is independent of age. **A.** The effect of reducing 'school' and 'work' contacts on the weighted average probability of a major outbreak (P), when 'other' contacts are reduced by 25% across all age groups. Red dotted lines indicate contours along which the major outbreak probability is constant. **B.** The analogous figure to A, but with a 50% reduction in 'other' contacts. **C.** The analogous figure to A, but with a 75% reduction in 'other' contacts. **D.** The effect of reducing 'other' contacts on the average major outbreak probability when 'school' and 'work' contacts are not reduced at all (dotted line) and when 'school' and 'work' contacts are reduced by 100% (solid line).

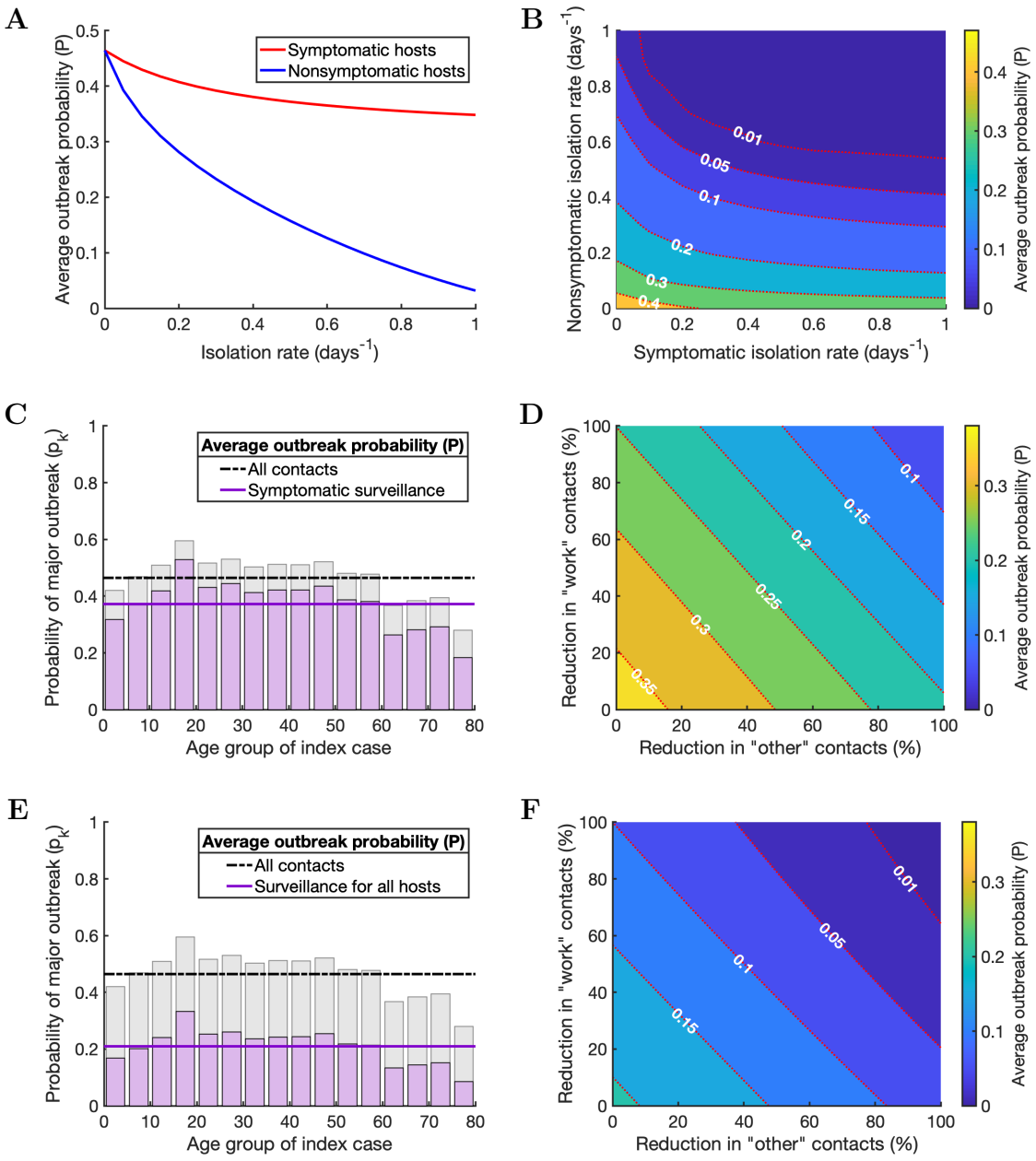


Figure B.5: Scenario A: Surveillance as part of a mixed strategy to reduce the major outbreak probability. For scenario A, susceptibility to infection and the proportion of hosts who experience a fully asymptomatic course of infection are independent of age. **A.** The effect of increasing the isolation rate of symptomatic (red line) or nonsymptomatic (blue line) infected hosts on the average probability of a major outbreak (P), in the absence of contact-reducing NPIs. The isolation rates ρ_k and σ_k are varied between 0 days^{-1} and 1 days^{-1}. **B.** The effect of simultaneously varying the isolation rate of symptomatic and nonsymptomatic hosts on the average probability of a major outbreak (P), again without contact-reducing NPIs. **C.** The age-dependent probability of a major outbreak when the isolation rate for symptomatic individuals is $\rho_k = 1/2$ days^{-1}, without contact-reducing NPIs or surveillance of nonsymptomatic infected individuals (purple bars and solid line). Pale grey bars and black dash-dotted line represent the major outbreak probabilities without any contact-reducing NPIs or enhanced surveillance (as in Fig 4.3A). **D.** The effect of reducing 'work' and 'other' contacts when the isolation rate of symptomatic individuals is $\rho_k = 1/2$ days^{-1}, as in C, without surveillance of nonsymptomatic infected individuals. **E,F.** The analogous figures to C,D, with enhanced surveillance of both symptomatic and nonsymptomatic infected hosts ($\rho_k = 1/2$ days^{-1} and $\sigma_k = 1/7$ days^{-1}).

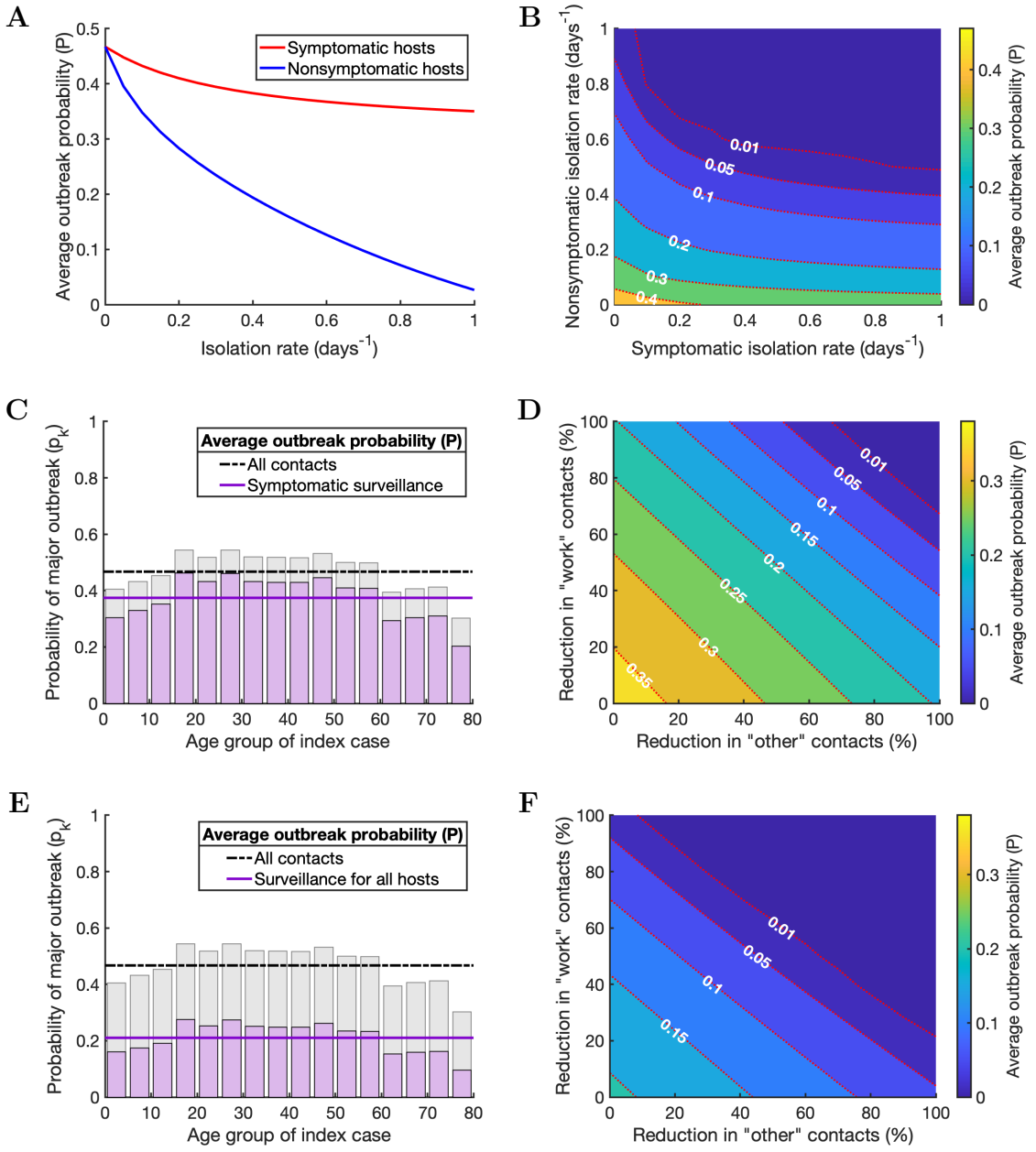


Figure B.6: Scenario B: Surveillance as part of a mixed strategy to reduce the major outbreak probability. For scenario B, susceptibility to infection varies with age but the proportion of hosts who experience a fully asymptomatic course of infection is independent of age. **A.** The effect of increasing the isolation rate of symptomatic (red line) or nonsymptomatic (blue line) infected hosts on the average probability of a major outbreak (P), in the absence of contact-reducing NPIs. The isolation rates ρ_k and σ_k are varied between 0 days^{-1} and 1 days^{-1}. **B.** The effect of simultaneously varying the isolation rate of symptomatic and nonsymptomatic hosts on the average probability of a major outbreak (P), again without contact-reducing NPIs. **C.** The age-dependent probability of a major outbreak when the isolation rate for symptomatic individuals is $\rho_k = 1/2$ days^{-1}, without contact-reducing NPIs or surveillance of nonsymptomatic infected individuals (purple bars and solid line). Pale grey bars and black dash-dotted line represent the major outbreak probabilities without any contact-reducing NPIs or enhanced surveillance (as in Fig 4.3B). **D.** The effect of reducing 'work' and 'other' contacts when the isolation rate of symptomatic individuals is $\rho_k = 1/2$ days^{-1}, as in C, without surveillance of nonsymptomatic infected individuals. **E,F.** The analogous figures to C,D, with enhanced surveillance of both symptomatic and nonsymptomatic infected hosts ($\rho_k = 1/2$ days^{-1} and $\sigma_k = 1/7$ days^{-1}).

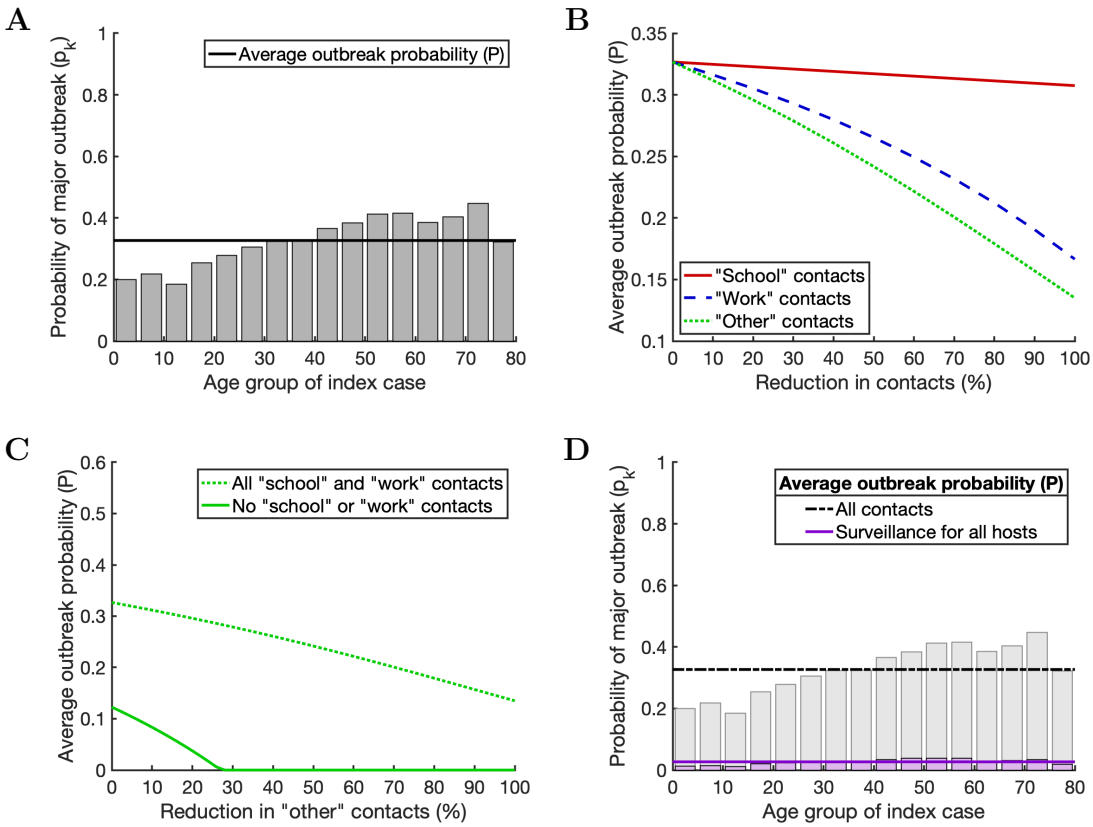


Figure B.7: Scenario C: The effect of reducing the basic reproduction number from $R_0 = 3$ (baseline value) to $R_0 = 2$. For scenario C, both susceptibility to infection and the proportion of hosts who experience a fully asymptomatic course of infection vary with age. **A.** Analogous to Fig 4.3C in the main text: the probability that a single infected individual in any given age group triggers a major outbreak (grey bars) and the weighted average major outbreak probability P (black horizontal line). **B.** Analogous to Fig 4.4D in the main text: partial reductions in 'school', 'work', and 'other' contacts, and the resulting reductions in the average major outbreak probability P (solid red, dashed blue, and dotted green lines respectively). **C.** Analogous to Fig 4.5D in the main text: the effect of reducing 'other' contacts on the average major outbreak probability when 'school' and 'work' contacts are not reduced at all (dotted line) and when 'school' and 'work' contacts are reduced by 100% (solid line). **D.** Analogous to Fig 4.6E in the main text: the age-dependent probability of a major outbreak with enhanced surveillance of both symptomatic and nonsymptomatic infected hosts ($\rho_k = 1/2$ days^{-1} and $\sigma_k = 1/7$ days^{-1}, without contact-reducing NPIs (purple bars and solid line). Pale grey bars and black dash-dotted line represent the major outbreak probabilities without any contact-reducing NPIs or enhanced surveillance (as in panel A.).

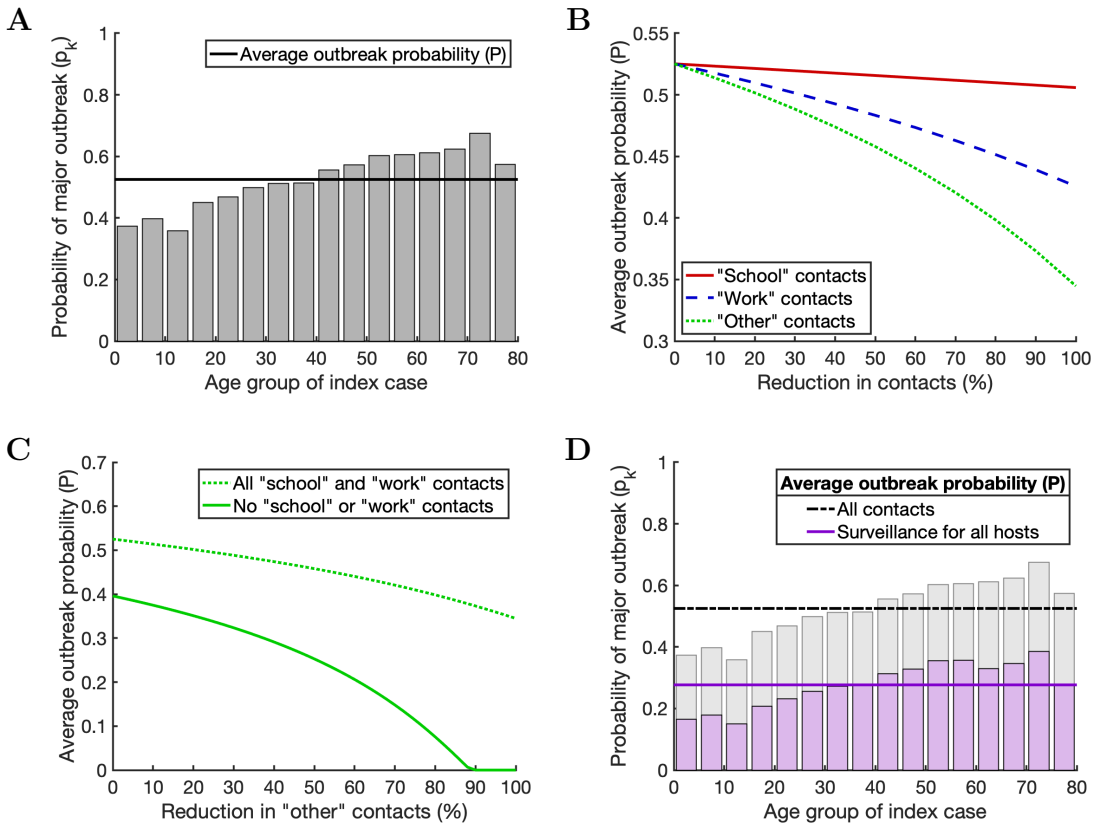


Figure B.8: Scenario C: The effect of increasing the basic reproduction number from $R_0 = 3$ (baseline value) to $R_0 = 4$. For scenario C, both susceptibility to infection and the proportion of hosts who experience a fully asymptomatic course of infection vary with age. **A.** Analogous to Fig 4.3C in the main text: the probability that a single infected individual in any given age group triggers a major outbreak (grey bars) and the weighted average major outbreak probability P (black horizontal line). **B.** Analogous to Fig 4.4D in the main text: partial reductions in 'school', 'work', and 'other' contacts, and the resulting reductions in the average major outbreak probability P (solid red, dashed blue, and dotted green lines respectively). **C.** Analogous to Fig 4.5D in the main text: the effect of reducing 'other' contacts on the average major outbreak probability when 'school' and 'work' contacts are not reduced at all (dotted line) and when 'school' and 'work' contacts are reduced by 100% (solid line). **D.** Analogous to Fig 4.6E in the main text: the age-dependent probability of a major outbreak with enhanced surveillance of both symptomatic and nonsymptomatic infected hosts ($\rho_k = 1/2$ days^{-1} and $\sigma_k = 1/7$ days^{-1}, without contact-reducing NPIs (purple bars and solid line). Pale grey bars and black dash-dotted line represent the major outbreak probabilities without any contact-reducing NPIs or enhanced surveillance (as in panel A.).

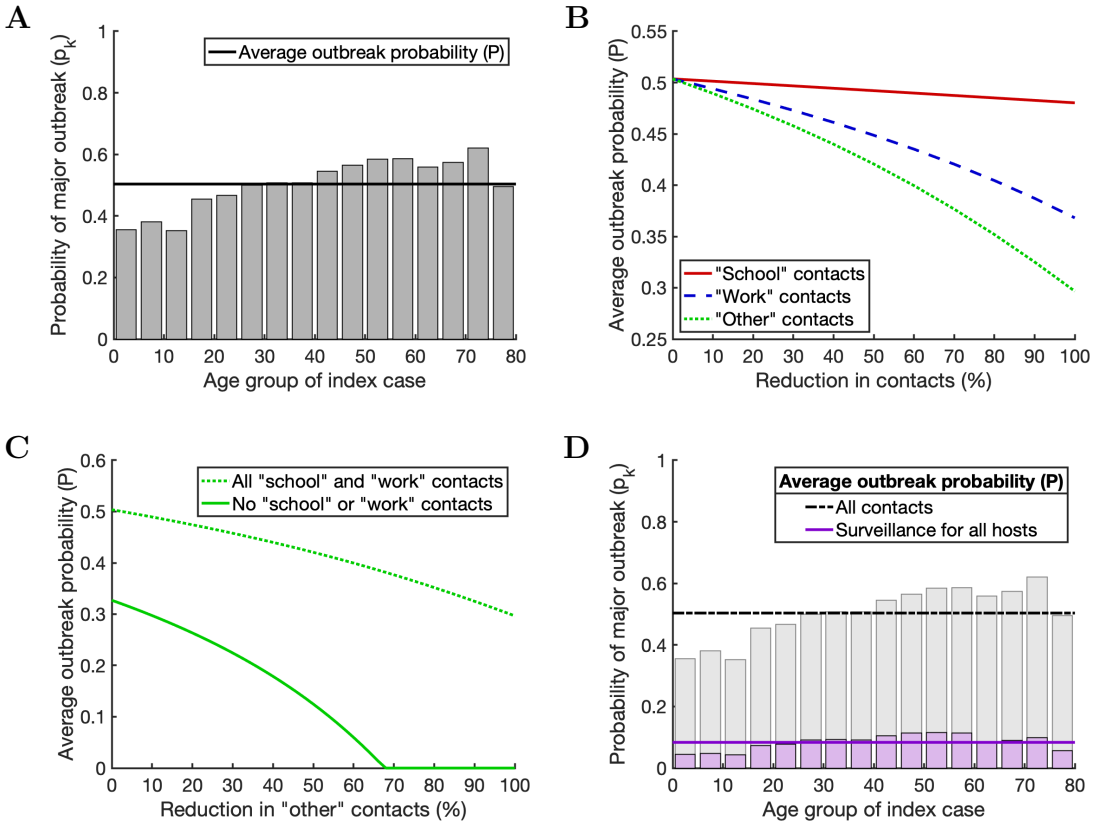


Figure B.9: Scenario C: The effect of reducing the proportion of infections that arise from presymptomatic hosts from $K_p = 0.489$ (baseline value) to $K_p = 0.25$. For scenario C, both susceptibility to infection and the proportion of hosts who experience a fully asymptomatic course of infection vary with age. **A.** Analogous to Fig 4.3C in the main text: the probability that a single infected individual in any given age group triggers a major outbreak (grey bars) and the weighted average major outbreak probability P (black horizontal line). **B.** Analogous to Fig 4.4D in the main text: partial reductions in 'school', 'work', and 'other' contacts, and the resulting reductions in the average major outbreak probability P (solid red, dashed blue, and dotted green lines respectively). **C.** Analogous to Fig 4.5D in the main text: the effect of reducing 'other' contacts on the average major outbreak probability when 'school' and 'work' contacts are not reduced at all (dotted line) and when 'school' and 'work' contacts are reduced by 100% (solid line). **D.** Analogous to Fig 4.6E in the main text: the age-dependent probability of a major outbreak with enhanced surveillance of both symptomatic and nonsymptomatic infected hosts ($\rho_k = 1/2$ days^{-1} and $\sigma_k = 1/7$ days^{-1}, without contact-reducing NPIs (purple bars and solid line). Pale grey bars and black dash-dotted line represent the major outbreak probabilities without any contact-reducing NPIs or enhanced surveillance (as in panel A.).

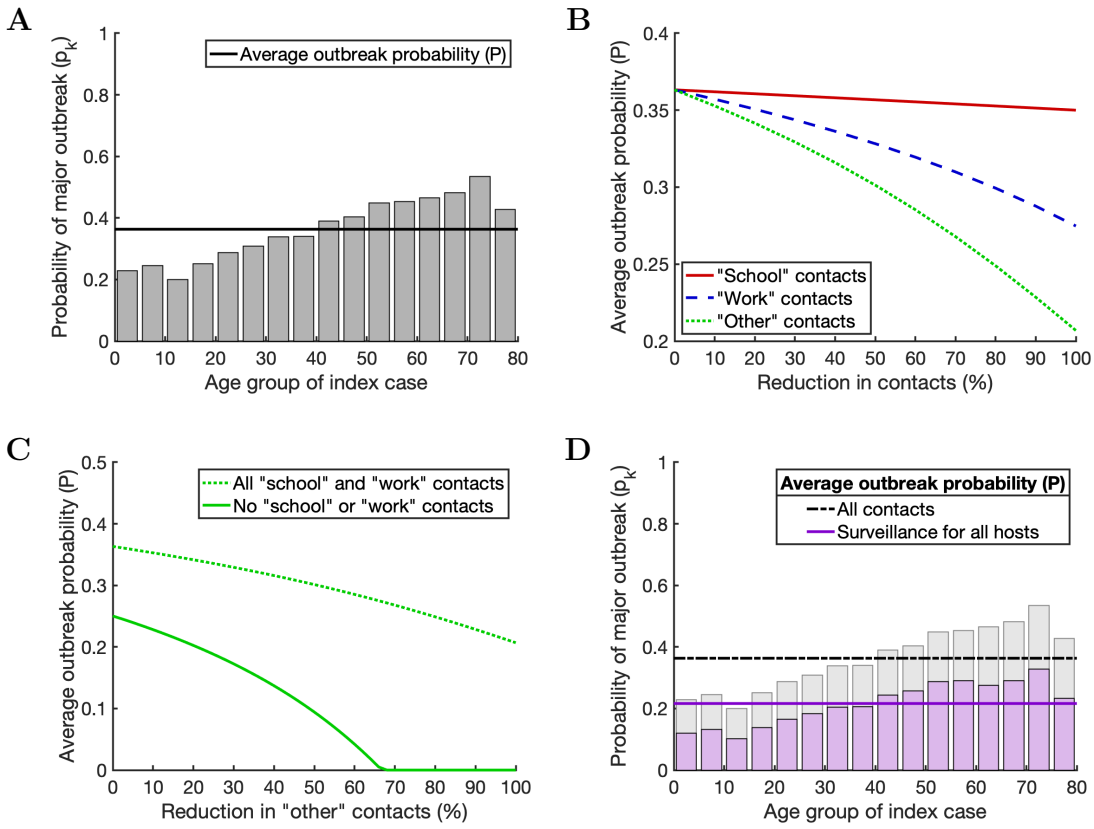


Figure B.10: Scenario C: The effect of increasing the proportion of infections that arise from presymptomatic hosts from $K_p = 0.489$ (baseline value) to $K_p = 0.75$. For scenario C, both susceptibility to infection and the proportion of hosts who experience a fully asymptomatic course of infection vary with age. **A.** Analogous to Fig 4.3C in the main text: the probability that a single infected individual in any given age group triggers a major outbreak (grey bars) and the weighted average major outbreak probability P (black horizontal line). **B.** Analogous to Fig 4.4D in the main text: partial reductions in 'school', 'work', and 'other' contacts, and the resulting reductions in the average major outbreak probability P (solid red, dashed blue, and dotted green lines respectively). **C.** Analogous to Fig 4.5D in the main text: the effect of reducing 'other' contacts on the average major outbreak probability when 'school' and 'work' contacts are not reduced at all (dotted line) and when 'school' and 'work' contacts are reduced by 100% (solid line). **D.** Analogous to Fig 4.6E in the main text: the age-dependent probability of a major outbreak with enhanced surveillance of both symptomatic and nonsymptomatic infected hosts ($\rho_k = 1/2$ days^{-1} and $\sigma_k = 1/7$ days^{-1}, without contact-reducing NPIs (purple bars and solid line). Pale grey bars and black dash-dotted line represent the major outbreak probabilities without any contact-reducing NPIs or enhanced surveillance (as in panel A.).

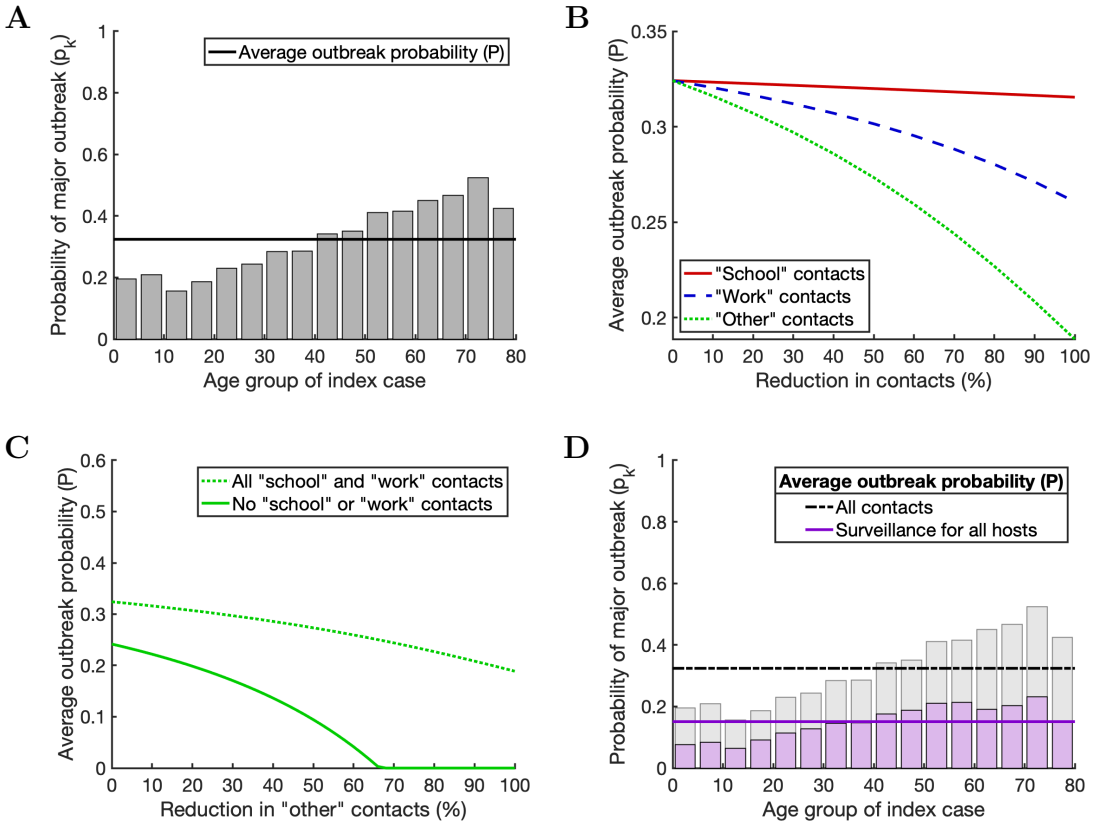


Figure B.11: Scenario C: The effect of reducing the proportion of infections that arise from asymptomatic hosts from $K_a = 0.106$ (baseline value) to $K_a = 0.01$. For scenario C, both susceptibility to infection and the proportion of hosts who experience a fully asymptomatic course of infection vary with age. **A.** Analogous to Fig 4.3C in the main text: the probability that a single infected individual in any given age group triggers a major outbreak (grey bars) and the weighted average major outbreak probability P (black horizontal line). **B.** Analogous to Fig 4.4D in the main text: partial reductions in 'school', 'work', and 'other' contacts, and the resulting reductions in the average major outbreak probability P (solid red, dashed blue, and dotted green lines respectively). **C.** Analogous to Fig 4.5D in the main text: the effect of reducing 'other' contacts on the average major outbreak probability when 'school' and 'work' contacts are not reduced at all (dotted line) and when 'school' and 'work' contacts are reduced by 100% (solid line). **D.** Analogous to Fig 4.6E in the main text: the age-dependent probability of a major outbreak with enhanced surveillance of both symptomatic and nonsymptomatic infected hosts ($\rho_k = 1/2$ days^{-1} and $\sigma_k = 1/7$ days^{-1}, without contact-reducing NPIs (purple bars and solid line). Pale grey bars and black dash-dotted line represent the major outbreak probabilities without any contact-reducing NPIs or enhanced surveillance (as in panel A.).

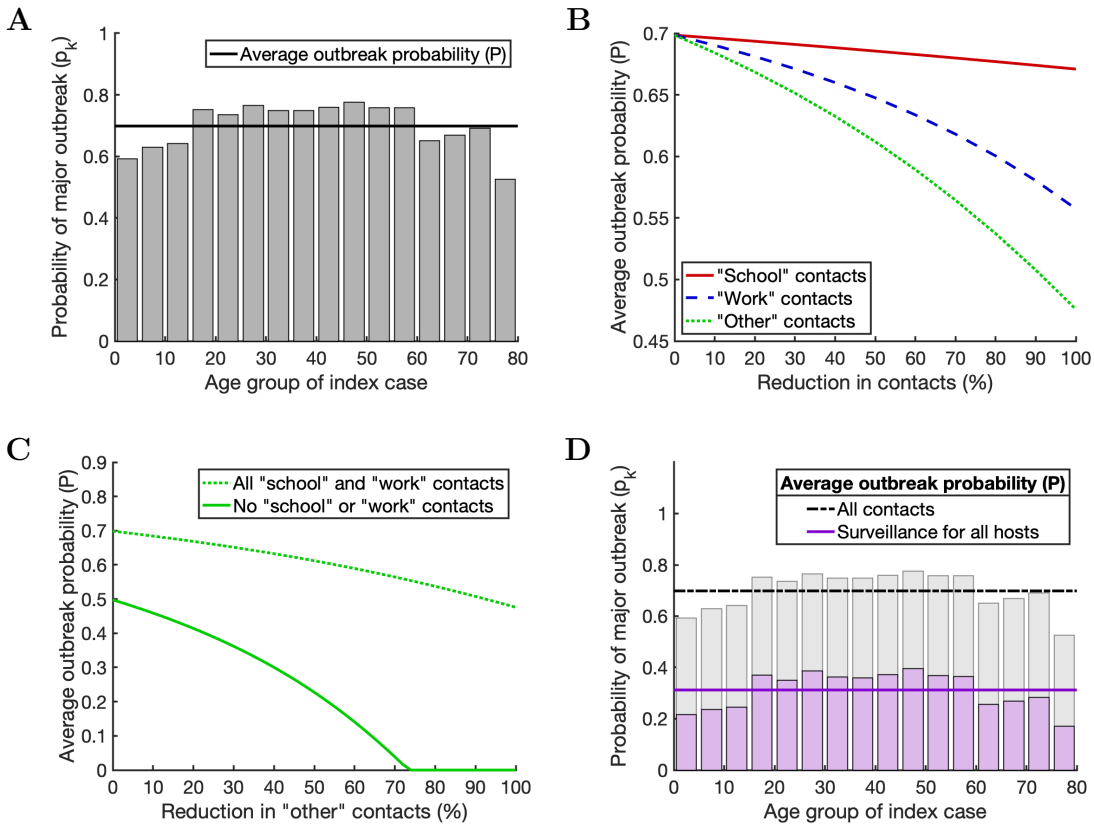


Figure B.12: Scenario C: The effect of increasing the proportion of infections that arise from asymptomatic hosts from $K_a = 0.106$ (baseline value) to $K_a = 0.5$. For scenario C, both susceptibility to infection and the proportion of hosts who experience a fully asymptomatic course of infection vary with age. **A.** Analogous to Fig 4.3C in the main text: the probability that a single infected individual in any given age group triggers a major outbreak (grey bars) and the weighted average major outbreak probability P (black horizontal line). **B.** Analogous to Fig 4.4D in the main text: partial reductions in 'school', 'work', and 'other' contacts, and the resulting reductions in the average major outbreak probability P (solid red, dashed blue, and dotted green lines respectively). **C.** Analogous to Fig 4.5D in the main text: the effect of reducing 'other' contacts on the average major outbreak probability when 'school' and 'work' contacts are not reduced at all (dotted line) and when 'school' and 'work' contacts are reduced by 100% (solid line). **D.** Analogous to Fig 4.6E in the main text: the age-dependent probability of a major outbreak with enhanced surveillance of both symptomatic and nonsymptomatic infected hosts ($\rho_k = 1/2$ days^{-1} and $\sigma_k = 1/7$ days^{-1}, without contact-reducing NPIs (purple bars and solid line). Pale grey bars and black dash-dotted line represent the major outbreak probabilities without any contact-reducing NPIs or enhanced surveillance (as in panel A.).

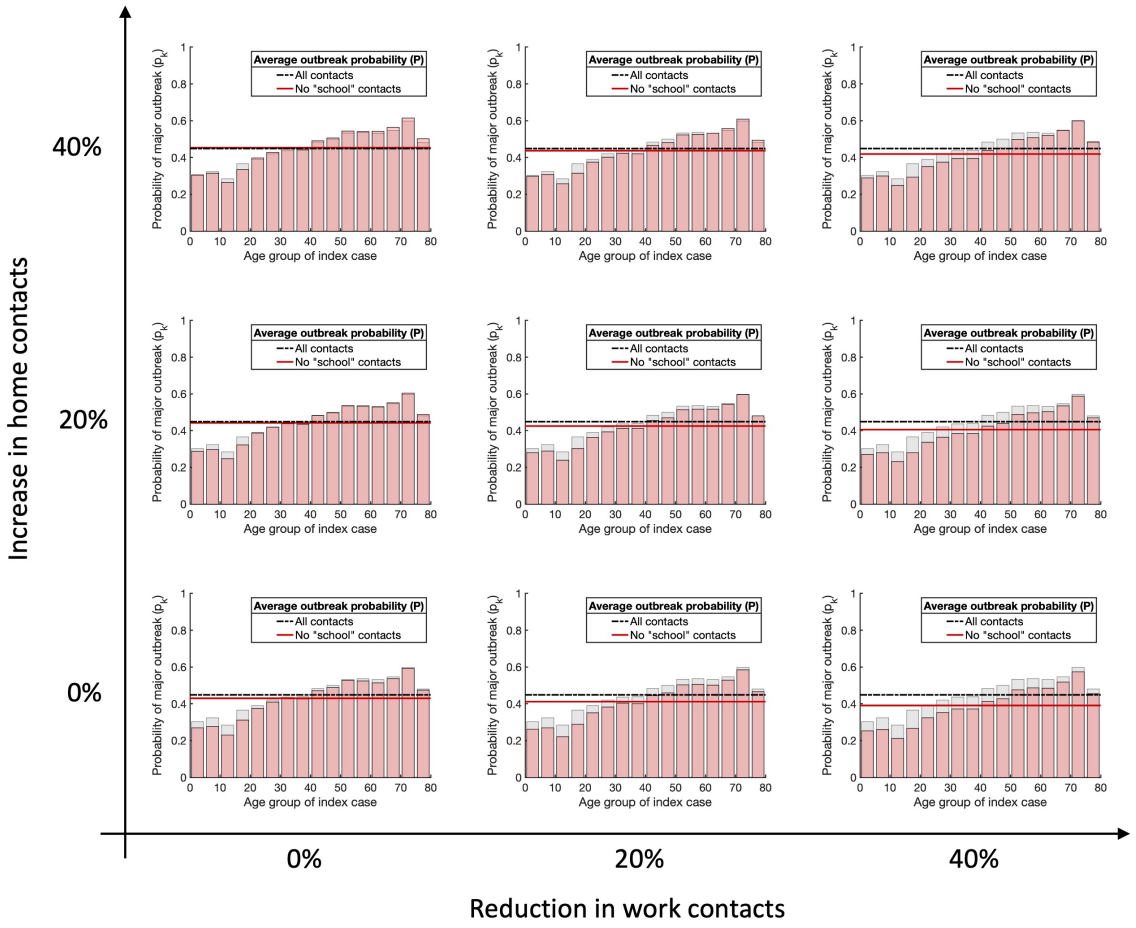


Figure B.13: Scenario C: The effects of school closures on the age-dependent major outbreak probability, allowing for possible secondary effects on 'work' and 'home' contacts. In each panel the age-dependent major outbreak probability is shown in the absence of all 'school' contacts, allowing also for a specified reduction in 'work' contacts and increase in 'home' contacts that may occur as a result of school closures. Columns left to right represent a 0%, 20%, and 40% reduction in 'work' contacts, respectively. Rows bottom to top represent a 0%, 20%, and 40% increase in 'home' contacts, respectively. In every case, we assume that 'other' contacts are unaffected and remain as shown in Fig 4.2F of the main text. Although the shape of the age-dependent risk profile in the absence of 'school' contacts is robust to these changes in 'work' and 'home' contacts, the weighted average major outbreak probability P (indicated by the solid red line in every case) varies. In particular, if the increase in 'home' contacts occurring as a result of school closures is high enough, this may counteract the benefits of reduced 'school' contacts. These results support our conclusion that school closures are unlikely to have a substantial impact on SARS-CoV-2 transmission when applied as the sole NPI.

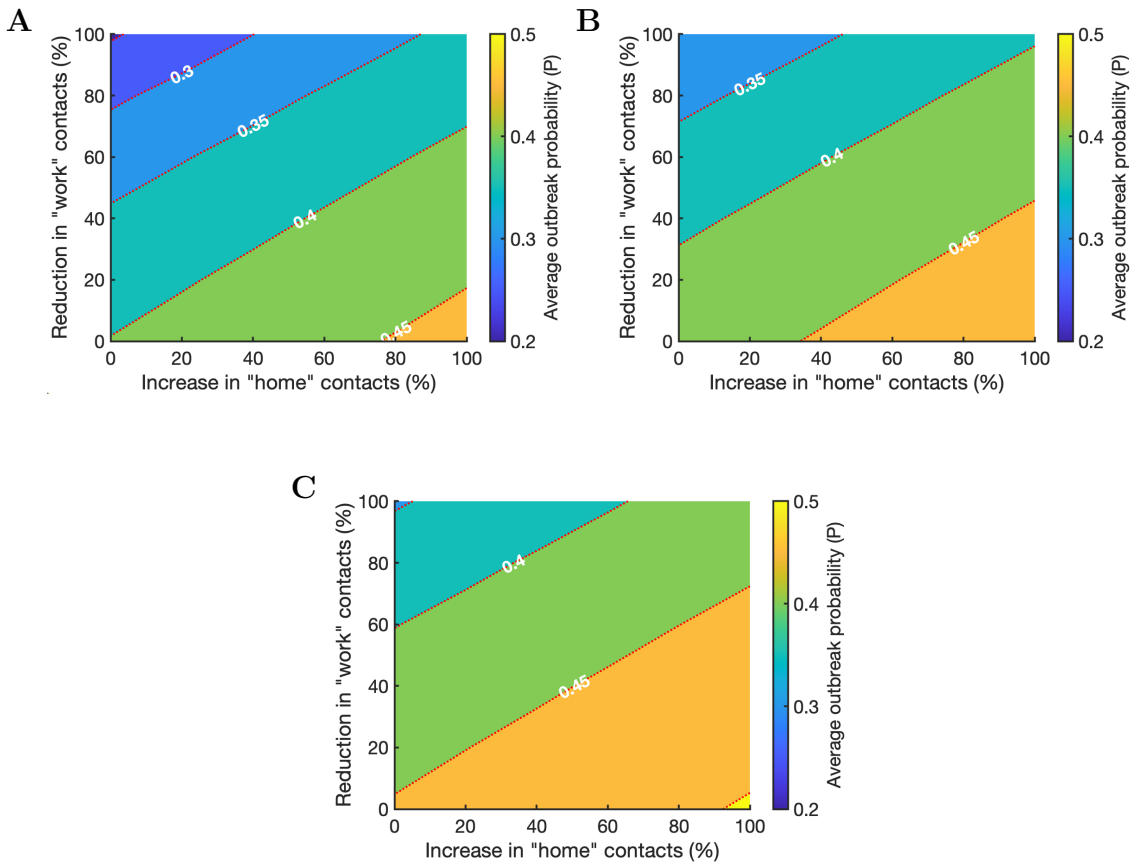


Figure B.14: Scenario C: The effects of school closures on the age-dependent major outbreak probability, allowing for possible secondary effects on 'work', 'home', and 'other' contacts. A. The effect of increasing 'home' contacts and reducing 'work' contacts on the weighted average probability of a major outbreak (P), when 'other' contacts are reduced by 20% and 'school' contacts are removed entirely. Red dotted lines indicate contours along which the major outbreak probability is constant. **B.** The analogous figure to A, but with a 0% change in 'other' contacts compared to the baseline level. **C.** The analogous figure to A, but with a 20% increase in 'other' contacts compared to the baseline level.

Part I: Summary and concluding remarks

When an infectious pathogen arrives in a susceptible population, there is a risk that it may trigger a major outbreak of disease, with potentially severe consequences for public health. To contain the spread of the pathogen, policy makers may decide to implement a variety of control interventions that aim to interrupt transmission, such as surveillance for infected hosts and their subsequent isolation, social distancing measures, school closures, and so on. Assessing the effectiveness of these control measures for a particular infectious pathogen is critical to ensure that the intervention package deployed is sufficient to control the outbreak, without employing unnecessarily restrictive measures. This becomes particularly important when there are heterogeneities in the infection and transmission characteristics of the population that influence how different groups of individuals contribute to pathogen spread. In that scenario, a policy maker must determine which groups within the population should be targeted with control measures, and how best to allocate limited resources for surveillance amongst the heterogeneous population.

One way to quantitatively assess the effectiveness of a control intervention for pathogen containment is to measure how that intervention influences the probability that an infected host introduced into the population initiates a major disease outbreak. This outbreak probability can be calculated straightforwardly for complex pathosystems under different control scenarios using the first-step analysis method, which has previously been used in a variety of disease-control contexts (see Chapter 2). In Part I of this thesis, we used SARS-CoV-2 as a case study to explore how the probability of a major outbreak can be calculated when heterogeneities in the population affect transmission. In particular, we extended the previously developed branching process methodology to explicitly account for the possibility of both presymptomatic and asymptomatic transmission, as well as incorporating a wide range of different age classes and the corresponding variations in susceptibility to infection and likelihood of developing clinical symptoms of COVID-19. Then, we demonstrated how these outbreak probability estimates could be used as a metric for assessing the effects of various types of control interventions, and considered how the population heterogeneities described above influence the optimal approach for control.

The results obtained in Chapters 3 and 4 have highlighted the following key messages regarding the effects of these heterogeneities on outbreak risks and optimal surveillance and control interventions.

- If transmission of an infectious pathogen can occur from infected individuals who are not showing clinical symptoms, those individuals may contribute substantially to the risk that a major outbreak occurs. If the proportion of transmission occurring from nonsymptomatic hosts is large, control interventions that target symptomatic hosts alone may not be able to eliminate the risk of major outbreaks. In that scenario, interventions that instead seek to identify and isolate nonsymptomatic infected hosts may play a crucial role in reducing the probability that a major outbreak occurs.

- The optimal division of surveillance resources between symptomatic and nonsymptomatic hosts depends not only on how much nonsymptomatic hosts contribute to infection, but also on the relative effectiveness of the identification methods. Even when identifying and isolating nonsymptomatic hosts is difficult relative to symptomatic hosts, a mixed strategy that targets both host types may be the optimal strategy (Supplementary Fig A.6).

- If susceptibility to infection and/or the likelihood of expressing clinical symptoms varies with age, the age groups of the early cases can make a substantial difference to the risk of observing a major outbreak. Therefore, the effectiveness of control measures that target different age groups within the population in different ways may be substantially influenced by these age-related heterogeneities. For this reason, policy makers should not assume that pathogens with a similar mode of transmission should be controlled in the same way, since they may have very different age-related infection characteristics (such as SARS-CoV-2 and influenza-A viruses, for example).

- Assessing the risk that different age groups will trigger a major outbreak is particularly important when the various age-related factors that influence pathogen transmission appear to act in conflicting ways. For example, in the case of SARS-CoV-2, before we conducted our analyses it was unclear whether the large numbers of social contacts observed amongst school aged individuals would outweigh their reduced susceptibility to infection and likelihood of developing clinical symptoms. Exploring the interacting effects of the various age-related factors that influence pathogen transmission is crucial to ensure that interventions are targeted at populations and locations where they will have the greatest effect.

In general, the work presented in Part I of this thesis highlights the importance of identifying relevant population heterogeneities that may influence transmission dynamics early in an outbreak, and incorporating them into models of pathogen transmission and control. Here, we focused on population heterogeneities relevant to SARS-CoV-2 transmission, but

other pathogens also exhibit different types of heterogeneity that play an important role in governing their transmission dynamics. For future emerging pathogens it will be crucial to focus on identifying such relevant heterogeneities early, so that control resources can be deployed optimally as soon as possible.

As well as emphasising the importance of accounting for these heterogeneities when assessing outbreak risks and the deployment of control interventions, our work here demonstrates how such complex heterogeneities can be incorporated into the existing branching process methodology. Our results highlight that the first-step analysis approach provides a flexible framework for assessing outbreak risks that can be straightforwardly extended to account for various model complexities, but which can nevertheless yield important insights into the effectiveness of control interventions for outbreak containment. When used in combination with a variety of other modelling techniques, first-step analyses for estimating outbreak risk probabilities have the potential to make valuable contributions to informing control interventions for future emerging and re-emerging pathogens.

Part II

'Sentinel' plants for surveillance: Improving early detection monitoring for invasive plant pathogens

Publication details and author contributions

Some of the work presented in Part II of this thesis has been adapted for publication in the following peer-reviewed journal article:

- **F.A. Lovell-Read**, S.R. Parnell, N.J. Cunniffe and R.N. Thompson, 'Using 'sentinel' plants to improve early detection of invasive plant pathogens'. *PLoS Comp. Biol., 2023, 19(2), e1010884. DOI: 10.1371/journal.pcbi.1010884.*

  CRediT author contributions for this work are as follows. Conceptualisation: All authors. Methodology: FALR, RNT. Investigation: FALR. Writing – original draft: FALR, RNT. Writing – review and editing: All authors. Supervision: RNT. All authors gave final approval for publication.

  Material published in this article comprises part of Chapter 5 and the majority of Chapter 6 of this thesis.

The work presented in Chapter 7 of this thesis has not currently been submitted or published elsewhere. However, we would like to further acknowledge the co-authors of the publication above for useful discussions relating to this study.

Chapter 5

Introduction to Part II

Infectious pathogens and the diseases they cause represent a severe and increasing threat to global plant health, with wide-ranging and devastating economic, environmental, and societal consequences [4, 5, 52, 217–228].

Among these, a key concern is the risk to global food security posed by plant pathogens that infect and damage nutritionally important crops, such as barley, maize, potato, rice, and wheat [4, 224, 226–229]. To highlight a famous historical example, in the mid-19$^{\text{th}}$ century the pathogen *Phytophthora infestans* arrived in Ireland, causing a devastating epidemic of potato late blight across the country that rendered infected plants rotten and inedible [4, 230]. Since at that time much of the Irish population was heavily dependent on potatoes for sustenance, the so-called 'Irish potato famine' resulted in around one million people dying of starvation, with another million forced to emigrate to avoid the same fate [4, 230]. Nowadays, although food security is less at risk in high income countries where many sources of nutrition are available, plant diseases still present a substantial threat to food production in lower income countries where locally grown crops typically form a staple part of the diet [4]. For example, cassava (a starchy root vegetable) is a very important crop plant in many low income countries, due to its tolerance to a range of environmental conditions that allows it to be grown reliably in regions where other crops cannot thrive [231]. However, in recent decades cultivation of cassava in sub-Saharan Africa has been threatened by the emergence of a particularly severe strain of the cassava mosaic virus, which causes substantial yield losses in infected plants [231]. The spread of the mosaic virus through Uganda, Kenya, Tanzania, and several other countries in East Africa has destabilised food security in that region, with many farmers forced to abandon cassava cultivation entirely [4].

In addition to the threat to food security, when an infectious pathogen invades and becomes established in a region in which commercial crops are grown, the resulting disease outbreak and any corresponding reduction in crop yield may lead to severe economic losses. For example, since citrus greening disease (a bacterial infection of citrus with causal agents

Candidatus Liberibacter spp.) was first observed in Florida in 2005, it has reduced citrus production in that state by approximately 74% [232]. Combined with the costs of surveillance and control measures put in place to prevent further spread, the annual cost to the Florida citrus industry has been estimated to be around USD $1 billion [233, 234]. Many other economically important crops are similarly under severe threat from infectious disease, including rice, wheat, soybean, tomato, sugar cane, grape, and banana [235–241]. Altogether, an estimated 16% of the global agricultural yield is lost annually to plant disease [242], resulting in total economic losses of USD $220 billion per year [243].

As well as crops that have clear economic and nutritional importance, plant diseases also affect many other species that play crucial roles in local ecosystems [225]. High-profile examples include devastating diseases of forest trees such as sudden oak death, ash dieback, and Dutch elm disease, which have each caused substantial damage to trees in affected areas [5, 51, 52, 54]. Loss of such trees can have a variety of detrimental effects on the natural environment, including reduction in biodiversity, habitat loss, increased soil erosion and flood risk, and reduced carbon sequestration [52, 225]. Furthermore, trees play an important role in human wellbeing, with key benefits including improved air quality, the cooling of urban environments, and the provision of shade [5]. The environmental and societal implications of plant disease can therefore also be substantial.

Pathogens that escape their native geographical range and arrive in new locations pose a particular threat to plant health in those regions. Whilst plants that have co-evolved with pathogens in their native location over thousands of years typically have some resistance to infection, potential hosts elsewhere may have a less effective innate immune response to those pathogens and may thus be susceptible to severe disease when infected [4, 52]. Unfortunately, the risk of non-native pathogens being introduced to new susceptible populations is extremely high, due to the large-scale global trade in live plant material [51–54]. Indeed, many recent severe plant disease outbreaks are thought to have originated in imported infected nursery stock [52]. Since plant trade represents the primary risk pathway for the introduction of invasive pathogens, many countries worldwide impose phytosanitary regulations on imported plant material, which typically consist of visual health checks conducted at the border. However, the vast scale of the plant trade network means that it is not feasible to thoroughly inspect every consignment of imported plants for signs of infection – it has been estimated that as little as 2-6% of plant shipments can realistically be screened in practice [52, 229]. Furthermore, since such checks typically rely on visual inspection for signs of disease, plants that are infected but not yet showing visible or clear symptoms may pass through undetected [54]. Therefore, it is all too easy for imported pathogens to evade detection and subsequently become established in new locations.

Given the high risk of infectious pathogens being introduced to new places, and the potentially severe consequences of the disease outbreaks that may follow, developing effective methods for plant disease surveillance and control is a vital area of current research [218, 244–248]. Since disease management strategies are typically resource-intensive and can themselves incur substantial economic costs, a key challenge for global governments and international organisations seeking to protect plant health is determining how management resources can be optimally deployed to achieve the maximum possible reduction in disease for a given level of control [249]. To that end, mathematical modelling is increasingly used to guide risk assessments, surveillance programmes, and intervention strategies for plant pathogens [218, 244, 250–252], helping policy makers understand how to direct limited resources for control to minimise the cost of outbreaks [35, 223, 253–256].

5.1 Mathematical modelling for post-detection control

Multiple studies of different pathogens have focused on the question of how to optimise control measures when a pathogen is known to be in a particular plant population (known as 'reactive' control). In that scenario, control interventions commonly include 'roguing' – the removal of infected hosts – to prevent further transmission of the pathogen in the population (see Cunniffe *et al.* [21] for examples of pathosystems in which roguing is commonly applied). However, since the number of plants that can be removed at any given time is typically limited due to logistical constraints, the success of roguing strategies relies on targeting those plants that are likely to contribute the most to onward transmission.

With this in mind, Adrakey *et al.* [249] developed a mathematical approach for prioritising hosts for removal based on the infection risk they pose to other plants. In their model framework, snapshots of observed epidemic data are used to assign every potential host plant a 'risk measure', representing the likelihood of being infected at the time the control strategy is implemented, as well as a 'hazard measure', which quantifies the likely contribution of that plant to onward transmission based on its proximity to other susceptible plants. These are combined to form a 'threat measure' that captures the overall likely contribution to infection of a given plant at a given time. The larger the threat measure, the greater the potential reduction in transmission risk that could be achieved by removing that plant, and therefore the higher the priority for its removal. Using citrus canker (a bacterial disease of citrus plants) in Florida as a case study, the authors demonstrated that the plants with the highest threat measure are typically located on the edge of a cluster of existing infection, where they are likely both to become infected and to transmit the infection to other nearby susceptible plants. Lower priority for roguing is placed on plants in the middle of an existing infection locus, since they are likely to lead to fewer secondary infections.

In addition to removing infected hosts that pose a high risk to nearby susceptible plants, roguing strategies may be extended to include the removal of all other plants within a specified 'cull radius' [21]. Although this approach aims to reduce the likelihood of onward transmission by targeting high risk potential hosts in the vicinity of infected plants, the effectiveness of such a strategy is strongly influenced by the choice of cull radius; too small, and nearby infected hosts may be missed, but too large, and many healthy plants will be removed unnecessarily. Mathematical modelling can be used to identify the optimal removal radius around confirmed infected hosts, balancing the reduction in the risk of further spread with the cost of destroying potentially healthy plants [34, 257, 258]. For example, Cunniffe *et al.* [34] showed that the optimal radius for control depends strongly on the rate of pathogen transmission, the dispersal scale of the pathogen, and the presymptomatic period of the infection, with large values of any of these factors favouring a greater cull radius. Additionally, they considered how the optimal strategy changes when the objective of the policy maker is to minimise the large scale impact of the epidemic rather than focusing on the local cost to the grower. In that case, more severe control strategies with greater cull radii are favoured despite the high cost at the local level, since the likelihood of the pathogen being exported to new locations – and thus the global epidemic cost – will be reduced.

If the eradication of a pathogen in a region of known infection is no longer feasible, policy makers may instead choose to deploy interventions that aim to contain the pathogen within the existing affected area. One approach for preventing the further spread of a pathogen beyond its current range is to perform intensive surveillance and control within a so-called 'buffer zone' surrounding the existing infected region, that acts as a 'firebreak' for infection [259]. As with the cull radius strategy described above, a key question for policy makers seeking to use this approach is how wide the buffer zone should be. It if is too narrow, the pathogen may be able to 'jump' from one side to the other via long-distance dispersal events, but if it is too wide, the resources that are available for control may be insufficient to allow intensive surveillance to take place throughout the designated region. White *et al.* [260] constructed a mathematical model of pathogen transmission and control to investigate how the risk of a pathogen escaping a defined buffer zone depends on the zone width and the intensity of surveillance, demonstrating how these factors can be selected optimally for a given surveillance resource in order to minimise the risk of escape. Their results suggest that when the budget for control is low, resources should be focused on increasing the intensity of surveillance; however, as the control budget increases, directing resources towards widening the buffer zone becomes increasingly beneficial.

5.2 Mathematical modelling for early detection

While reactive control has been well-studied, a key additional challenge in plant disease management is designing surveillance programmes that are specifically intended to achieve early detection of an invading pathogen, thus increasing the chance of eradicating the pathogen from the landscape before it becomes widespread, or containing it within the existing affected region ('preventive' control) [261]. Surveillance programmes for plant disease monitoring typically rely on regularly selecting a sample of plants from the at-risk population and inspecting them for visible signs of infection [262, 263]. Following the detection of the pathogen, measures for its eradication or containment such as the 'cull radius' or 'buffer zone' strategies described above can then be implemented. However, although early detection of a pathogen is crucial for the success of subsequent control measures, the frequent large-scale surveillance required to achieve it is both expensive and logistically challenging. For a grower seeking to implement a monitoring programme, a key question is therefore how much resource they should dedicate towards surveillance. If the sampling intensity is low, it is likely that the pathogen will reach a high prevalence in the population before it is discovered, making its subsequent eradication or containment unfeasible, and increasing the long term cost to the grower. However, if the surveillance programme is too intensive, valuable resources that could be used elsewhere (e.g. directed to other management strategies) may be wasted.

To explore this trade-off, Parnell *et al.* [264] developed a mathematical approach for estimating the expected prevalence of disease in a population at the time of discovery, based on a monitoring programme in which random samples of size N are selected from the population every Δ days and inspected for signs of infection. They demonstrated that, as long as the pathogen is detected whilst still in its exponential growth phase, the expected detection prevalence can be approximated by the expression $r\Delta/N$ (where r represents the intrinsic epidemic growth rate of the pathogen). Therefore, the faster the spread of the pathogen, the longer the interval between samples, or the smaller the samples that are taken, the greater the number of plants that will be infected before the pathogen is discovered. As well as providing a straightforward estimate of the expected prevalence of infection at the time of detection for a given sampling strategy (choice of N and Δ), this simple 'rule of thumb' can be used to design a monitoring programme capable of detecting a pathogen before it has reached a given prevalence in the population, provided an estimate of the intrinsic growth rate r is available. The authors subsequently tested this rule of thumb against estimates of the detection prevalence obtained through computational simulations of pathogen spread and surveillance, using two commercially important pathogens of citrus as case studies [265]. They demonstrated that their simple analytic approximation for the detection prevalence was surprisingly accurate when compared to the results obtained through stochastic simulations, despite the fact that it did not account for various model complexities such as the spatial resolution of the epidemic. These results highlight that simple heuristic measures such as

this 'rule of thumb' can be very beneficial for guiding disease management policy, with the key advantage that they are straightforward to understand for stakeholders who are not modelling experts.

The approach of Parnell *et al.* was subsequently extended by Mastin *et al.* [61] to apply to a setting in which pathogen transmission via vectors is modelled explicitly, and a policy maker must determine how to divide limited surveillance resources between the plant and vector populations. In that study, the authors derived 'rule of thumb' expressions for the expected prevalences of infection in the host and vector populations at the time of discovery given the sampling effort applied to either group, and explored how sampling resources can be deployed to achieve a specified prevalence of infection in either group at the time of detection. Interestingly, they found that the sampling effort required to achieve a given detection prevalence was minimised when either plants or vectors were exclusively sampled (rather than a mixture of both), where the optimal group to sample from was the one with the greatest proportion of infected hosts during the initial period of epidemic growth. Then, the authors further extended their framework to account for the costs of surveillance, investigating how the relative cost of inspecting hosts and vectors influences the optimal sampling strategy. They derived a simple rule for determining whether it is more cost effective to sample exclusively from plants or from vectors, based upon i) the relative prevalence of infection in the plant and vector populations during the initial growth phase, and ii) the relative cost of sampling plants and vectors. When the former (i) exceeds the latter (ii), the cost of surveillance is minimised when only plants are sampled, and vice versa.

As discussed previously, a key route by which invasive pathogens are introduced to plant populations is through the trade of live plants that are raised in nurseries and exported to new locations for planting. With this in mind, Alonso-Chavez *et al.* [51] adapted the probabilistic approach derived by Parnell *et al.* to explore the use of within-nursery surveillance for reducing the risk of growers selling infected plant material. Specifically, they considered the interaction between a surveillance strategy in which N plants are randomly selected and inspected for disease every Δ days, and a selling programme in which M randomly selected plants are sold every τ days, exploring the probability that an infected plant is sold from the nursery before the surveillance programme is able to detect the presence of the pathogen. Their results demonstrated that the probability of not having sold infected material before detection occurs is small, even if sampling is intensive, suggesting that within-nursery surveillance is unlikely to be a feasible approach for preventing the export of plant pathogens to new locations. This is particularly the case if large numbers of plants are regularly being sold, the intrinsic transmission rate of the pathogen is high, or the pathogen has a long period of presymptomatic infection before visible symptoms develop.

5.3 Presymptomatic transmission and 'sentinel' plants

For many plant pathogens, a major obstacle to achieving early detection is a long presymptomatic period (the time between initial infection and symptom onset)[1] [51, 251, 266–270]. For example, symptoms of Panama disease (also known as Fusarium wilt) in banana take two to six months to develop [271], whilst citrus greening disease has an incubation period that may last anywhere from a few months to multiple years [268, 269]. If transmission by infected hosts can occur prior to symptom onset, the pathogen may spread widely within the susceptible population before any visible signs of infection appear. This presents a substantial hindrance for disease monitoring programmes that rely upon visual inspection of potential hosts for signs of infection (as are standard across plant health [262, 263]), which may fail to identify the presence of a pathogen before widespread transmission has occurred. Indeed, Alonso-Chavez *et al.* [51] showed that if presymptomatic transmission occurs and the pathogen is discovered early in an outbreak, the expected prevalence of infection in the population at the time of discovery increases exponentially with the duration of the incubation period.

Interestingly, the incubation period of a pathogen may vary substantially between different host species, or between hosts of different ages [266–268, 270, 272]. For citrus greening disease, for example, younger citrus plants are likely to develop symptoms more quickly than mature trees (after 6-12 months compared to 1-2.5 years, respectively) [268, 272]. This suggests that in some cases there may be the potential to use alternative plant species with relatively short incubation periods as 'sentinels' to detect new outbreaks. In this context, we refer to a 'sentinel' as a susceptible plant species specifically chosen to have a short incubation period, which is placed amongst the crop plants and regularly monitored for visible signs of infection[2]. In this way, sentinel plants could act as early warning beacons for the presence of a pathogen in a plant population – the canaries in the proverbial coal mine [274]. Indeed, for this reason it has been common practice for viticulturists to plant rose bushes at the ends of rows of grapevine since the middle of the 19$^{\text{th}}$ century [275, 276]. The story goes that in 1851 a group of monks in Burgundy successfully managed to save their vines from an invasion of powdery mildew fungus by observing early symptoms of infection on the rosebushes surrounding the vineyard, alerting them to the presence of the fungus and allowing them to treat it before it spread too far [276]. Now, we know that the pathogens that cause powdery mildew in roses and grapevines are two different species of fungi, so this is not a true example of a sentinel system in the sense described above; however, the

[1] This is also commonly referred to in the literature as the cryptic period, the incubation period, or the asymptomatic period. Here, we use the term 'presymtomatic period' for consistency with the terminology used in Part I of this thesis.

[2] We note that, in the literature, the term 'sentinel' has alternatively been used to refer to plant species that are grown outside of their natural habitat and monitored to assess whether pests native to the new location pose a risk to that species, which is distinct from the context considered in this thesis [273].

two fungal pathogens have very similar climatic preferences and simultaneous infections commonly occur in the same vineyards, which is why the practice can be beneficial and continues to this day [277].

Generally speaking, the rapid onset of visible symptoms in sentinel plants could result in earlier detection of the pathogen in the population. If this leads to a reduced prevalence of infection upon discovery, this could reduce the cost of reactive control. However, sentinels may also have drawbacks. The increased rate of disease progression in sentinel plants could result in more rapid onward transmission, which may counteract the positive effect of early detection and lead to an increased detection prevalence. Therefore, research is needed to understand this trade-off and infer the conditions under which sentinel plants are likely to be beneficial for reducing the detection prevalence. If sentinel plants are deployed, one key consideration is how many sentinel plants should be added to the population to provide sufficient opportunity for early detection whilst limiting the concurrent increase in the transmission rate. Another important question is how to divide limited surveillance resources optimally between crop and sentinel plants. Is it more effective to sample preferentially from the available sentinel population, or should a mixture of crop and sentinel plants be inspected?

Although using sentinel plants (as defined in the context of this thesis) to facilitate early detection of a range of invasive pathogens has been suggested as a possibility previously (by Mansfield *et al.* [273] and Eschen *et al.* [278], for example, as well as by 19$^{\text{th}}$ century winegrowers) the question of how to design effective large scale surveillance strategies using sentinel plants targeted at a specific pathogen in a given region has not yet been addressed [273]. In the second part of this thesis, we explore the potential for sentinel plants to aid early outbreak detection, using a plant pathogen of significant current importance as a case study.

5.4 Case study: *Xylella fastidiosa*

An important example of a pathogen for which the development of effective surveillance strategies is currently critical is *Xylella fastidiosa* [279], a vector-borne bacterial pathogen transmitted by xylem sap-feeding insects of order Hemiptera, sub-order Auchenorrhyncha [267, 280]. First isolated from infected grapevines in 1978 [281], *X. fastidiosa* is now recognised to be one of the most dangerous threats to plant health worldwide, highlighted by the fact that it became the first plant pathogen to have its genome fully sequenced [266, 282, 283]. *X. fastidiosa* is known to be pathogenic to over 600 plant species [284], including many economically important crops such as olive, grapevine, almond, citrus, and coffee, as well as a variety of ornamental species and shade trees [267, 280, 283]. Depending on the host species and specific bacterial subspecies (hereafter abbreviated to 'subsp.'),

symptoms of *X. fastidiosa* infection typically include leaf tissue necrosis (leaf scorch), stunted plant growth, decrease in fruit production, dieback of twigs and branches, and eventual death [267, 280, 283]. Outbreaks of *X. fastidiosa*-related diseases in commercial crops therefore have substantial negative economic effects [223, 240, 285].

For example, *X. fastidiosa* subsp. *fastidiosa* causes a dangerous disease in grapevine known as Pierce's Disease (PD), which is prevalent across the USA [267, 283, 286]. PD has been a recognised problem for viticulturists since the late 1800s [287], although *X. fastidiosa* was not implicated as the etiological agent until it was isolated a century later [281]. Infected grapevines experience severe leaf scorch and leaf drop as well as shrivelling and dessication of their fruits (descriptively known as 'raisining') [283, 287], and typically die within a few years of infection. Therefore, Pierce's disease has a significant negative impact on the quantity and quality of grape production in the USA, with severe implications for the country's USD $30 billion wine industry [286]. Indeed, the endemicity of *X. fastidiosa* in the south-eastern states of the USA has already rendered grapevine production economically unfeasible there [267], and PD also threatens the grapevine industry in California, where it leads to estimated annual costs of USD $104 million per year [240].

Over the past few decades, *X. fastidiosa* has also presented a substantial obstacle for citrus farming in South America. In all major cultivars of sweet orange trees, *X. fastidiosa* subsp. *pauca* causes a disease known as citrus variegated chlorosis (CVC) [283, 288]. First observed in São Paolo in 1987, by 2005 CVC was observed in all the major citrus growing regions in Brazil, and it 2012 it was estimated that around 40% of all Brazilian citrus plants were infected [267, 285, 288]. Symptoms of CVC include the leaf scorch and stunted plant growth typically associated with *X. fastidiosa* infection, as well as the variegated yellowing of leaves that gives the disease its name [285, 288]. Although infected citrus trees do not usually die, the size and weight of their fruits are significantly reduced, resulting in potential yield losses of up to 80% [267, 285, 288]. Additionally, infected fruits become hard, and the content and quality of their juice is reduced [285, 289]. Since the Brazilian citrus industry accounts for 30% of orange production worldwide, as well as 85% of frozen orange juice concentrate [290], the economic implications of CVC are substantial; in 2012 it was estimated that the annual economic loss due to *X. fastidiosa* infection in Brazilian citrus was around USD $120 million [288].

Although *X. fastidiosa* is native and endemic to the Americas [291–293], the large scale global trade in live plant material means that the risk of the pathogen being introduced to new locations via imported infected hosts is high [259, 294]. Additionally, the range of potentially suitable xylem sap-feeding insect vectors of *X. fastidiosa* encompasses a variety of different species, that have a wide geographical distribution and appear able to

transmit different strains of the bacterium with little vector-strain specificity [267, 295]. Therefore, sustained local transmission of imported strains is a clear possibility in many parts of the world that have suitable climatic conditions for *X. fastidiosa* growth and multiplication.

X. fastidiosa infection has indeed previously been identified in many countries outside of the Americas, including France, Iran, Israel, Italy, Portugal, Spain, and Taiwan (Fig 5.1) [267, 292]. In some instances, these appearances have been localised and transient, but in others, the pathogen has taken hold in its new environment with worrying consequences. Of particular current concern is a strain of *X. fastidiosa* subsp. *pauca* known as CoDiRO (Complesso del Disseccamento Rapido dell'Olivo – loosely, 'rapid drying disease of olive trees'), which was discovered in the Apulia region of south-east Italy in 2013 [296]. Believed to have been introduced to the region via an infected ornamental coffee plant imported from Central America (most likely Costa Rica) in around 2008 [297, 298], the CoDiRO strain was subsequently identified as the causative agent of the Olive Quick Decline Syndrome (OQDS) affecting European olive trees (*Olea europaea*, see Fig 5.2A) in that area [260, 296]. Primarily transmitted by the meadow spittlebug *Philaenus spumarius*, CoDiRO spreads rapidly, is difficult to contain, and has devasting effects on infected trees; similarly to Pierce's Disease of grapevine, symptoms of OQDS typically include leaf scorch and defoliation, the dessication of twigs, branches, and fruits, and overall tree decline, resulting in death within a few years after symptom onset [280, 283]. OQDS therefore results in significant crop loss in affected olive groves.

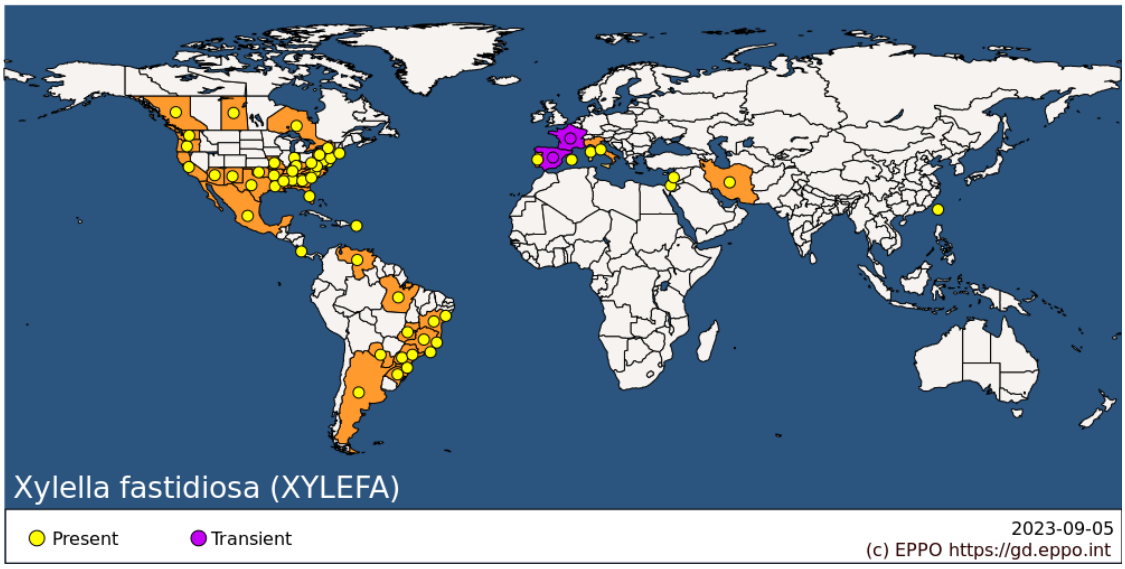


Figure 5.1: World map indicating countries and US states in which *X. fastidiosa* is present (orange shading, yellow dots) or has been recorded as transiently present (purple shading and dots), as of 5th September 2023. Map generated using the EPPO Global Database species distribution tool [299].

Since its discovery in Europe, the pathogen has spread through over 23,000 hectares of olive trees in the Salento peninsula [259] and continues to spread northwards, threatening olive farming throughout Apulia and beyond [300]. The areas at highest risk of *X. fastidiosa* invasion are olive-growing regions with climatic conditions similar to Apulia – in particular, the rest of Italy, Greece, and Spain, which together account for 93% of European olive production [280]. Recent projections indicate that the economic impact on olive farming in these countries alone could exceed EUR €24 billion over the next 50 years, if feasible control strategies are not devised [223]. The widespread loss of olive trees in the Mediterranean basin and the effect on the natural landscape may also threaten the economically crucial tourism industry in that region, as well as disrupting the local ecosystem [294, 301]. Since olive trees also have substantial cultural significance in these countries, the additional societal impacts associated with their loss cannot be ignored [283, 294, 301, 302].

In addition to olive production, the extremely wide host range of *X. fastidiosa* and the high risk of human mediated spread through the transportation of infected plant material means that this pathogen also threatens European agriculture on a much larger scale [259, 266, 280]. Widespread transmission of *X. fastidiosa* within the European Union (EU) territory could risk the production of many other economically valuable crops including grapevine, almond, cherry, and plum [266], leading to a projected annual production loss of EUR €5.5 billion and threatening over 300,000 jobs [303, 304]. For this reason, *X. fastidiosa* has been designated a 'priority pest' by the EU, with strict regulations governing the importation and movement of potential hosts within the EU territory, as well as procedures for demarcating infected areas and removing confirmed and potential infected hosts [303].

Despite intensive ongoing research, there is currently no effective treatment for OQDS, or *X. fastidiosa* infection more generally [280, 283]. Control methods in affected regions mainly consist of roguing confirmed infected plants and removing healthy plants in their vicinity, and reducing vector transmission using insecticides [260]. However, these interventions are costly, and must be swift in order to be effective [34, 259, 305]. The sooner control interventions are implemented, the less opportunity there is for the pathogen to be exported to new locations [34, 218]. Devising appropriate monitoring programmes to facilitate early detection is therefore critical to the success of containment strategies that aim to reduce the economic, ecological, and societal impacts of this disease [266].

Unfortunately, *X. fastidiosa* is a prime example of a pathogen for which the incubation period can provide a major obstacle to achieving early detection. Of particular relevance to the situation in Apulia, *X. fastidiosa* subsp. *pauca* has a long incubation period in *O. europaea* with a mean duration of around 15 months [266], and transmission by infected hosts can occur prior to symptom onset [267]. Therefore, the pathogen may spread widely

amongst olive trees in an affected area before the grower is alerted to the presence of infection. Although molecular tests are able to detect *X. fastidiosa* at low bacterial titres and could therefore theoretically be used to identify infection before symptom onset occurs [306, 307], the costs of large-scale presymptomatic sampling and laboratory testing are prohibitive [263]. Therefore, surveillance strategies for OQDS typically rely upon visual inspection of potential hosts as a first line of defence [263, 280]. The efficacy of current infection monitoring programmes is thus severely limited by the problem of presymptomatic transmission. However, despite its long incubation period in *O. europaea*, there is substantial variation in the incubation period of *X. fastidiosa* across its large host range, depending on factors such as plant species and age, pathogen subspecies, and climatic conditions [266, 267]. In particular, symptoms typically develop faster in small ornamental shrubs [266]. For this reason, the use of sentinel plants for surveillance of *X. fastidiosa* is a clear possibility, and has been identified as a key research area by the G20 Meetings of Agricultural Chief Scientists [267].

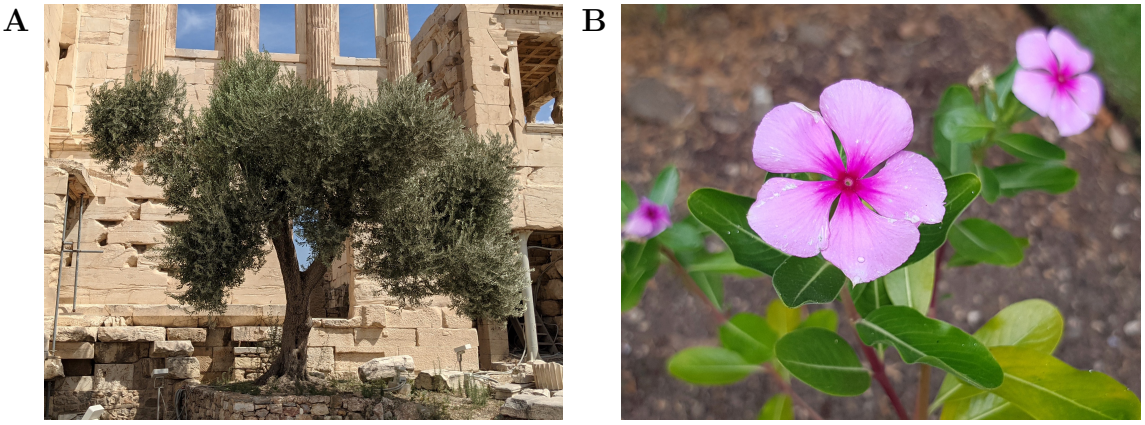


Figure 5.2: The plant species considered in our *X. fastidiosa* case study. A. *Olea europaea* (European olive) at the Acropolis (Athens, Greece). **B.** *Catharanthus roseus* (Madagascan periwinkle) at the University of Oxford Botanic Garden (Oxford, UK). *(Personal photographs)*

5.5 The contribution of Part II to the literature

The use of sentinel plants for surveillance has the potential to be a valuable strategy for plant disease management when presymptomatic transmission limits the effectiveness of existing monitoring programmes that rely on visual inspection as a means of detection. However, quantitative research into the effects of incorporating sentinel plants into surveillance programmes is lacking. In Part II of this thesis, we address this gap in the literature by conducting a mathematical investigation into how sentinel-based surveillance strategies could be used to provide a cost-effective approach to early detection monitoring for invasive plant pathogens, using *X. fastidiosa* infection in *O. europaea* (European olive) as a current

high-profile case study.

For OQDS, a natural candidate for a sentinel species is *Catharanthus roseus*, a herbaceous flowering plant commonly known as Madagascan periwinkle (Fig 5.2B). *C. roseus* is a known host of *X. fastidiosa* subsp. *pauca*, with a mean time from infection to symptom onset of around seven weeks (a factor of nine times shorter than *X. fastidiosa* in *O. europaea*) [266]. Although native to Madagascar, *C. roseus* has a wide geographical distribution and is already found across Italy and the rest of Europe, making it an ecologically suitable choice for a sentinel species [308]. Additionally, the small size of *C. roseus* plants allows them to be intercropped in olive groves. For these reasons, in this thesis we will consider *C. roseus* as the potential sentinel host, noting that our computational framework may easily be adapted to consider alternative sentinel species by adjusting the model parameters accordingly.

We begin in Chapter 6 by constructing a compartmental model of pathogen transmission that accounts for two different plant species, parameterised to be consistent with *X. fastidiosa* infection in *O. europaea* and *C. roseus*. Using this model, we perform stochastic simulations of pathogen transmission and surveillance to explore whether including sentinel plants in a surveillance programme could lead to a reduction in the expected prevalence of disease in the population at the time of detection, compared to a non-sentinel based monitoring programme of equivalent sampling effort. In particular, we examine the key trade-off between the potential benefits and drawbacks of sentinel plants – namely, whether the reduced detection prevalence that could be achieved by including sentinels in the population is sufficient to offset the corresponding increased rate of pathogen transmission. With this trade-off in mind, we demonstrate how the number of sentinel plants to include in the population can be optimised based on the amount of sampling resource available, and consider how limited surveillance resources can be optimally allocated between the crop and sentinel plants that are available for sampling.

In Chapter 7 of this thesis, we then extend the model developed in Chapter 6 to incorporate additional realism, accounting for seasonally varying factors that influence pathogen transmission and symptom expression. In particular, we consider seasonal fluctuations in the insect vector population, which lead to variations in the likelihood of pathogen introduction and the rate of pathogen transmission throughout the year. Additionally, we incorporate seasonal variations in temperature and precipitation, which influence the rate at which symptoms develop and the strength with which those symptoms are expressed. As well as confirming that our previous conclusions regarding the usefulness of sentinel plants to facilitate early detection still hold when these seasonal factors are accounted for, we adapt our sampling algorithm to allow for temporally varying surveillance strategies, and explore how limited sampling resources can be optimally allocated throughout the year. Specifically,

we seek to identify the optimal window of the year in which to focus surveillance effort, and compare this to the existing EU guidelines regarding the recommended time at which to perform visual inspection for *X. fastidiosa* infection.

Since a quantitative investigation of the potential for sentinel plants to improve early detection monitoring for invasive pathogens has not previously been attempted, the work presented in Part II of this thesis represents an important novel contribution to the existing literature on plant disease surveillance. Although here we use *X. fastidiosa* infection in *O. europaea* as a case study, the modelling framework we develop in this thesis could also be applied to investigate the use of sentinels against other invasive pathogens for which presymptomatic infection hinders existing monitoring approaches. Our results therefore provide a key evidence base for future researchers to build on when conducting investigations into sentinel-based surveillance strategies for a range of pathosystems.

Chapter 6

'Sentinel' plants as an early warning system for *Xylella fastidiosa* infection in *Olea europaea*

6.1 Introduction

As discussed in Chapter 5, the use of sentinel plants for surveillance represents an attractive possibility for improving early detection monitoring for invasive pathogens. However, although the idea of using sentinel plants for surveillance is an intuitive one, it is not immediately clear whether sentinels will actually have a substantive benefit when their advantages and disadvantages are weighed up. As mentioned previously, although the rapid onset of visible disease symptoms in sentinel plants may allow a pathogen to be detected quickly after invasion has occurred, introducing sentinel plants to a population also has a key potential downside: the resulting increase in population density and the rapid progression of disease in sentinels may lead to a corresponding increased rate of pathogen transmission, which could counteract the benefits of rapid detection. Therefore, achieving 'early detection' in the sense of rapidly identifying the pathogen following invasion does not necessarily equate to achieving 'early detection' in the sense of identifying the pathogen when it is still at a low prevalence in the population (the outcome that is practically beneficial to a grower or a policy maker).

There are many factors that could play an important role in determining whether a sentinel-based surveillance strategy is able to facilitate low-prevalence detection. Intrinsic epidemiological factors that influence pathogen transmission and symptom expression, such as the relative infectiousness of sentinel plants compared to crop plants and the relative durations of their presymptomatic periods, will naturally play a critical role in determining the effectiveness of a sentinel-based strategy. However, whether or not sentinel plants are able to facilitate detection at a low prevalence will also be influenced by practical choices made by the grower that determine how sentinel plants are incorporated into their monitoring programme. For example, a key question for a grower looking to implement a sentinel-based

strategy is how many sentinel plants should be added to the population. If too few sentinels are deployed, they are likely to be of limited benefit. On the other hand, including too many is likely to be detrimental due to the resulting increase in the transmission rate, particularly if the number of sentinels far exceeds the available capacity for sampling so that they cannot be fully exploited. The success of a sentinel-based strategy will therefore rely on the grower selecting an appropriate sentinel population size.

A second key question for a grower seeking to implement a sentinel-based monitoring programme is how to divide limited resources for sampling and inspection between the crop and sentinel plants in the population. Although an intuitive choice may be to preferentially allocate surveillance resources to sampling sentinel plants, it is not immediately clear whether this is sensible. In particular, if the number of sentinels in the population does not substantially exceed the sampling capacity, preferential sampling of sentinels could lead to the frequent repeated inspection of the same group of plants. Instead, it may be more beneficial for the sample selected in each surveillance round to consist of a mixture of crop and sentinel plants, that are selected at random and vary between sampling rounds. In some scenarios, by sampling different plants in each sampling round, more information may be gained by the grower. In order to design a successful sentinel-based surveillance strategy, the interacting effects of the different choices available to the grower need to be evaluated.

In this chapter, we conduct a first investigation into the potential for sentinel plants to facilitate low-prevalence detection of invasive plant pathogens, using *X. fastidiosa* infection in *O. europaea* and *C. roseus* as a case study. Since the effects of incorporating sentinel plants into a disease monitoring programme have never previously been assessed using epidemiological models, our key objective here is to construct the simplest possible quantitative framework for modelling pathogen transmission and surveillance in a two-species system, and to use it to investigate whether monitoring programmes that include sentinel plants can result in demonstrable differences in the detection prevalence compared to non-sentinel based surveillance strategies. Additionally, we aim to demonstrate how the optimal sentinel-based strategy can be identified for a given amount of surveillance resource, and explore how the optimal strategy is influenced by changes in that resource or in the intrinsic epidemiological properties of the pathosystem.

To that end, we construct a stochastic compartmental model of pathogen transmission that includes two different types of plants (crops and sentinels), and consider monitoring programmes in which, at regular intervals, a fixed number of plants are sampled at random and inspected for visible disease symptoms. A surveillance strategy is defined by: i) the number of sentinel plants added to the population; ii) the number of crops and sentinels to be examined in each sampling round, and; iii) the time interval between successive sampling

rounds. For a given surveillance strategy, we use model simulations to calculate the expected detection prevalence (EDP) in the crop population at the time of discovery. We investigate the conditions under which including sentinel plants in a surveillance programme allows us to attain a lower EDP than standard monitoring (i.e., the analogous surveillance strategy but without any sentinel plants). For a given choice of sample size and sample interval, we apply a Bayesian optimisation algorithm to determine the minimum attainable EDP and the surveillance strategy for which this is achieved.

We demonstrate that including sentinel plants in a surveillance programme has the potential to reduce the EDP compared to a standard monitoring programme of equivalent sampling effort. Sentinel plants are particularly beneficial when limited resources are available for plant disease surveillance. We show that both the total number of sentinels deployed and the division of the sample between crop and sentinel plants are crucial in determining the effectiveness of a surveillance strategy, and that it can be preferable to sample a mixture of both crop and sentinel plants, rather than exclusively sampling sentinels. Overall, our results demonstrate that sentinel plants are a useful tool to improve early detection monitoring, and encourage further research to identify the range of host-pathogen systems for which sentinel plants can reduce the damage caused by invading plant pathogens.

6.2 Methods

6.2.1 Transmission model

We considered a compartmental model of pathogen transmission in which plants are classified as 'Healthy' (H), 'Undetectable' (U) or 'Detectable' (D). The model includes two types of plant – crops (denoted by subscript C) and sentinels (denoted by subscript S). Healthy crops (H_C) and sentinels (H_S) are uninfected plants that are susceptible to infection. Undetectable crops (U_C) and sentinels (U_S) are plants that have been infected (and are infectious) but are not currently displaying visible symptoms. Once visible symptoms develop, plants progress into the Detectable class (D_C or D_S for crops and sentinels, respectively)[1]. A schematic illustrating the different compartments for both crop and sentinel plants is shown in Fig 6.1A.

We denote the total number of crop and sentinel plants in the population by $P_C = H_C + U_C + D_C$ and $P_S = H_S + U_S + D_S$, respectively, with a total population size of $P = P_C + P_S$. In each of our model simulations, fixed values of P, P_C, and P_S are used,

[1] The 'Healthy', 'Undetectable', and 'Detectable' classes in this model therefore respectively correspond to the 'Susceptible', 'Presymptomatic infectious', and 'Symptomatic infectious' classes used in the branching process models in Part I. The choice of labels 'Healthy', 'Undetectable', and 'Detectable' serves both to improve notational clarity (H, U, and D are clearly distinguishable from the subscripts C and S), and to emphasise to the reader which plants can and cannot be succesfully identified as infected via visual inspection at any given time. Note that in this model we assume that all infected plants go on to develop symptoms at some point during their infection, and therefore we do not include an 'Asymptomatic infectious' class here.

since we consider only the time until first detection and not the subsequent period during which infected plants and other nearby plants may be removed.

Undetectable and Detectable plants may generate new infections in any Healthy plant, with crops and sentinels assumed to be equally susceptible. The mode of transmission (insect vectors in the case of *X. fastidiosa*) is captured implicitly through the model parameterisation, rather than modelled explicitly. The parameters β_C and β_S represent the daily per capita rates at which individual infected Detectable crop and sentinel plants generate new infections, respectively. We also introduce the scaling parameters ε_C and ε_S to represent the relative infectiousness of Undetectable crops and sentinels compared to Detectable ones, so that the daily rates at which Undetectable crop and sentinel plants generate new infections are $\varepsilon_C\beta_C$ and $\varepsilon_S\beta_S$, respectively. In reality, Undetectable plants are unlikely to become infectious immediately after they are infected themselves. However, in the absence of detailed information on the time from infection to onset of infectiousness, we make a simplifying assumption that Undetectable plants are equally infectious throughout their presymptomatic period. To reflect the fact that Undetectable plants are likely to be less infectious than Detectable ones, we select $0 < \varepsilon_C, \varepsilon_S < 1$. The different transmission routes and the corresponding rates at which infections occur are illustrated in Fig 6.1B. The mean duration of the crop and sentinel Undetectable periods are given by the parameters γ_C and γ_S, respectively.

The resulting compartmental differential equation model representing pathogen transmission is given by:

$$\begin{aligned}
\frac{\mathrm{d}H_C}{\mathrm{dt}} &= -\phi H_C, & \frac{\mathrm{d}H_S}{\mathrm{dt}} &= -\phi H_S,\\
\frac{\mathrm{d}U_C}{\mathrm{dt}} &= \phi H_C - \frac{1}{\gamma_C}U_C, & \frac{\mathrm{d}U_S}{\mathrm{dt}} &= \phi H_S - \frac{1}{\gamma_S}U_S,\\
\frac{\mathrm{d}D_C}{\mathrm{dt}} &= \frac{1}{\gamma_C}U_C, & \frac{\mathrm{d}D_S}{\mathrm{dt}} &= \frac{1}{\gamma_S}U_S,
\end{aligned}$$

where the force of infection, ϕ, acting on each healthy plant is

$$\phi = \beta_C(\varepsilon_C U_C + D_C) + \beta_S(\varepsilon_S U_S + D_S). \tag{6.1}$$

This model is density dependent – that is, the transmission rate parameters β_C and β_S are not scaled by the size of the population. Therefore, any increase in the population density will result in an increased rate of pathogen transmission (in contrast to a frequency dependent model, in which the transmission rate is independent of the population density). This type of model is appropriate here, because although we are considering a crop population of fixed size, which is spread throughout a fixed area, when sentinel plants are added to this area the

density of the overall population increases – and the rate of pathogen transmission should increase accordingly. Indeed, this increase in transmission rate that occurs as a result of adding sentinel plants to the population is the key potential drawback of sentinels that we seek to investigate in this thesis, and as such this density dependence is a crucial aspect of our model.

In our analyses, we run simulations of the analogous stochastic model using the direct method version of the Gillespie SSA (Algorithm 1.1), as described in Section 6.2.5.

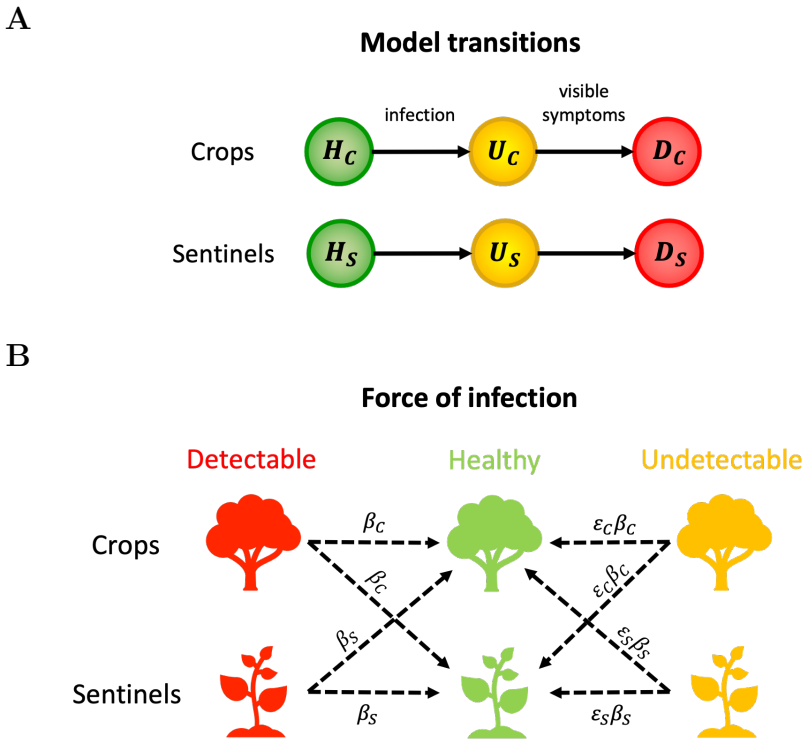


Figure 6.1: Schematic diagrams illustrating the compartments in the model (and how individual plants move between them) and the rates at which infections occur. A. Model transitions. Plants of either type (crop or sentinel), begin in the Healthy compartment (H_C or H_S) before moving to the corresponding Undetectable compartment (U_C or U_S) upon infection. Undetectable plants progress to the appropriate Detectable compartment (D_C or D_S) once visible symptoms develop. **B.** Force of infection. The rates at which different infectious hosts generate new infections. A Detectable crop plant can infect a Healthy crop or sentinel at rate β_C, whilst an Undetectable crop infects Healthy plants at the scaled rate $\varepsilon_C\beta_C$. Similarly, Detectable and Undetectable sentinels infect Healthy plants at rates β_S and $\varepsilon_S\beta_S$, respectively.

6.2.2 The 'baseline' case – reduced model in the absence of sentinel plants

In the absence of sentinel plants ($P_S = 0$), the transmission model reduces to:

$$\frac{\mathrm{d}H_C}{\mathrm{dt}} = -\phi_b H_C, \quad \frac{\mathrm{d}U_C}{\mathrm{dt}} = \phi_b H_C - \frac{1}{\gamma_C} U_C, \quad \frac{\mathrm{d}D_C}{\mathrm{dt}} = \frac{1}{\gamma_C} U_C,$$

where the force of infection, ϕ_b, acting on each Healthy plant is now given by

$$\phi_b = \beta_C(\varepsilon_C U_C + D_C). \tag{6.2}$$

Throughout this chapter, this reduced system is what we will refer to as the 'baseline case' for a particular choice of model parameterisation or surveillance strategy. It provides a point of reference for the EDP, allowing us to determine whether the EDP is reduced or increased when sentinel plants are introduced. Specifically, the effect of including sentinel plants in a surveillance strategy may then be assessed relative to the baseline case with equivalent parameterisation and sampling effort (i.e. the same total sample size and sample interval). When considering the baseline case, we again run model simulations of the analogous stochastic model using the direct method version of the Gillespie SSA (Algorithm 1.1).

6.2.3 Sampling and detection

We considered a monitoring programme in which different random samples of N plants are taken from the population every Δ days and inspected for symptoms of disease. We chose random sampling since, in a 'well-mixed' model that assumes spatial homogeneity in the infection risk across the landscape, it outperforms repeated sampling of the same individuals (see Supplementary Text C.1 and Supplementary Fig C.1), whilst being straightforward to implement computationally. For a given surveillance strategy (choice of N and Δ), we fix the number of crop plants and sentinel plants in the sample so that $N = N_C + N_S$, where N_C is the crop sample size and N_S is the sentinel sample size (in the baseline case, $N_S = 0$ and $N_C = N$). To reflect the introduction of the pathogen at a random time relative to the sampling scheme, we begin sampling from our model disease system at a time selected uniformly at random from the interval $[0, \Delta]$. We then sample every Δ days until detection occurs. In our analyses, we assume that Detectable plants in a sample are always correctly identified as being infected (visual inspection is 100% sensitive), Undetectable plants are never correctly identified as being infected (visual inspection is 0% sensitive), and Healthy plants are always correctly identified as uninfected (visual inspection is 100% specific) (Table 6.1). Therefore, detection occurs at a given time if and only if at least one Detectable plant (crop or sentinel) is included in the sample selected at that time. The effects of varying the detection sensitivity for Detectable crop and sentinel plants are considered in a supplementary analysis (see Supplementary Text C.2 and Supplementary Figs C.2, C.3, C.4).

Model class	Sensitivity	Specificity
Detectable (D_C, D_S)	100%	N/A
Undetectable (U_C, U_S)	0%	N/A
Healthy (H_C, H_S)	N/A	100%

Table 6.1: The assumed sensitivity and specificity of visual inspection as a means of detection for each of the three model classes (Detectable, Undetectable, and Healthy).

6.2.4 Model parameterisation

We selected the epidemiological parameters of our model (Table 6.2) to be consistent with *X. fastidiosa* subsp. *pauca* infection in the crop plant *O. europaea* (European olive) and the sentinel plant *C. roseus* (Madagascan periwinkle) [266, 309]. In our main analyses, we assumed that the transmission coefficient for Detectable sentinels (β_S) was equal to that for Detectable crops (β_C). In other words, infected sentinel plants were assumed to generate the same number of infections (on average) as infected crop plants. Whilst some studies have suggested that *C. roseus* may be more susceptible to *X. fastidiosa* infection than *O. europaea* [310], others have indicated that the rate at which the vector acquires the pathogen from *C. roseus* is lower than from *O. europaea* [311], and detailed information on how these conflicting factors affect the overall transmission rates is lacking. We were also required to assume the relative infectiousness of Undetectable sentinels compared to Detectable sentinels. The value chosen ($\varepsilon_S = 0.1$) is a conservative choice compared to the equivalent parameter for crop plants ($\varepsilon_C = 0.015$), since it assumes that Undetectable sentinels are substantially more infectious than Undetectable crops. This choice aims to provide a lower bound on the utility of sentinel plants. Due to the uncertainty in these parameter values, we performed extensive sensitivity analyses to determine how the choice of parameterisation affected our results (see Section 6.3.4, Supplementary Text C.3, and Supplementary Figs C.5-C.13). In each case that we considered, our overall finding – that sentinel plants can be helpful to reduce the EDP – was unchanged.

Throughout, we considered a crop population of $P_C = 1,000$ plants. The size of the sentinel population, P_S, varied between simulations. We initialised each simulation with a single Undetectable infected host, with the probability of that host being a crop or a sentinel plant determined by their respective proportions within the population.

Parameter	Meaning	Value	Justification
β_C	Transmission coefficient for Detectable crops	5×10^{-5}	Chosen so that $\beta_C P_C = 0.05$ [309]
β_S	Transmission coefficient for Detectable sentinels	5×10^{-5}	Assumed
ε_C	Transmission scaling factor for Undetectable crops	0.015	[309]
ε_S	Transmission scaling factor for Undetectable sentinels	0.1	Assumed
γ_C	Mean duration of crop Undetectable period	452 days	[266]
γ_S	Mean duration of sentinel Undetectable period	49 days	[266]

Table 6.2: The epidemiological parameters of the model, their meanings, and their baseline values chosen to be consistent with *X. fastidiosa* infection in *O. europaea* (crop) and *C. roseus* (sentinel). Other model parameter values are considered in Section 6.3.4, Supplementary Text C.3, and Supplementary Figs C.5-C.13.

6.2.5 Computational implementation

As noted above, for a given number of sentinels added to the population (P_S), we performed simulations of pathogen spread using the direct method version of the Gillespie SSA (Algorithm 1.1), generating stochastic epidemic curves in which the numbers of Undetectable and Detectable crops and sentinels were tracked over time until the entire population became infected (Fig 6.2A). Then, given the remaining parameters defining the surveillance strategy (values of N, N_C, N_S, and Δ), we implemented the corresponding monitoring programme on these simulated epidemics as described in Section 6.2.3 (Figs 6.2B, 6.2C). For each sampling run and subsequent detection completed on a unique epidemic curve, we recorded the total prevalence of infection in crop plants when the pathogen was discovered (i.e., we recorded the value of $U_C + D_C$ on discovery, as shown in Fig 6.2C). Repeatedly implementing the same surveillance strategy on many simulated epidemic curves, we obtained a distribution on the detection prevalence amongst crop plants for that surveillance strategy and computed the EDP as the mean value of that distribution (note that this does not include the prevalence amongst sentinel plants, since we assume that damage to the crop population is the primary concern for commercial growers) (Fig 6.2D). For any given choice of N and Δ, the baseline EDP for that strategy is computed in the same way, setting P_S (and thus also N_S) equal to zero. Throughout, the EDP for a given surveillance strategy (choice of P_S, N_C, N_S, and Δ) was calculated by performing sampling on $25,000$ simulated epidemic curves.

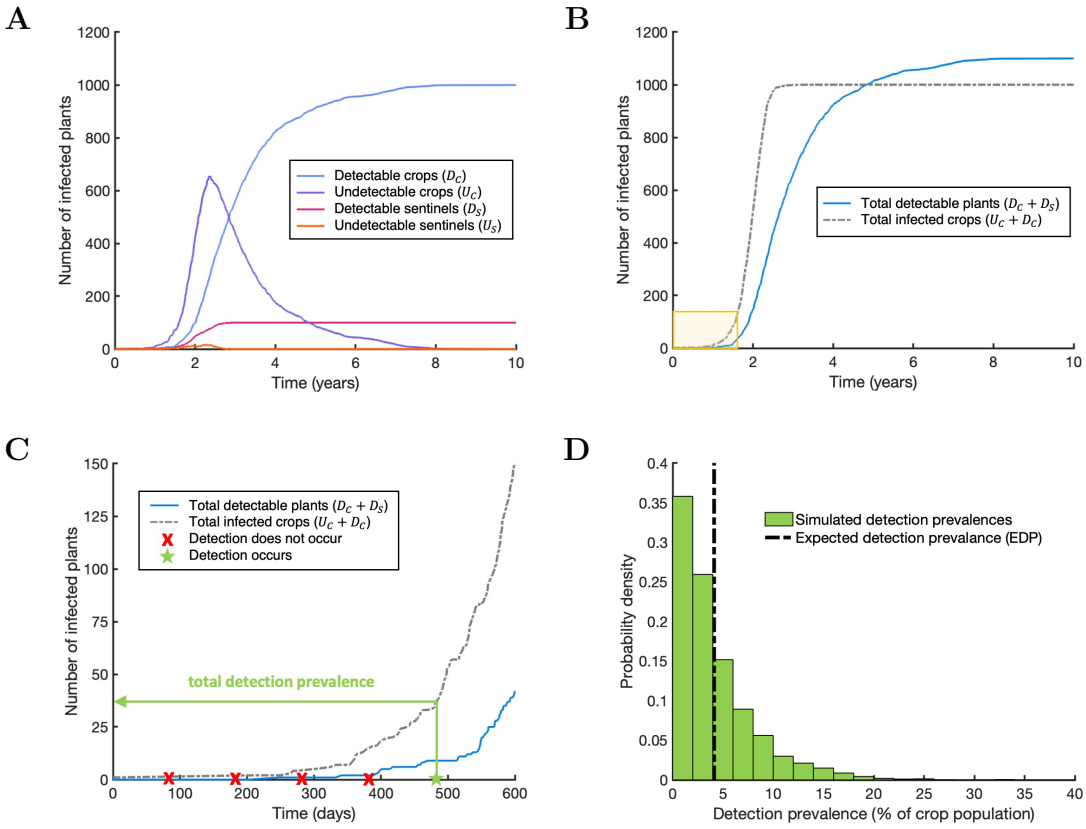


Figure 6.2: Schematic summarising how the EDP was obtained for an example sampling strategy. A. Stochastic simulations of pathogen spread in the model system were performed using the Gillespie SSA (direct method). In each simulation the number of Undetectable and Detectable crops and sentinels were tracked over time until the entire population became infected. In the simulations shown, we considered a population of $P_C = 1,000$ crops and $P_S = 100$ sentinels, beginning with a single Undetectable individual. All epidemiological parameters were as specified in Table 6.2. **B.** The total number of Detectable plants in the population ($D_C + D_S$; blue solid line) and the total number of infected crop plants in the population ($U_C + D_C$; grey dash-dotted line). Yellow box highlights the region that is shown magnified in panel C. **C.** A magnification of the yellow region highlighted in panel B, showing how the sampling strategy (defined in terms of the parameters N, N_C, N_S, and Δ) is implemented on an individual epidemic curve. Random samples of size $N = N_C + N_S$ were taken from the population every Δ days, with detection occurring if and only if a Detectable plant (D_C or D_S) was present in the sample. In this example, we used $N_C = 50$, $N_S = 50$, $\Delta = 100$ days. The time of the first sample was selected uniformly at random from the interval $[0, \Delta]$. When detection occurred, the total detection prevalence amongst crop plants ($U_C + D_C$) was recorded. **D.** The sampling strategy illustrated in panel C was repeated on $10,000$ simulated epidemic curves. Due to the inherent randomness of the spread and sampling simulations, different values for the total detection prevalence were obtained each time, leading to a probability distribution for this prevalence (green bars). The EDP was calculated as the mean of this distribution (black dash-dotted line).

Initially, we considered the effects of implementing a monitoring programme without sentinel plants (the baseline case described in Section 6.2.2). In the absence of sentinel plants, the monitoring programme requires selecting a random sample of $N_C = N$ plants at regular time intervals Δ, and checking for the presence of Detectable plants (D_C) in the sample. Doing so,

we computed the baseline EDP as outlined above, for sample sizes $N_C = 25, 30, 35, \cdots, 200$ and sample intervals $\Delta = 30, 35, 40, \cdots, 150$ days. The results of this analysis are shown in Section 6.3.1 (Fig 6.4).

Next, we considered three fixed values for the number of sentinels P_S added to the population: $P_S = 50$, $P_S = 100$, and $P_S = 200$. We allowed the total sample size N to take values $N = 25, 50, 75, \cdots, 200$, and considered five values for the sample interval: $\Delta = 30, 60, 90, 120$, and 150 days. In each case, we allowed the number of sentinels included in the sample (N_S) to be chosen in the range $[0, \min(P_S, N)]$, and applied a Bayesian optimisation algorithm (see Section 6.2.6) to identify the choice of N_S for which the EDP was maximally reduced compared to the baseline level. The choice of upper limit on N_S ensures that the number of sentinels sampled does not exceed the total sample size (N) or the total number of sentinels available (P_S). The results of this analysis are shown in Section 6.3.2 (Figs 6.5, 6.6, and 6.7).

Finally, we considered varying the number of sentinels in the population (P_S) and in the sample (N_S) simultaneously, and sought to identify the pair of values (P_S^*, N_S^*) that maximised the reduction in EDP compared to the baseline. In that case, we allowed the total sample size N to take values $N = 25, 30, 35, \cdots, 200$, and the sample interval to take values $\Delta = 30, 35, 40, \cdots, 150$ days. For each pair of (N, Δ) values, we allowed the total number of sentinels added to the population (P_S) to vary in the range $[0, 350]$, and the number of sentinels included in the sample (N_S) to vary in the range $[0, \min(P_S, N)]$. The upper bound of 350 on P_S was selected following trial simulations that indicated this would be sufficient to identify the optimal value of P_S across the (N, Δ) range considered. For each (N, Δ) pair, we applied a Bayesian optimisation algorithm to determine the optimal values of P_S and N_S, and constrained it to ensure that the number of sentinels included in the sample could not exceed the total number of sentinels available in the population (see Section 6.2.6). The results of this analysis are shown in Section 6.3.3 (Figs 6.8, 6.9, 6.10).

6.2.6 Bayesian optimisation and its application to this thesis

Bayesian optimisation is an efficient technique for finding the extrema of objective functions that are expensive to evaluate [312]. It is an iterative process in which successive (possibly noisy) observations of an objective function are taken at carefully selected trial points and used to update our beliefs about its likely form. Beginning from a prior distribution on the objective function (typically a Gaussian process prior [312, 313]), on each iteration the objective function is evaluated at a trial point and Bayes' theorem is applied to incorporate the observed result into a posterior distribution. The extrema of this posterior approximate those of the objective function with increasing accuracy on each iteration.

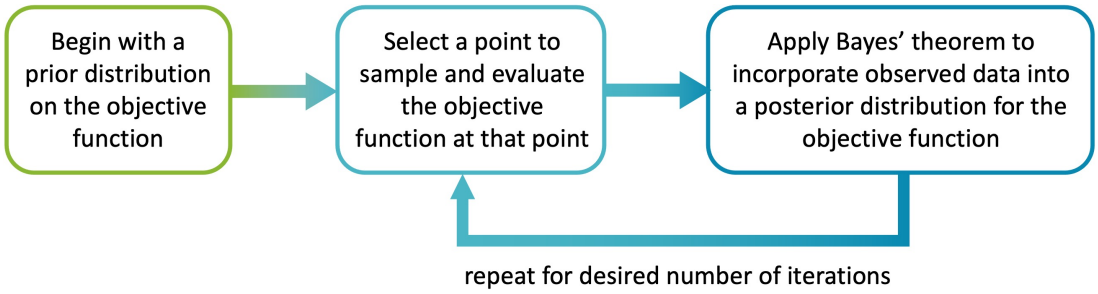


Figure 6.3: Schematic illustrating the basic process underlying Bayesian optimisation.

Successive trial points (points at which the objective function is evaluated) are selected by maximising an acquisition function that quantifies the potential gain of sampling at a particular point. When seeking to minimise the objective function, the acquisition function is typically defined so that it takes high values: i) near to previously sampled points where the objective function is known to be small (known as exploitation), and ii) where there is a large amount of uncertainty, and the objective function has the potential to be small (known as exploration). In contrast to the objective function, the acquisition function is cheap to evaluate and maximise, so the best trial point can be found easily. The judicious selection of trial points based on their utility makes Bayesian optimisation a particularly efficient method in terms of the number of evaluations of the objective function required, which is very useful when sampling is expensive [312]. Furthermore, Bayesian optimisation can be applied to an objective function that does not have a known analytic form, as long as the function can be evaluated computationally for a given choice of input parameters. It is therefore an ideal optimisation technique for this study, since our objective function is both analytically inaccessible and expensive to evaluate computationally.

In this thesis, we implemented Bayesian optimisation in MATLAB using the inbuilt function '*bayesopt*' [314, 315]. Since '*bayesopt*' seeks to minimise (rather than maximise) a specified objective function, we defined our objective to be the percentage change in EDP compared to the baseline level for a given (N, Δ) pair. Therefore, smaller (more negative) values of the objective corresponded to greater reductions in the EDP.

Initially, we fixed the number of sentinels added to the population (P_S) and considered the number of sentinels included in the sample (N_S) as the single optimisable variable, allowed to vary between 0 and the (constant) upper bound $\min(P_S, N)$ (see Section 6.3.2). We used the '*expected-improvement-plus*' acquisition function, which selects the next trial point (in this case, the value of N_S) so as to maximise the expected reduction in the objective function compared to the current estimated minimum, whilst also avoiding overexploitation of any given area [314]. For each choice of (N, Δ) we performed 30 iterations of the Bayesian optimisation algorithm (30 iterations is the *'bayesopt'* default); that is, we evaluated the

objective function for 30 successively selected values of N_S (not necessarily distinct). Each evaluation of the objective function involved computing the resultant EDP for the sampling strategy under consideration by performing sampling on $25,000$ simulated epidemic curves (as shown in Fig 6.2), and comparing it to the baseline EDP for the specified values of N and Δ.

Subsequently, we allowed both the number of sentinels added to the population (P_S) and the number of sentinels included in the sample (N_S) to vary simultaneously, and adapted the Bayesian optimisation algorithm to treat each of these as an optimisable variable (see Section 6.3.3). We allowed P_S to vary between 0 and 350, and N_S to vary between 0 and $\min(P_S, N)$ (now variable). Since the default implementation of '*bayesopt*' requires constant bounds for the optimisable variables, we first specified that N_S could vary between 0 and the total sample size N. Then, we further restricted the feasible region from which trial points could be drawn by enforcing the constraint $N_S \leqslant P_S$. This form of constraint (a deterministic function of the optimisable variables) can be implemented in MATLAB using the '*XConstraintFcn*' option in '*bayesopt*' [316].

6.2.7 Code availability

All computing code used to implement the above methods was written in MATLAB (version R2022a), and is available at https://github.com/francescalovellread/sentinel_plants.

6.3 Results

6.3.1 The baseline case – a monitoring programme without sentinel plants

We first considered the effects of implementing a monitoring programme without sentinel plants (the baseline case described in Section 6.2.2). A schematic of the reduced version of the model system in this case is shown in Fig 6.4A. In the absence of sentinel plants, surveillance involves selecting a random sample of $N_C = N$ crop plants at regular time intervals Δ, and checking for the presence of Detectable crops (D_C) in the sample (Fig 6.4B). As expected, lower EDPs are achieved with larger sample sizes N_C (inspecting more plants) and/or smaller sample intervals Δ (inspecting more frequently) (Fig 6.4C). These baseline values provide a point of comparison that we will use to evaluate the relative effects of sentinel-based strategies in subsequent sections. We also performed sensitivity analyses for different baseline parameter values (Supplementary Text C.3, and Supplementary Figs C.5, C.6). In every case that we considered, the qualitative behaviour of the baseline EDP as N and Δ were varied was unchanged.

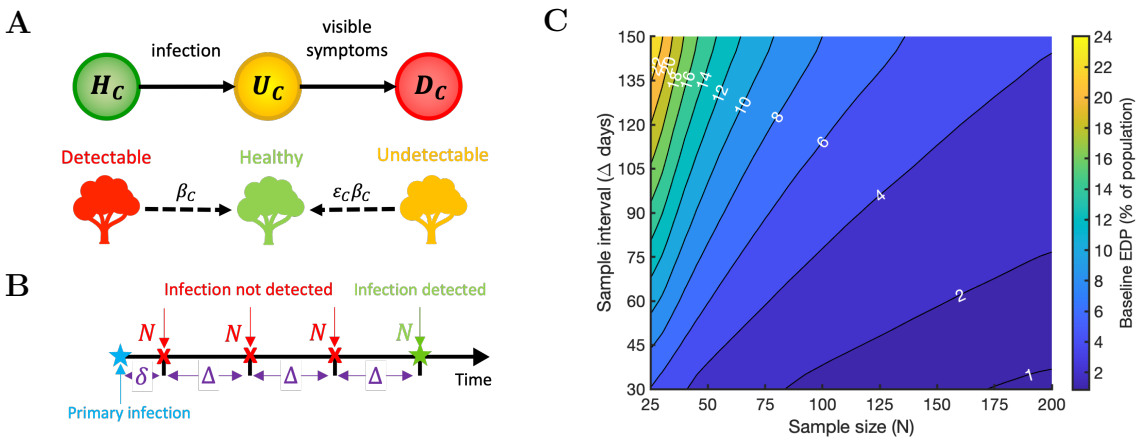


Figure 6.4: The baseline case – the model in the absence of sentinel plants. A. Schematic illustrating how crop plants progress through the model compartments, and the rates at which transmissions occur. Individual plants begin in the Healthy class (H_C), move to the Undetectable class (U_C) upon infection, and progress to the Detectable class (D_C) once visible symptoms develop. A Detectable crop infects Healthy crops at per capita rate β_C, whilst an Undetectable crop generates infections at the scaled per capita rate $\varepsilon_C\beta_C$. **B.** Schematic illustrating the implementation of the monitoring programme. Monitoring begins at a random time δ relative to the time of primary infection, where δ is drawn from a U$[0, \Delta]$ distribution. Random samples of size N are subsequently selected from the population at regular time intervals Δ. Infection is detected at a given time if and only if a Detectable plant is contained in the sample selected at that time. **C.** The baseline EDP, expressed as a percentage of the total crop population size, as the sample size ($N = N_C$) and sample interval (Δ) vary.

6.3.2 Introducing sentinel plants – choosing P_S and N_S carefully is critical

We next considered introducing sentinel plants to the population using the full model described in Section 6.2.1. This raises two important questions.

1. **How many sentinels should we add to the population (P_S)?** Although the relatively fast development of symptoms in sentinels facilitates the rapid detection of disease, this is only beneficial if the earlier discovery time corresponds to a lower EDP. Since adding sentinels will also increase the rate of pathogen transmission, including too many sentinels may negate the benefits of fast detection, particularly if (as assumed here) sentinel plants are more infectious than crop plants when Undetectable.

2. **How many of those sentinels should we include in the sample (N_S)?** Although a natural choice may be to sample preferentially from the available sentinel population (i.e. to include as many sentinels as possible in the sample), this is not necessarily optimal. For example, if the number of sentinels in the population is close to the sample size, this strategy would lead to frequent repeated sampling of the same set of plants, resulting in a reduction in the information gained per sample (see Supplementary Text C.1 and Supplementary Fig C.1).

In this section, we demonstrate how choosing P_S and N_S carefully is critical to avoid the introduction of sentinel plants having a negative effect, and to instead achieve the maximum possible reduction in EDP for a given sampling effort. As described in Section 6.2.5, we considered three fixed values for the number of sentinels added to the population ($P_S = 50$, $P_S = 100$, and $P_S = 200$) and a range of sample sizes (N) and sample intervals (Δ). For each combination of (P_S, N, Δ) that we considered, we ran a Bayesian optimisation algorithm (see Section 6.2.6) to identify the choice of N_S corresponding to the greatest reduction in EDP compared to the baseline value for that (N, Δ) pair. We denoted this optimal choice of N_S by N_S^*. For example, in the case $P_S = 50$, $N = 50$, and $\Delta = 30$ days, the optimisation indicated that the maximum reduction in EDP was achieved when $N_S^* = 17$ sentinels were included in each sampling round (out of a total possible maximum of 50) (Fig 6.5). This choice of sampling strategy (indicated by the green circle) led to a 16% reduction in the EDP compared to the baseline value. When N_S was instead chosen to take another of the values considered, smaller reductions (or even increases) in the EDP were achieved. In the extreme case in which none of the available sentinels were sampled, an 11% increase in the EDP was observed (y-intercept of Fig 6.5), demonstrating the drawback of adding sentinel plants to the population (an increase in transmission rate) if they are not sampled appropriately. The optimal number of sentinels N_S^* to include in the sample across the range of sample sizes (N) and sample intervals (Δ) is shown for $P_S = 50, 100$, and 200 in Figs 6.6A, 6.6C, and 6.6E respectively, with the corresponding reductions in the EDP compared to the baseline shown in Figs 6.6B, 6.6D, and 6.6F.

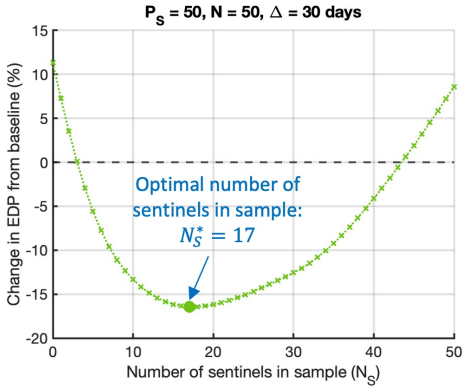


Figure 6.5: The effect of varying the number of sentinels included in the sample (N_S) on the percentage change in the EDP compared to the baseline level, in the case $P_S = 50$, $N = 50$, $\Delta = 30$ days. The number of sentinels in the sample for which the reduction in EDP is maximised ($N_S^* = 17$) is indicated by the green circle. Black dashed line marks the baseline EDP.

The optimal number of sentinel plants to include in the sample depended strongly on the sample interval and on the relationship between the sample size and the total number of sentinels available (Figs 6.6A, 6.6C, 6.6E). When $\Delta = 90$ days, 120 days, or 150 days, the

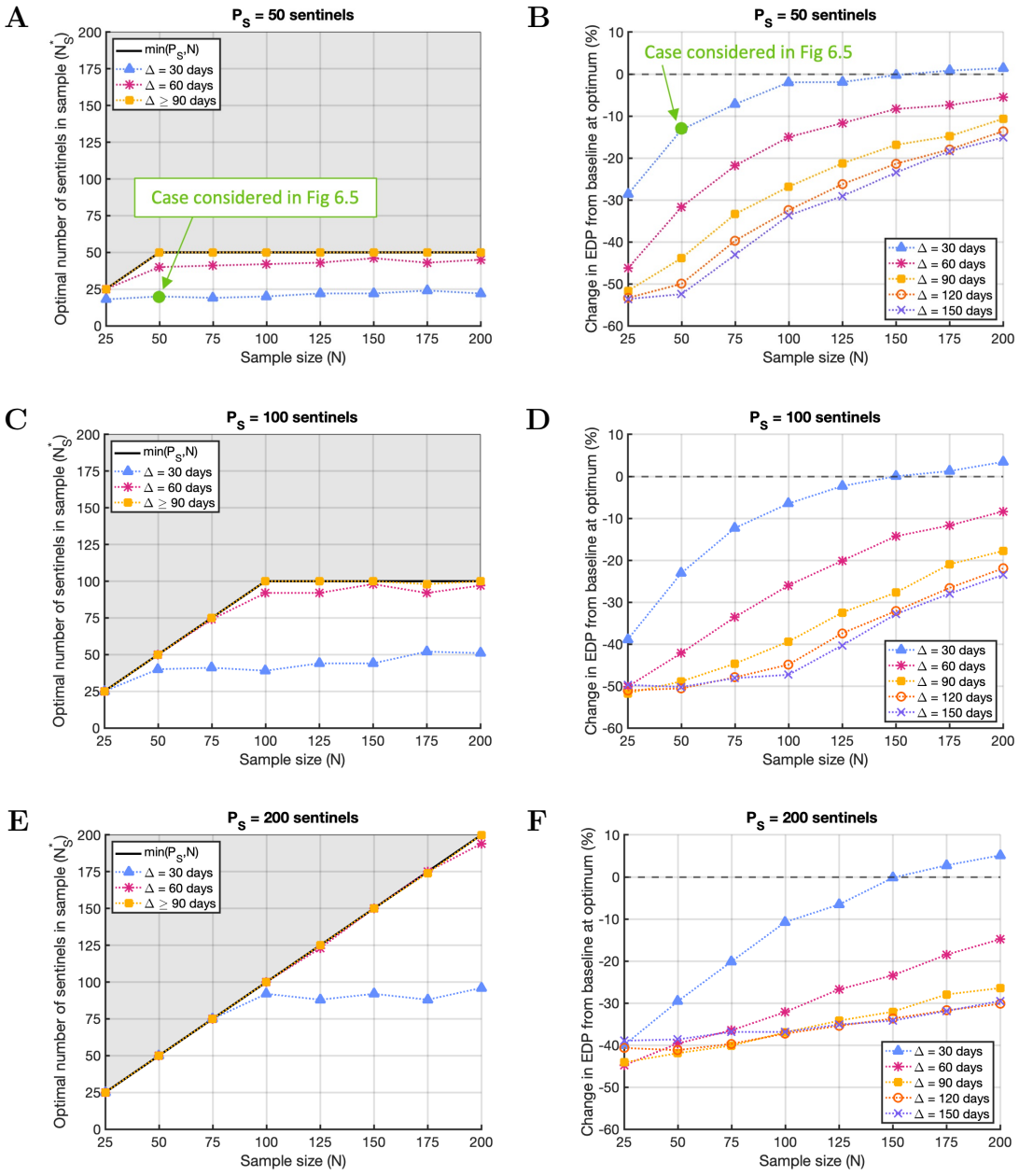


Figure 6.6: The optimal number of sentinel plants to include in the sample, and the correponding reductions in the EDP compared to the baseline level. A. The optimal number of sentinels N_S^* to include in the sample when $P_S = 50$, as the sample size (N) and sample interval (Δ) vary. Solid black line marks the maximum possible number of sentinels that could be sampled at any time ($\min(P_S, N)$). Grey shading marks the unfeasible region in which N_S exceeds this maximum. Green circle marks the case considered in Fig 6.5 ($P_S = 50$, $N = 50$, $\Delta = 30$ days). **B.** The best achievable percentage changes in the EDP compared to the baseline level for each (N, Δ) pair when $P_S = 50$, corresponding to the optimal strategies identified in Fig 6.6A. Green circle again marks the case considered in Fig 6.5. Note that the baseline level depends on N and Δ (Fig 6.4C, Supplementary Fig C.14A), so the relative changes in EDP shown here are not a measure of the resultant EDP. The resultant EDP decreases with sampling effort (Supplementary Figs C.14B, C.14C, C.14D). **C,D.** The analogous figures to panels A and B, respectively, but with $P_S = 100$ sentinels added to the population. **E,F.** The analogous figures to panels A and B, respectively, but with $P_S = 200$ sentinels added to the population.

optimal strategy in every case we considered was to sample the maximum possible number of sentinels (i.e. $N_S^* = \min(P_S, N)$). Since this result is identical for all three of those cases, they are represented by the single yellow line in Figs 6.6A, 6.6C, and 6.6E. However, for the shorter sample intervals $\Delta = 30$ and 60 days (blue and pink lines, respectively) the optimal monitoring strategy involved sampling a combination of sentinel plants and crop plants. In other words, in those scenarios it was preferable to sample fewer than the maximum allowable number of sentinels (i.e. $N_S^* < \min(P_S, N)$) for a range of choices of P_S and N, particularly when the total number of sentinels in the population was not substantially larger than the sample size.

These results may be explained by noting that, if the total number of sentinels available to sample from (P_S) is not substantially larger than the sample size (N), then sampling the maximum allowable number of sentinels ($\min(P_S, N)$) results in many or all of the same plants being repeatedly selected in every sampling round. If the sample interval is short, this leads to the frequent re-inspection of plants whose disease-free status was already established in the recent past, limiting the information gained per sampling round (see Supplementary Text C.1 and Supplementary Fig C.1). However, this effect diminishes as the sample interval increases, because the disease status of plants inspected in the previous sample becomes less informative of their state at the next sample time. Thus, when the sample interval is large, sampling the maximum possible number of sentinels is the optimal strategy ($N_S^* = \min(P_S, N)$) regardless of the sample size (N) or the total number of sentinels available (P_S). These results confirm the need to consider the division of the sample between crops and sentinels as a variable quantity that should be chosen carefully based on the precise conditions under which surveillance is taking place. If the number of sentinels included in the sample is suboptimal, smaller reductions in the EDP will be achieved, and sentinel plants may even have a detrimental effect (Fig 6.5).

For almost all values of P_S, N, and Δ that we considered, when the number of sentinels included in the sample was optimised (as indicated in Figs 6.6A, 6.6C, and 6.6E), a reduction in the EDP compared to the baseline value was achieved (Figs 6.6B, 6.6D, and 6.6F). Exceptions to this occurred when the sample interval was $\Delta = 30$ days and the sample size was $N \geqslant 150$; in those cases we observed a slight increase in the EDP compared to the baseline (blue lines in Figs 6.6B, 6.6D, and 6.6F). This can be attributed to the fact that, in this initial analysis, we considered only a limited number of values of P_S (specifically, $P_S = 50, 100$, or 200). If instead P_S can be chosen to take any value, the baseline EDP would never be exceeded at the optimised value of P_S, as at worst the baseline EDP can always be achieved by setting $P_S = 0$. These results demonstrate that, in addition to the proportion of sentinels included in the sample, the total number of sentinels added to the population must also be carefully selected: if a suboptimal value of P_S is chosen, sentinel plants can be detrimental rather than beneficial.

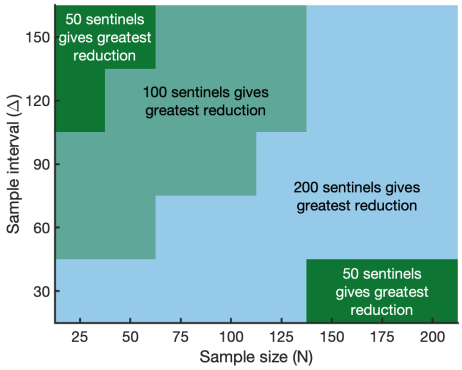


Figure 6.7: The optimal number of sentinels to include in the population. Combinations of the sample size N and sample interval Δ for which adding $P_S = 50$ (dark green), $P_S = 100$ (light green), or $P_S = 200$ (blue) sentinels to the population led to the greatest reduction in the EDP compared to the baseline level (of the three values of P_S considered).

As expected, the resultant EDP following the implementation of the optimal sentinel strategy decreased with greater sampling effort: taking larger samples and/or sampling more frequently always led to a lower EDP (Supplementary Fig C.14). However, larger percentage reductions in the EDP relative to the baseline level were mostly achieved when the sampling effort was low (i.e. when the sample size was small and/or the sample interval was large) (Figs 6.6B, 6.6D, 6.6F). This is because, when the sampling effort was low, the baseline EDP was much higher to begin with (Fig 6.4C). In those cases, the potential for the use of sentinel plants to lead to a large relative improvement in the EDP was greater than when the sampling effort was high and the baseline EDP was already low.

As well as affecting the magnitude of the reduction in EDP compared to the baseline, the choice of sample size and sample interval also affected the total number of sentinels for which the greatest reduction was achieved (Fig 6.7). For example, when the sample size was $N = 25$ and the sample interval was $\Delta = 150$ days, choosing $P_S = 50$ led to the greatest reduction in EDP of the three values considered (54%, compared to a 50% reduction when $P_S = 100$ and a 39% reduction when $P_S = 200$). However, for $N = 100$ and $\Delta = 30$ days, choosing $P_S = 200$ gave the greatest reduction in EDP (11%, compared to 2% and 6% when $P_S = 50$ and 100 respectively). Overall, introducing fewer sentinels was preferable when the sampling effort was either low or very high, with larger numbers preferable for intermediate sampling efforts (Fig 6.7).

This variation in the optimal number of sentinels for different values of (N, Δ) reflects the crucial trade-off between the benefits and drawbacks of sentinel plants. Although adding sentinels to the population helps to facilitate early detection, it also leads to an increased rate of pathogen transmission (particularly if sentinels are more infectious than crop plants when

Undetectable, as assumed here). Therefore, including more sentinel plants is only beneficial if the advantage gained from sampling them outweighs the impact of increased transmission. For small sample sizes N, the capacity to exploit large numbers of sentinel plants is limited. Although increasing the number of sentinels is beneficial up to a point, since it allows for sampling without frequently inspecting the same sentinel plants, there is a threshold number of sentinels to introduce beyond which there will be no further improvement in detection to counterbalance the concurrent increase in overall transmission. The benefit of increasing P_S is more limited when the sample interval Δ is large, since in that scenario a past negative sample is less likely to indicate that the current sample will be negative. Thus, the same sentinels may be resampled without a substantial correlation between successive samples. Smaller numbers of sentinels are therefore preferable when the sampling effort is low (Fig 6.7). At the opposite extreme, when sampling is very intensive (large sample size N and small sample interval Δ) the baseline EDP is low (Fig 6.4C) and the potential for sentinel plants to reduce it is limited. In such a case, including a very large number of sentinels in the population is also not optimal, since this limited reduction is outweighed by the consequent higher rate of transmission. Therefore, smaller numbers of sentinels are also preferable when the sampling effort is very high (Fig 6.7). However, for intermediate sampling efforts, larger numbers of sentinels perform better, since the capacity to exploit them and the scope to reduce the EDP compared to the baseline are less restricted. These results emphasise that judicious selection of the total number of sentinel plants added to the population is required to ensure that the benefits of including them are sufficient to offset their drawbacks in terms of increasing transmission. This emphasises the need for an epidemiological modelling framework as provided here to guide the number of sentinel plants to introduce, and we explore how the number of sentinel plants can be optimised in the next section.

6.3.3 Optimising the number of sentinel plants included in the population

We next considered optimising the total number of sentinel plants added to the population (P_S) and the number of sentinels included in the sample (N_S) simultaneously. The total number of sentinels added to the population (P_S) was allowed to vary in the range $[0, 350]$, whilst the number of sentinels included in the sample (N_S) was allowed to vary in the range $[0, \min(P_S, N)]$ (since N_S cannot exceed either the total number of sentinels available to sample from, P_S, or the total sample size, N). As outlined in Section 6.2.5, for each (N, Δ) pair considered, we applied a constrained Bayesian optimisation algorithm to identify the values (P_S^*, N_S^*) that reduced the EDP most compared to the baseline value.

Including sentinel plants in the population was beneficial across the range of sampling strategies considered, with the optimal number of sentinels added to the population (P_S^*) greater than zero for all values of (N, Δ) (Fig 6.8). This shows that deploying sentinel plants has the potential to reduce the EDP. However, the optimal numbers of sentinels to use varied

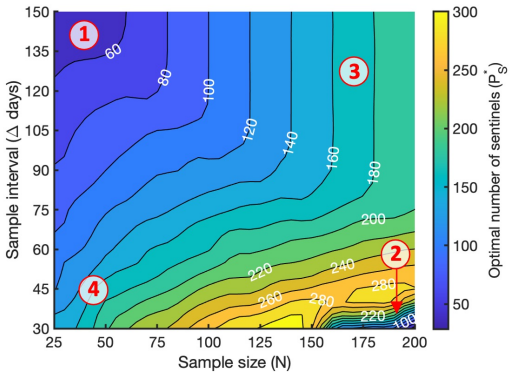


Figure 6.8: The optimal number of sentinel plants to include in the population, P_S^*, for which the maximal reduction in the EDP compared to the baseline level is achieved (if N_S is also chosen optimally). Region 1: When the sampling effort was small, P_S^* was low. P_S^* increased for larger sample sizes (moving to the right on the figure) and smaller sample intervals (moving downwards on the figure). **Region 2:** P_S^* dropped again when the sampling effort was very high. In that region, the baseline EDP was very low (Fig 6.4C), and the scope for reducing it insufficient to offset the increase in transmission rate caused by adding large numbers of sentinel plants to the population. **Region 3:** When the sample interval (Δ) was large, the optimal number of sentinel plants to include in the population was equal to the sample size (N) (contour lines are vertical). In that region, the sample interval was long enough to allow for repeated sampling of the same plants, eliminating the need for P_S to exceed the sample size. **Region 4:** When the sample interval (Δ) was small, the optimal total number of sentinels in the population was substantially larger than the sample size. In that region, a large sentinel population was necessary to avoid frequent repeated sampling of the same plants.

substantially with the sample size (N) and sample interval (Δ). Consistent with Fig 6.7, P_S^* was low when the sampling effort was small (region 1 in Fig 6.8), with the optimal number of sentinels increasing for larger sample sizes and smaller sample intervals. This is again due to the benefit of avoiding repeated sampling of the same plants when the sampling effort is high. By including more sentinels, the chance of repeatedly sampling the same sentinel plants is reduced. As in Fig 6.7, P_S^* dropped again when the sampling effort was very high (region 2 in Fig 6.8). In that region, the baseline EDP was very low (Fig 6.4C), and the scope for reducing it insufficient to offset the increase in transmission rate caused by adding large numbers of sentinel plants to the population. When the sample interval (Δ) was large, the optimal number of sentinel plants to include in the population was equal to the sample size (N) (region 3 in Fig 6.8 in which contour lines are vertical). In that region, the sample interval was long enough to allow for repeated sampling of the same plants, eliminating the need for P_S to exceed the sample size.

For almost all of the (N, Δ) values considered, the optimal number of sentinels to include in the sample (N_S^*) was the maximum possible (i.e. $N_S^* = \min(P_S, N)$) (Fig 6.9). That is, it was almost always preferable to first fill the sample with as many sentinel plants as possible, before using any remaining capacity to sample crop plants. The only exceptions to this

occurred when the sample size N was large, and the sample interval Δ was small (region 1 in Fig 6.9). To explain these results, we recall from Section 6.3.2 that the key reason why preferential sampling of crop plants may not be optimal is that it may lead to the frequent repeated sampling of the same set of plants. When the sampling effort was very high (large N and small Δ; region 1 in Fig 6.9), the total number of available sentinels dropped below the sample size (see region 2 in Fig 6.8). Therefore, preferentially filling the sample with sentinels would result in repeated sampling of the same plants on each sampling round. In that case, the small sample interval would result in a high correlation between successive samples, reducing the amount of information gained per sample. However, since the sample size substantially exceeded the number of sentinels in that region (see region 2 in Fig 6.8), randomly selected crop plants still made up the majority of the sample. Therefore, repeated sampling of the same sentinels had a relatively small effect and the consequent reduction in the optimal proportion of sentinels to sample was not large. When the sample interval and sample size were both small (region 2 in Fig 6.9), the optimal total number of sentinels in the population was substantially larger than the sample size (see region 4 in Fig 6.8). Therefore, preferential sampling of sentinels was optimal, since different sentinels could be used to fill the sample on each sampling round and repeated sampling of the same plants could be avoided. Finally, when the sample interval (Δ) was large (region 3 in Fig 6.9), information about the infection status of a given plant that was gained on the previous sampling round became much less informative of the current status of that plant. Therefore, the same plants could be sampled repeatedly on each sampling round, whilst still gaining new information about pathogen transmission each time. Thus, for large Δ, preferentially sampling sentinels was always optimal, regardless of the relative sizes of the sample and the sentinel population.

When P_S and N_S were chosen optimally, reductions in the EDP compared to the baseline value were achieved for almost all (N, Δ) values considered (Fig 6.10A). Of course, since the baseline itself (Fig 6.4C) can theoretically always be achieved by choosing $P_S = 0$, we would not expect the optimal resultant EDP to exceed the baseline substantially for any choice of N and Δ. However, in cases where the optimal sentinel strategy has little effect on the EDP, small increases in the EDP compared to the baseline may still occur due to the stochasticity of our simulations. This was observed for some very large values of N and very small values of Δ (Fig 6.10A). As in all previous cases, the resultant EDP decreased with greater sampling effort: increasing the sample size (N) and/or decreasing the sample interval (Δ) always led to a smaller EDP (Fig 6.10B).

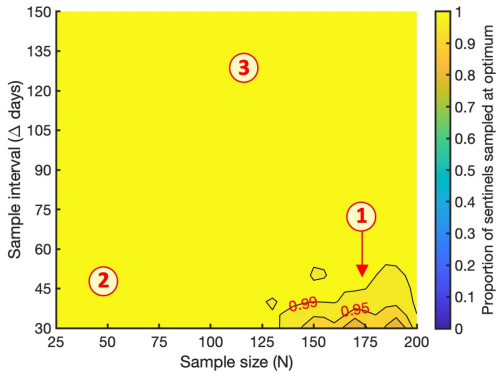


Figure 6.9: The optimal proportion of available sentinels to include in the sample as the sample size (N) and sample interval (Δ) vary, when $P_S = P_S^*$ as shown in Fig 6.8. For each (N, Δ) pair, this proportion is computed as the optimal number of sentinels to include in the sample (N_S^*) divided by the maximum possible number of sentinels that could be included ($\min(P_S^*, N)$). **Region 1:** When the sampling effort was very high (large N and small Δ), the number of sentinels available in the population was low. Therefore, sampling fewer than the maximum possible number of sentinels was preferable in order to reduce the impact of repeatedly sampling the same plants. **Region 2:** When the sample interval and sample size were both small, the optimal total number of sentinels in the population was substantially larger than the sample size, and preferential sampling of sentinels was optimal. **Region 3:** when the sample interval was large, preferentially sampling sentinels was always optimal. Here, the correlation between successive samples was low and thus the same plants could be sampled repeatedly on each sampling round without loss of information.

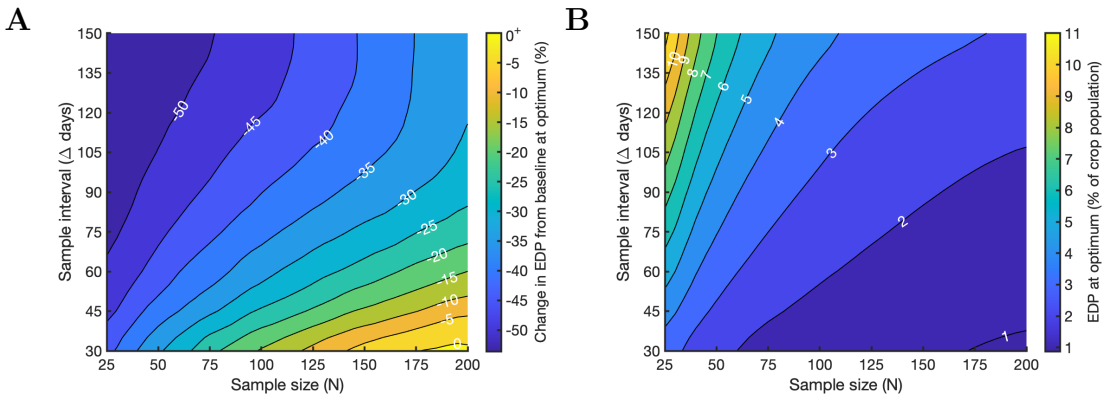


Figure 6.10: The effect of optimising the number of sentinels to include in the population and in the sample on the EDP. A. The percentage change in the EDP compared to the baseline value at the optimum, achieved when $P_S = P_S^*$ and $N_S = N_S^*$. **B.** The resultant value of the EDP at the optimum, expressed as a percentage of the total crop population.

6.3.4 Robustness of the results to the parameter values used

As far as possible, the epidemiological parameters used in our analyses were chosen based on literature estimates for *X. fastidiosa* infection in *O. europaea* and *C. roseus* (Table 6.2). However, reported estimates were not available for all parameters. In particular, we were required to assume the transmission coefficient for Detectable sentinels (β_S) and the transmission scaling factor for Undetectable sentinels (ε_S) (see Section 6.2.4). Therefore, we also conducted supplementary analyses to determine how variation in the model parameters affected our results. These analyses are crucial for demonstrating the generalisability of our results. As well as varying β_S and ε_S, we considered variation in the transmission scaling factor for Undetectable crops (ε_C), the mean durations of the crop and sentinel Undetectable periods (γ_C and γ_S), the crop population size (P_C), and the initial number of Undetectable infected individuals (U_0). For each of these, we generated plots analogous to Figs 6.8, 6.10A, and 6.10B for two different values of the relevant parameter (Supplementary Text C.3 and Supplementary Figs C.7-C.13).

Although the optimal total number of sentinel plants to include in the population (P_S^*; panels A and B in each of Supplementary Figs C.7-C.13) exhibited some variation with the changes in parameter values, the overall results remained qualitatively similar in most respects. In half of the cases we considered, the main qualitative change was that the drop in P_S^* observed in region 2 of Fig 6.8 did not occur (Supplementary Figs C.7A, C.8A, C.8B, C.11A, C.12B, C.13A, C.13B). In general, this arose in cases in which sentinel plants became relatively more beneficial due to the parameter change – for example, when the duration of the sentinel Undetectable period (γ_S) or the relative infectiousness of Undetectable sentinels (ε_S) was reduced. In some other cases, when the parameter change led to sentinel plants becoming less beneficial, such as increasing the transmission coefficient for Detectable sentinels (β_S) or the relative infectiousness of Undetectable sentinels (ε_S), then smaller numbers of sentinels were preferable (Supplementary Figs C.7B, C.9B, C.10B, C.11B). The optimal number of sentinels was also predictably reduced when the crop population size was halved ($P_C = 500$; Supplementary Fig C.12A), confirming that sentinel-based strategies must be assessed in context and tailored to the specific crop population being considered.

As expected, larger percentage reductions in the EDP compared to the baseline were observed for parameter changes that increased transmission amongst crop plants or that made sentinel plants relatively more beneficial (Supplementary Figs C.7C, C.8C, C.8D, C.11C, C.13C, C.13D). This included increasing the relative infectiousness of Undetectable crops, increasing the initial number of infected hosts, and decreasing the infectiousness of sentinels or the duration of their Undetectable period. Similarly, for parameter changes that made sentinel plants relatively less beneficial, such as increasing the duration of their Undetectable period or their relative infectiousness, smaller reductions in the EDP were observed (Supplementary

Figs C.7D, C.9D, C.11D).

The resultant EDP at the optimum (panels E and F in each figure) remained qualitatively similar with variation in each parameter. In every case, increasing the sampling effort (increasing the sample size N and/or reducing the sample interval Δ) led to a smaller resultant EDP. The resultant EDP also remained quantitatively similar for almost all parameter changes, with the only substantial change occurring when the transmission scaling factor for Undetectable crops was increased from $\varepsilon_C = 0.015$ to $\varepsilon_C = 0.25$ (Supplementary Fig C.8F).

6.4 Discussion

An important challenge in plant disease management is detecting invading pathogens before they become widespread in the population. In this chapter, we have considered how early detection of invasive pathogens can be aided by the introduction of sentinel plants, which are alternative species that display symptoms of infection more quickly than the main at-risk species. We have explored the trade-off between faster detection of the pathogen using sentinel plants and the potential for sentinel plants to increase transmission, which is likely if the onset of symptoms is associated with higher infectivity. Overall, our results indicated that sentinel plants have the potential to facilitate substantial reductions in the EDP compared to the baseline level for a wide range of sampling efforts (Figs 6.6B, 6.6D, 6.6F, 6.10A).

The practical benefits of a reduced EDP are multiple. A lower prevalence of infection upon pathogen discovery minimises the number of plants that must subsequently be removed, with benefits including lowering the total value of damaged crops, a lower reduction in crop yield, and a reduced logistical cost of plant removal. Fast discovery of an invading pathogen also limits the opportunity for dispersal to other locations, meaning that incorporating sentinel plants into surveillance programmes may be beneficial for disease containment on a broader scale [34, 218]. Discovering the pathogen at a low prevalence may increase the chance that local eradication is feasible, although evaluating whether the benefits of sentinel plants are sufficient to allow this would require control strategies to be accounted for explicitly in our model.

In addition to the overall potential of sentinel plants to facilitate early detection, our results highlight that both the optimal number of sentinel plants to introduce to the population and the optimal proportion of sentinel plants to include in each sample vary according to the sampling resources available (Figs 6.6, 6.7, 6.8, 6.9). The deployment of sentinel plants can be substantially less beneficial, or even detrimental, if either of these quantities are chosen

without carefully assessing the impact of the monitoring programme in advance (Figs 6.5, 6.6B, 6.6D, 6.6F).

In this chapter, our focus was investigating the degree to which including sentinel plants in a surveillance programme could reduce the EDP. Although our analyses suggest that sentinel plants may be effective for this purpose, there is clear motivation to consider these results within a wider economic context. In practice, an important consideration is the cost of implementing the surveillance programme, which must be compared against the benefits of fast detection of the invading pathogen. Further research is needed to quantify the economic costs of surveillance and control, accounting for the cost of sampling as well as the cost of removing infected plants and the resulting loss in plant value. The specific objective of the control strategy also needs to be carefully considered. For example, if reducing the opportunity for the pathogen to be exported to a new location prior to detection is a particular focus, the decision maker may wish to consider the 'global' epidemic cost associated with pathogen exportation in addition to the 'local' cost incurred in the controlled region [34, 253], or to seek surveillance strategies that minimise the detection prevalence in the sentinel population as well as the crop population (for a preliminary analysis, see Supplementary Text C.4 and Supplementary Fig C.15).

Although here we considered *X. fastidiosa* infection in *O. europaea* and *C. roseus* as a case study, our model provides a general framework that can be used to assess sentinel-based surveillance strategies in other scenarios. For example, we could consider alternative sentinel species by adjusting the model parameters appropriately. One possibility for an alternative sentinel species is alfalfa, a known host of many *X. fastidiosa* strains [311]. A benefit of alfalfa is that it is itself economically valuable, which may increase its attractiveness to growers. However, alfalfa has a longer presymptomatic period than *C. roseus*, of around 6-9 months [317], which could limit its effectiveness as a sentinel plant. An exploration of this trade-off would also require the incorporation of economic considerations in our epidemiological modelling approach. Our framework could similarly be adapted to apply to alternative pathosystems. For example, as mentioned in Chapter 5, monitoring programmes for citrus greening disease are hindered by a long period of presymptomatic infection in mature trees [268, 269]. Younger trees develop symptoms more quickly [268, 272], but also tend to support more rapid growth rates in the vector population [318, 319]; our model could be used to explore their suitability for use as sentinel plants, taking this drawback into account.

An important limitation of our modelling framework is that we did not include vector dynamics explicitly. Since our goal was to use the simplest possible model to investigate the potential of sentinel plants to facilitate early detection of invasive pathogens, we instead

assumed that vector behaviour was captured implicitly through the model parameterisation (specifically, through the parameters β_C and β_S, which characterise the rate at which the pathogen is transmitted between hosts). Whilst a similarly implicit characterisation of vector behaviour has been used in previous modelling studies (see, for example, [34, 260, 309]), it does limit the possibility of altering the assumptions underlying the vector dynamics and observing the resulting effects on epidemiological dynamics. For example, our model assumes that vectors are equally attracted to crop and sentinel plants. If instead vectors exhibit a strong preference for crop plants, or are simply more likely to feed on them due to their much larger size, the relative rate at which sentinel plants become infected may be too low for them to be useful as early warning beacons for infection. On the other hand, if vectors preferentially feed on sentinel plants their benefit may be greater, as the rate at which crop plants become infected will be reduced. Similarly, we assume that vectors make no distinction between hosts that are displaying symptoms and hosts that are not. This is not necessarily the case – for example, vectors of *X. fastidiosa* in grapevine appear to prefer feeding on nonsymptomatic hosts rather than on symptomatic ones [320]. Such a preference may alter the transmission dynamics between crop and sentinel plants, since sentinels are assumed to remain nonsymptomatic for a shorter period of time than crop plants. These dynamics can be approximated in our model by altering the transmission rate parameters (see Supplementary Fig C.16 for an example in which the susceptibility to infection of sentinels is half that of crop plants). However, extending our model to include explicit vector compartments would allow us to perform a thorough investigation into the influence of vector ecological dynamics on the effectiveness of sentinel-based surveillance strategies.

Throughout, we assumed that visual inspection was perfectly sensitive for identifying infection in Detectable plants and not at all sensitive for Undetectable plants, with no distinction between crop and sentinel plants (see Section 6.2.3). However, allowing for false negatives in Detectable plants and/or positives in Undetectable plants would affect the EDP and the optimal number of sentinels to include in the population for a given sample size and sample interval. If the chance of testing errors differed between crop and sentinel plants, the optimal allocation of sampling effort between crops and sentinels would vary. Although a detailed exploration of the impact of varying detection sensitivities and specificities is beyond the scope of this chapter, a preliminary analysis in which we consider the effects of allowing the sensitivity of detection to vary between Detectable crop and sentinel plants is presented in the Supplementary material (see Supplementary Text C.2 and Supplementary Figs C.2, C.3, C.4).

Our model could also be extended to incorporate additional realism in other ways. For example, we could extend the transmission model considered here to account for the spatial structure of the population, allowing the likelihood of transmission between any two plants to depend on the distance between them [34, 35, 249, 252, 260, 265]. Considering a spatially

heterogeneous model would raise additional questions regarding the optimal spatial placement of sentinel plants and their selection as part of a monitoring programme, particularly if the risk of primary invasion also varied in space [321]. In addition to spatial heterogeneities, we could also consider temporal heterogeneities in the probability of invasion and detection that arise due to seasonal effects that impact both vector dynamics [77, 259], and the level of symptoms displayed by infected hosts [267, 322, 323]. These heterogeneities motivate considering temporally varying sampling strategies that allow for surveillance resources to be optimally allocated throughout the year, which is the focus of the next chapter of this thesis. Further avenues for investigation include incorporating a latent period (time from when infection occurs to when the plant becomes infectious) in the model, or more generally a graduated progression through model compartments in which symptom expression and/or the likelihood of detection increase between successive compartments [211, 324, 325].

In summary, the results described in this chapter represent a key step towards understanding how sentinel plants may be used to facilitate the early detection of invasive plant pathogens. As we have shown, sentinels have the potential to reduce the expected prevalence of disease upon pathogen discovery substantially, thereby lowering the cost of reactive control and increasing the chance of pathogen containnment. These results encourage further research into the economic and logistical viability of using sentinel plants as a means of combating the problem of presymptomatic infection, as well as into the precise epidemiological characteristics of particular sentinel-crop-pathogen combinations.

Appendix C

Supplementary material for Chapter 6

C.1 Comparing random sampling and repeated sampling

When considering the use of sentinel plants for surveillance, as well as deciding how many sentinels to add to the population (P_S), an important question is how many of those sentinels to include in the sample (N_S). A natural choice is to sample the maximum possible number of sentinels – that is, choosing $N_S = \min(P_S, N)$. However, in some circumstances this would result in frequent resampling of the same plants. For example, if the total sample size, N, is equal to the total number of sentinels in the population, P_S, selecting N_S in this way would mean inspecting an identical set of plants on every sampling round. Since the information we gain about the disease status of a plant we have already inspected in the recent past is less than the information we gain by looking at a plant we have not previously inspected, repeated sampling of this kind may reduce the effectiveness of the surveillance strategy compared to a situation in which the sample selection has an element of randomness.

Here, we use a simple example to demonstrate how repeated sampling of the same plants can lead to a worse outcome than selecting plants to sample at random. We constructed a standard one-species Susceptible-Infected (SI) model in which plants progress from the 'Susceptible' class to the 'Infected' class at rate βSI (where β is the transmission coefficient, S is the number of 'Susceptible' plants, and I is the number of 'Infected' plants). Setting $\beta = 5 \times 10^{-6}$, we generated simulated epidemic curves in a population of size $1,000$, beginning from a single infected individual. We considered a monitoring strategy in which samples of size N were taken from the population every Δ days, with detection occurring if and only if an 'Infected' plant was included in the sample. We implemented this monitoring strategy on our simulated epidemic curves for a range of sample sizes N and sample intervals Δ. For each choice of N and Δ we considered both random sampling (in which the sampled plants were chosen at random on every sampling round), and repeated sampling (in which the same N plants were inspected on every sampling round).

For the chosen model parameterisation, random sampling outperformed repeated sampling across the range of (N, Δ) values considered (Supplementary Fig C.1). This result is conceptually equivalent to the phenomenon observed in the main text – that sampling as many sentinels as possible is not always the best strategy, if repeated sampling is required to achieve this. Note that the difference between the two sampling approaches tested here was smaller for large sample intervals Δ, since the longer the time between samples the less relevant the information gained from a previous sample is to the current state of the system. Therefore, less information is lost by looking at the same plants. The difference between the two approaches also decreased for larger sample sizes N, since detection occurred within fewer sampling rounds (thus less repetition occurred).

C.2 Varying the sensitivity of visual inspection

Throughout the analyses conducted in Chapter 6, we assumed that visual inspection was a perfectly sensitive means of detection for all Detectable plants. That is, we assumed that Detectable plants that were included in a sample were always correctly identified as being infected. In particular, we made no distinction between the detection sensitivity for crop and sentinel plants. However, perfect sensitivity is unlikely to be a realistic assumption, since early symptoms of *X. fastidiosa* infection may easily be overlooked or mistaken for other conditions such as water stress. Additionally, for a large host plant such as *O. europaea*, careful inspection of the entire host is not feasible. Therefore, if symptoms are not yet expressed throughout the plant, successful detection depends upon a symptomatic section of the plant being selected for inspection. For this reason, the sensitivity of visual inspection is also likely to differ between crop and sentinel plants. Our candidate sentinel plant *C. roseus* is much smaller in size than *O. europaea*, meaning that a thorough inspection of each sampled host is more feasible. This could result in a higher detection sensitivity for sentinel plants. Allowing for imperfect detection sensitivity for Detectable plants would increase the baseline EDP, which could affect the optimal number of sentinels to include in the population for a given sample size and sample interval. Additionally, if the sensitivity of detection indeed differed between crop and sentinel hosts, the optimal allocation of sampling resources between crops and sentinels would vary, as well as the overall potential for sentinel plants to facilitate early detection.

Here, we conducted a preliminary analysis to investigate the effects of varying the sensitivity of visual inspection for Detectable crop and sentinel plants. We introduced two additional parameters to represent the detection sensitivities for crop and sentinel plants, denoted θ_C and θ_S respectively, and adjusted the sampling algorithm applied in Chapter 6 so that a Detectable crop (respectively sentinel) plant included in a sample at any given time is correctly identified as infected with probability θ_C (respectively θ_S). We allowed θ_C and θ_S

to independently take each of the three values 0.25, 0.5, and 1. This resulted in a total of nine scenarios, that included the following cases:

- The detection sensitivity is perfect for both crop and sentinel plants ($\theta_C = \theta_S = 1$), as considered in the main text (included here for comparison with alternative values);
- The detection sensitivity is reduced equally for crop and sentinel plants ($\theta_C = \theta_S < 1$);
- The detection sensitivity is higher for sentinels than crops ($\theta_S > \theta_C$);
- The detection sensitivity is lower for sentinels than crops ($\theta_S < \theta_C$) (included for completeness, although we deem this scenario less likely to be realistic due to the much smaller size of sentinel plants compared to crop plants).

For each of the nine scenarios considered, we repeated the analyses described in Section 6.3.3 and produced plots analogous to those displayed in Figs 6.8, 6.10A, and 6.10B, showing the optimal total number of sentinels to include in the population (Supplementary Fig C.2), the corresponding reduction in the EDP compared to the baseline level (Supplementary Fig C.3), and the resultant EDP (Supplementary Fig C.4).

In every case that we considered for which the detection sensitivity for sentinel plants was greater than or equal to the detection sensitivity for crop plants ($\theta_S \geqslant \theta_C$; all panels with a white background in Supplementary Figs C.2, C.3, C.4), including sentinel plants in the population remained beneficial across the range of (N, Δ) values considered (Supplementary Figs C.2, C.3). The higher the sensitivity of detection for sentinel plants relative to crop plants, the greater the reductions in the EDP compared to the baseline level that were achieved (Supplementary Fig C.3). Since it is likely that visual inspection as a means of detection is indeed more sensitive for sentinel plants than for crop plants, these results provide further support for our previous conclusions regarding the potential of sentinel plants to facilitate substantial reductions in the EDP. However, as expected, sentinel plants were less beneficial when $\theta_S < \theta_C$ (panels with a grey background in Supplementary Figs C.2, C.3, C.4). Although we deem this scenario less likely to be realistic, this result highlights the importance of considering the effect of detection sensitivities in future studies that investigate the use of sentinel plants for surveillance.

C.3 Variation in parameter values

Where possible, the epidemiological parameters used in our main analyses were chosen based on literature estimates for *X. fastidiosa* infection in *O. europaea* and *C. roseus* (Table 6.2 of the main text). However, due to uncertainty in these reported values and the lack of available estimates for some parameters, we also conducted supplementary analyses to investigate the robustness of our results to parameter variation. We performed these analyses

both for the baseline case, and with sentinels included. The parameters that we varied, their meanings, their values taken in the main text, and the alternative values considered in our supplementary analyses are given in Supplementary Table C.1 (baseline case) and Supplementary Table C.2 (with sentinels included). In almost all cases considered, the selected parameter was varied whilst all other parameters remained fixed at their main text values. The exception to this was when varying the crop population size (P_C); in those cases, we simultaneously scaled the transmission coefficients for Detectable crops and sentinels[1] (β_C and β_S respectively) to ensure that the epidemic growth rate remained fixed at $\beta_C P_C = 0.05$, whilst keeping $\beta_C = \beta_S$ [309]. In the baseline case, for each alternative parameter value considered we present plots analogous to Fig 6.4C in the main text, showing the baseline EDP as a percentage of the total crop population size (Supplementary Figs C.5, C.6). When sentinels were included, for each alternative parameter value considered we present plots analogous to Figs 6.8, 6.10A, and 6.10B in the main text, showing the optimal total number of sentinels in the population, the percentage change in EDP compared to the baseline level, and the resultant EDP (Supplementary Figs C.7-C.13).

Parameter	**Meaning**	**Main text value**	**Alternative values considered**
β_C	Transmission coefficient for Detectable crops	5×10^{-5}	2.5×10^{-5}, 1×10^{-4}
ε_C	Transmission scaling factor for Undetectable crops	0.015	0.1, 0.25
γ_C	Mean duration of crop Undetectable period	452 days	350 days, 550 days
P_C	Number of crop plants in the population	$1,000$	500, $1,500$*
U_0	Initial number of Undetectable infected individuals	1	2, 4

* value of β_C scaled accordingly to ensure $\beta_C P_C = 0.05$.

Table C.1: Parameter variation in the baseline case: the parameters that we varied, their meanings, their values used in the main text, and the alternative values that we considered in our Supplementary analyses.

[1] This is because we wish to consider a population that is larger (or smaller) both in terms of size and in terms of area, but with the same overall population density. A larger population spread over a larger area with the same density should not result in an inherently greater transmission rate, as would be the case if a larger population were concentrated into the same total area, increasing the density. Since the density remains constant, in this specific instance we scale β_C and β_S by the population size (in contrast to when the population size and density change due to the introduction of sentinels, when we do not).

Parameter	Meaning	Main text value	Alternative values considered
β_S	Transmission coefficient for Detectable sentinels	5×10^{-5}	2.5×10^{-5}, 1×10^{-4}
ε_C	Transmission scaling factor for Undetectable crops	0.015	0.1, 0.25
ε_S	Transmission scaling factor for Undetectable sentinels	0.1	0.02, 0.5
γ_C	Mean duration of crop Undetectable period	452 days	350 days, 550 days
γ_S	Mean duration of sentinel Undetectable period	49 days	28 days, 70 days
P_C	Number of crop plants in the population	$1,000$	500, $1,500$*
U_0	Initial number of Undetectable infected individuals	1	2, 4

* values of β_C, β_S scaled accordingly to ensure that $\beta_C P_C = 0.05$ and $\beta_C = \beta_S$.

Table C.2: Parameter variation with sentinels included: the parameters that we varied, their meanings, their values used in the main text, and the alternative values that we considered in our Supplementary analyses.

C.4 Variation in the objective function

Throughout our main analyses, we investigated how including sentinel plants in a surveillance strategy could reduce the EDP compared to the baseline level, where the EDP was defined to be the expected detection prevalence in the crop population at the time of discovery. In doing so, we assumed that minimising the prevalence of infection amongst valuable crop plants was likely to be the primary objective of the grower, and did not consider the prevalence of infection in the sentinel population. However, in practice a grower may also wish to take the expected detection prevalence in the sentinel population (EDP_{sent}) into account in some way when evaluating the effects of surveillance strategies. Reasons for this include reducing the cost of removing infected plants post-detection, and reducing the likelihood that the pathogen is exported to a new location before control can be implemented. Additionally, in some pathosystems the sentinel plants may have some intrinsic value – for example, young orange trees used as sentinels for greening disease in citrus groves.

Our analyses can be adapted to explore any balance of interest between crop and sentinel detection prevalence by replacing the EDP with $\Omega = \text{EDP} + (\omega_S \times \text{EDP}_{\text{sent}})$ for some weight $\omega_S \in [0, 1]$. Computationally speaking, this simply requires a small adjustment to the objective function used in our Bayesian optimisation algorithm. Our main analyses

correspond to the case $\omega_S = 0$ (Figs 6.8, 6.9, 6.10); here, we display analogous results for the case $\omega_S = 0.5$ (Supplementary Fig C.15). This represents a hypothetical situation in which an infected sentinel plant is considered half as concerning as an infected crop plant (rather than not being considered at all). In this case, introducing a large number of sentinel plants is penalised by contributing to an increased EDP_{sent}. The optimal number of sentinels to include in the population is therefore reduced compared to the case in which $\omega_S = 0$, apart from when the sample interval Δ is large and the optimal number of sentinels is equal to the sample size (i.e. not in excess) (Supplementary Fig C.15A). The achievable percentage reduction in Ω compared to the baseline level is also reduced compared to the case in which we consider solely the EDP, since infected sentinel plants are no longer discounted (Supplementary Fig C.15B). This confirms that the specific objective a grower wishes to achieve should be carefully considered before the results of sentinel-based surveillance strategies are evaluated.

C.5 Supplementary Figures

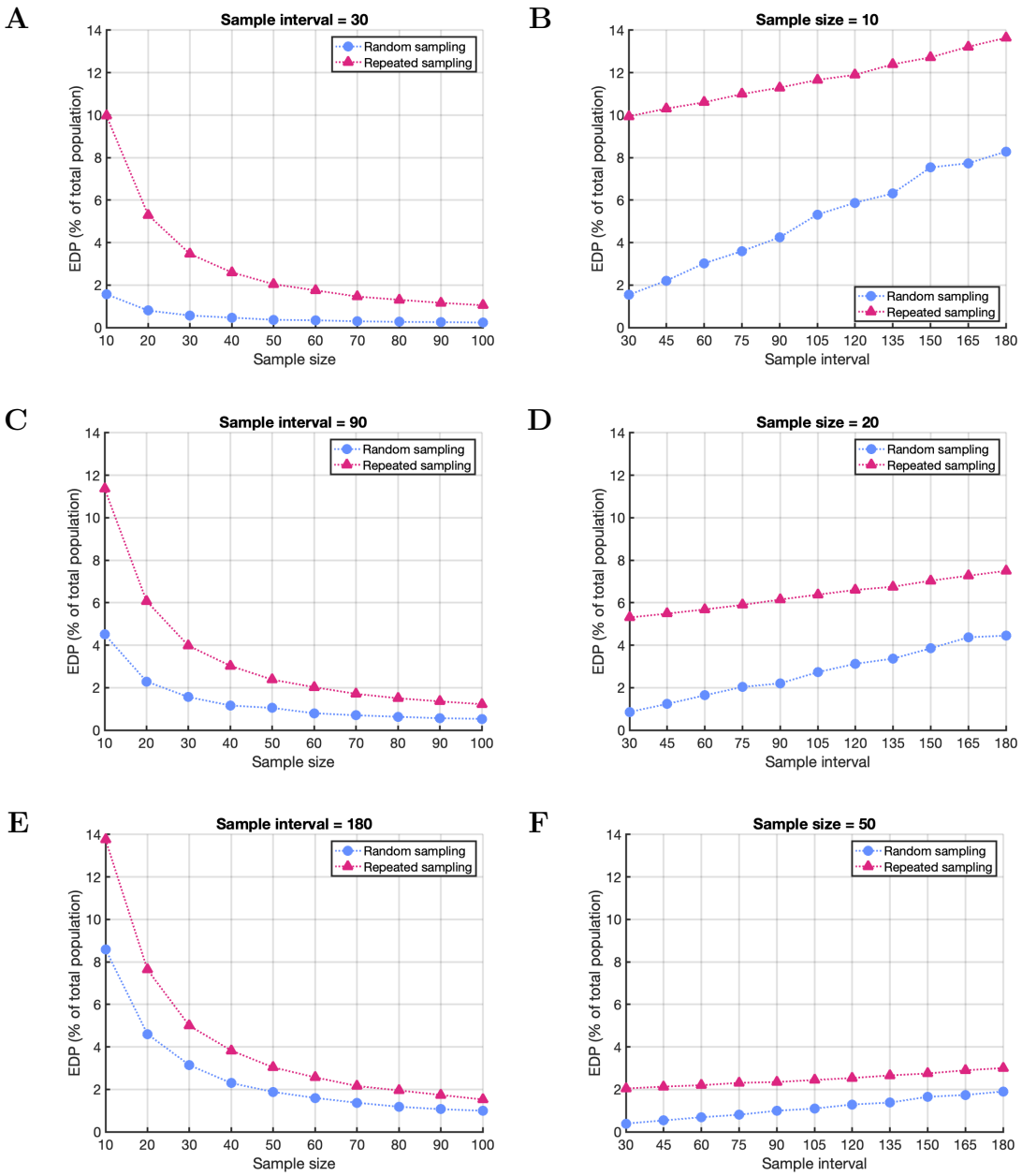


Figure C.1: Resultant EDPs (as a percentage of the total population) for the example system described in Supplementary Text C.1, illustrating how random sampling (blue circles) outperforms repeated sampling (pink triangles). A,C,E. The resultant EDP for sample intervals $\Delta = 30$ days (**A**), 90 days (**C**), and 180 days (**E**), as the sample size N varies. **B,D,F.** The resultant EDP for sample sizes $N = 10$ (**B**), $N = 20$ (**D**), and $N = 50$ (**F**), as the sample interval varies.

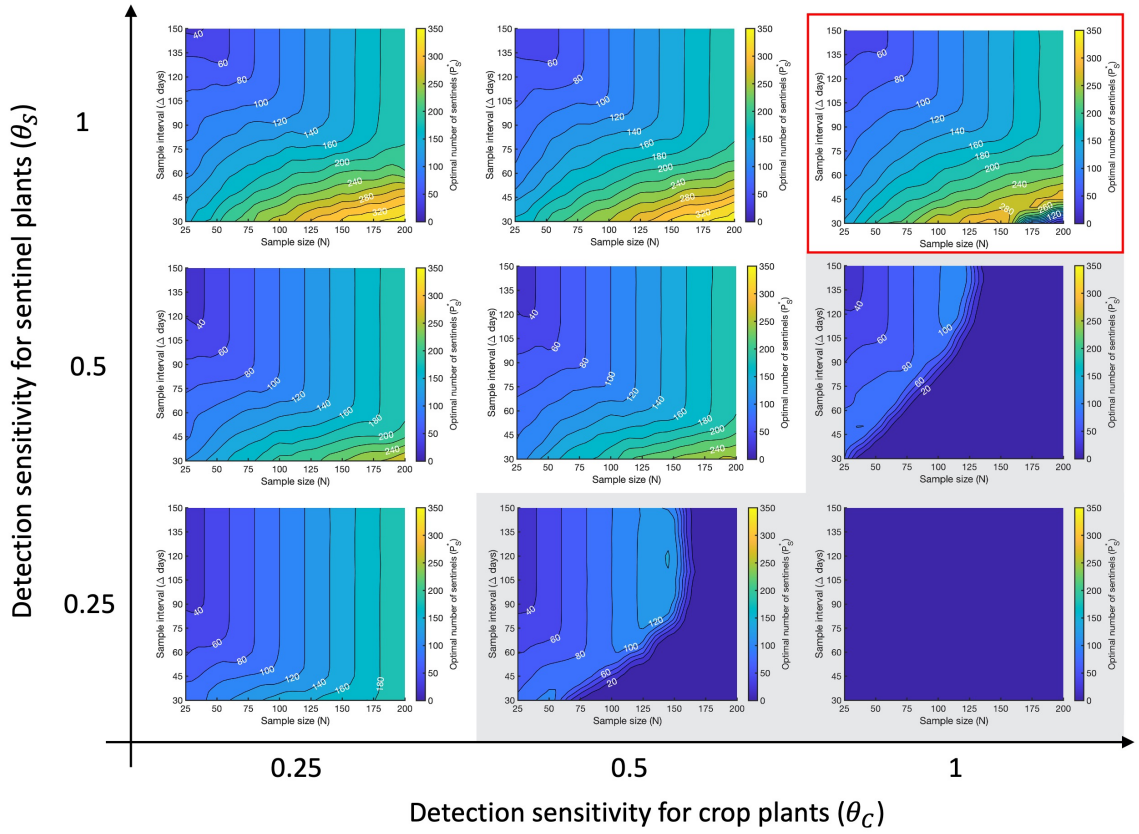


Figure C.2: The optimal number of sentinel plants to include in the population (P_S^*) as the sensitivity of visual inspection for Detectable crop plants (θ_C) and for Detectable sentinel plants (θ_S) vary. The results shown here are analogous to those shown in Fig 6.8, but allowing for variation in the detection sensitivity for crop and sentinel plants as described in Supplementary Text C.2. Columns left to right represent a detection sensitivity for crop plants of $\theta_C = 0.25, 0.5$, and 1, respectively. Rows bottom to top represent a detection sensitivity for sentinel plants of $\theta_S = 0.25, 0.5$, and 1, respectively. All other model parameters remain at their values given in Table 6.2. Red box indicates the scenario $\theta_C = \theta_S = 1$ (perfect detection sensitivity for both crop and sentinel plants) that was considered in the main text, included here for ease of comparison with alternative values of θ_C and θ_S. Panels with a grey background are those for which $\theta_S < \theta_C$ (i.e. detection sensitivity is less for sentinels than for crops), which we consider to be less likely to reflect reality.

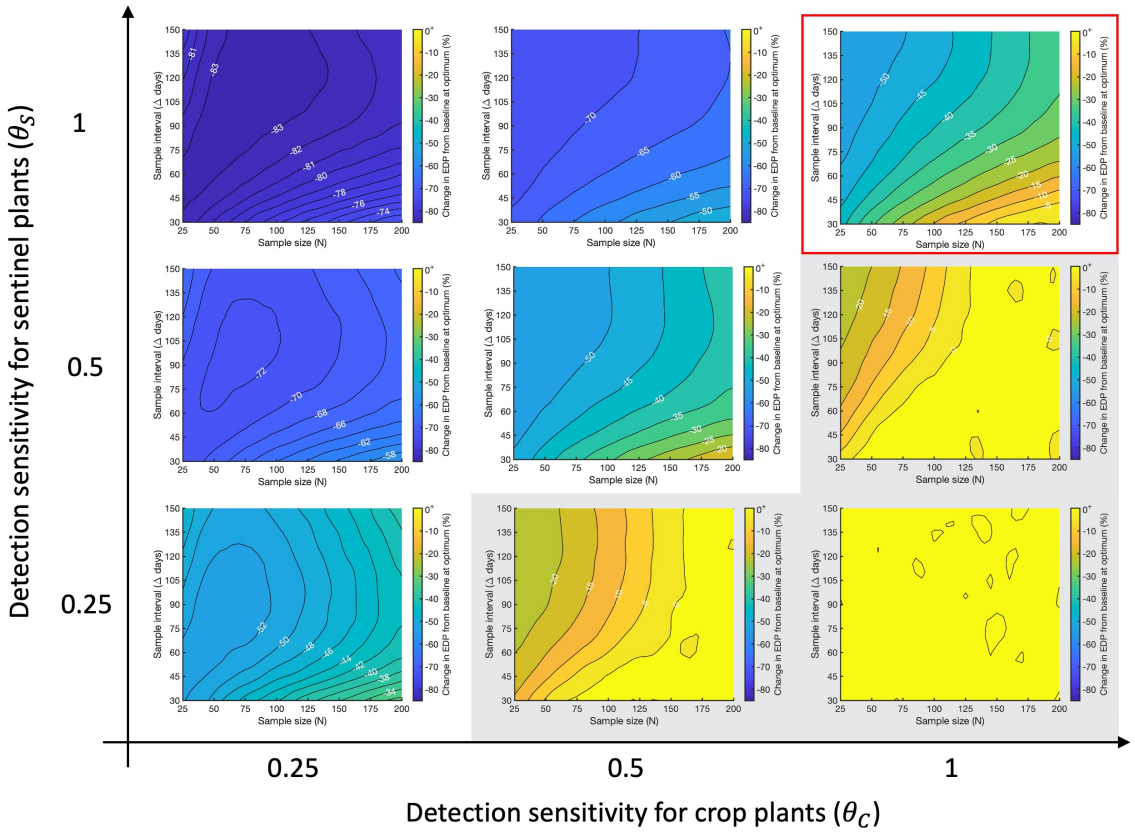


Figure C.3: The optimal percentage change in the EDP compared to the baseline value as the sensitivity of visual inspection for Detectable crop plants (θ_C) and for Detectable sentinel plants (θ_S) vary. The results shown here are analogous to those shown in Fig 6.10A, but allowing for variation in the detection sensitivity for crop and sentinel plants as described in Supplementary Text C.2. Columns left to right represent a detection sensitivity for crop plants of $\theta_C = 0.25, 0.5$, and 1, respectively. Rows bottom to top represent a detection sensitivity for sentinel plants of $\theta_S = 0.25, 0.5$, and 1, respectively. All other model parameters remain at their values given in Table 6.2. Red box indicates the scenario $\theta_C = \theta_S = 1$ (perfect detection sensitivity for both crop and sentinel plants) that was considered in the main text, included here for ease of comparison with alternative values of θ_C and θ_S. Panels with a grey background are those for which $\theta_S < \theta_C$ (i.e. detection sensitivity is less for sentinels than for crops), which we consider to be less likely to reflect reality.

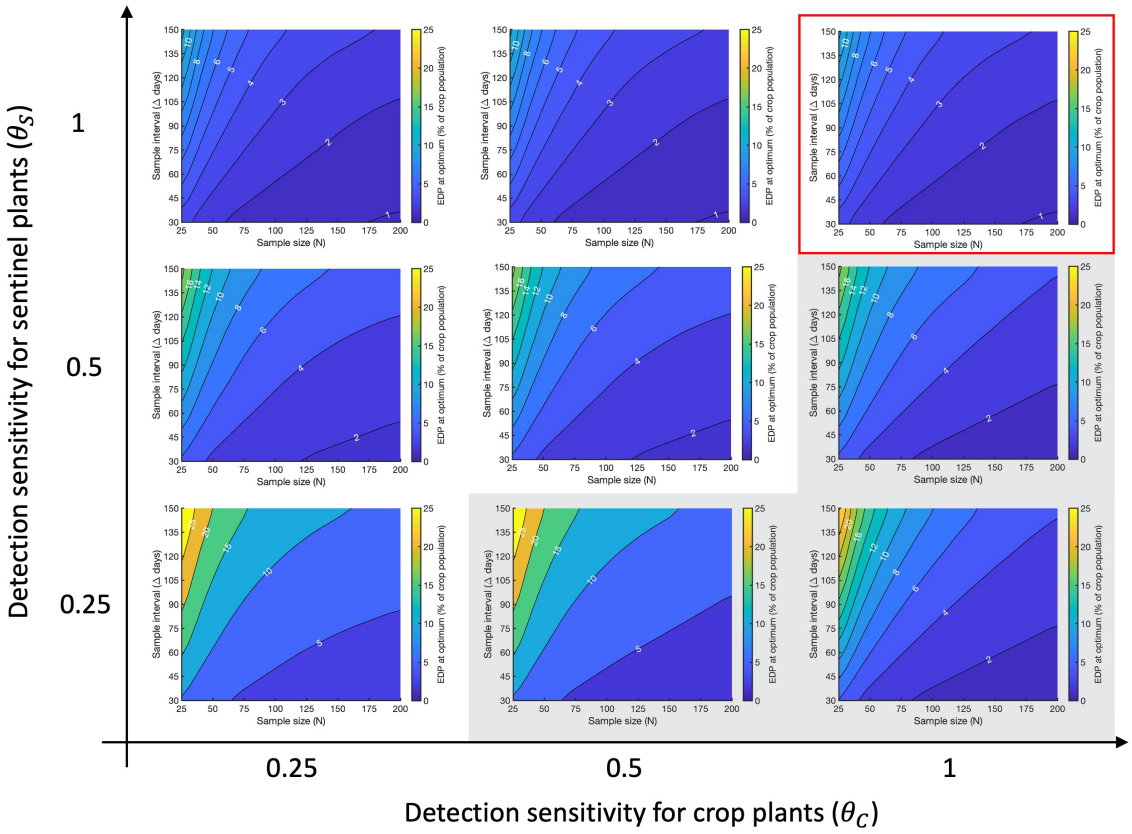


Figure C.4: The resultant value of the EDP at the optimum, expressed as a percentage of the total crop population, as the sensitivity of visual inspection for Detectable crop plants (θ_C) and for Detectable sentinel plants (θ_S) vary. The results shown here are analogous to those shown in Fig 6.10B, but allowing for variation in the detection sensitivity for crop and sentinel plants as described in Supplementary Text C.2. Columns left to right represent a detection sensitivity for crop plants of $\theta_C = 0.25, 0.5$, and 1, respectively. Rows bottom to top represent a detection sensitivity for sentinel plants of $\theta_S = 0.25, 0.5$, and 1, respectively. All other model parameters remain at their values given in Table 6.2. Red box indicates the scenario $\theta_C = \theta_S = 1$ (perfect detection sensitivity for both crop and sentinel plants) that was considered in the main text, included here for ease of comparison with alternative values of θ_C and θ_S. Panels with a grey background are those for which $\theta_S < \theta_C$ (i.e. detection sensitivity is less for sentinels than for crops), which we consider to be less likely to reflect reality.

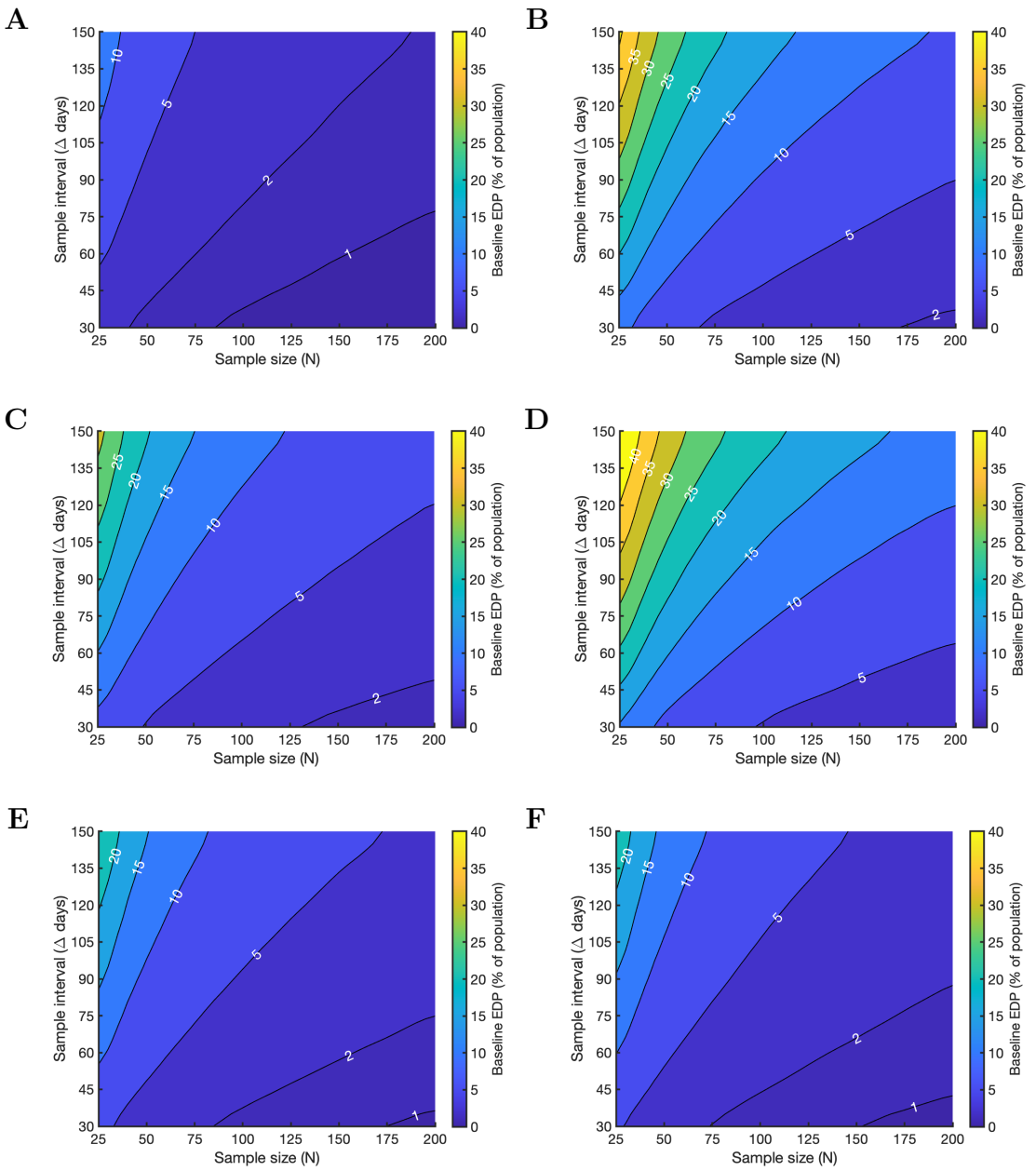


Figure C.5: The effect of varying the model parameters on the baseline EDP. Panels analogous to Fig 6.4C in the main text. A,B. The effect of varying the transmission coefficient for Detectable crops (β_C) from its baseline value ($\beta_C = 5 \times 10^{-5}$) to $\beta_C = 2.5 \times 10^{-5}$ (**A**) and to $\beta_C = 1 \times 10^{-4}$ (**B**). **C,D.** The effect of varying the transmission scaling factor for Undetectable crops (ε_C) from its baseline value ($\varepsilon_C = 0.015$) to $\varepsilon_C = 0.1$ (**C**) and to $\varepsilon_C = 0.25$ (**D**). **E,F.** The effect of varying the mean duration of the crop Undetectable period (γ_C) from its baseline value ($\gamma_C = 452$ days) to $\gamma_C = 350$ days (**E**) and to $\gamma_C = 550$ days (**F**).

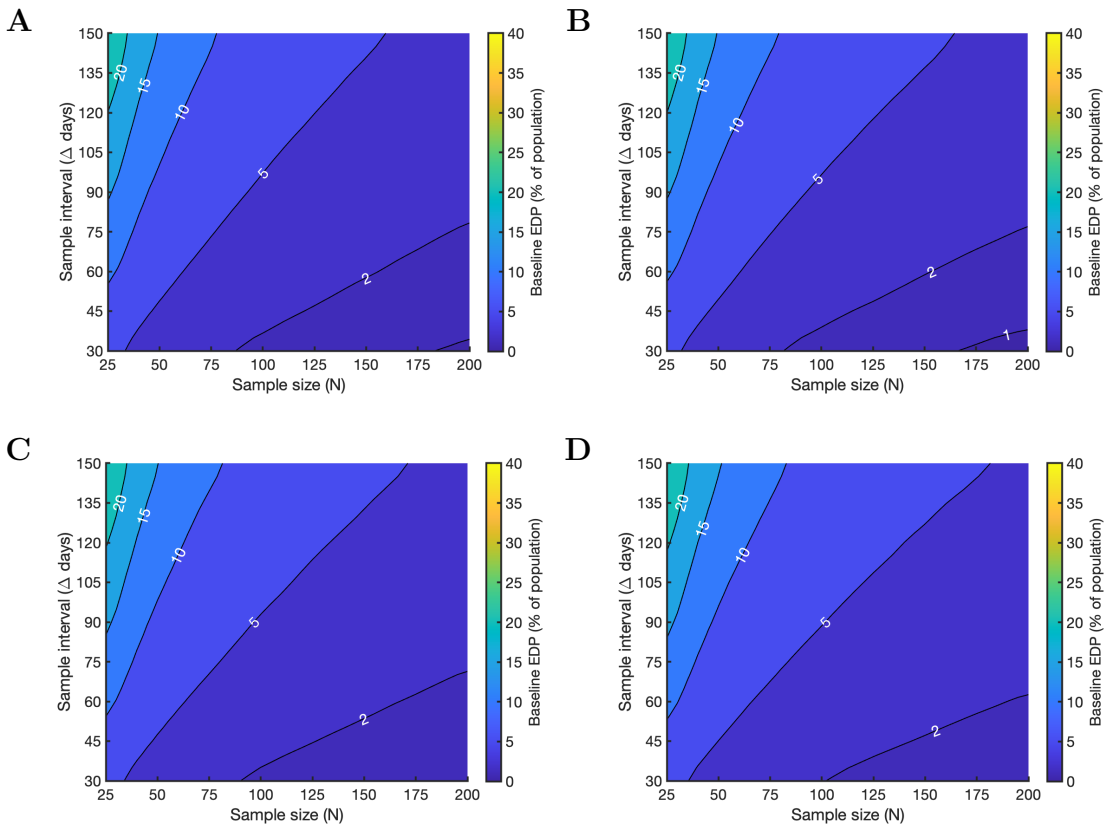


Figure C.6: The effect of varying the model parameters on the baseline EDP. Panels analogous to Fig 6.4C in the main text. A,B. The effect of varying the crop population size (P_C) from its baseline value ($P_C = 1,000$) to $P_C = 500$ (**A**) and to $P_C = 1,500$ (**B**). **C,D.** The effect of varying the initial number of Undetectable infected hosts (U_0) from its baseline value ($U_0 = 1$) to $U_0 = 2$ (**C**) and to $U_0 = 4$ (**D**).

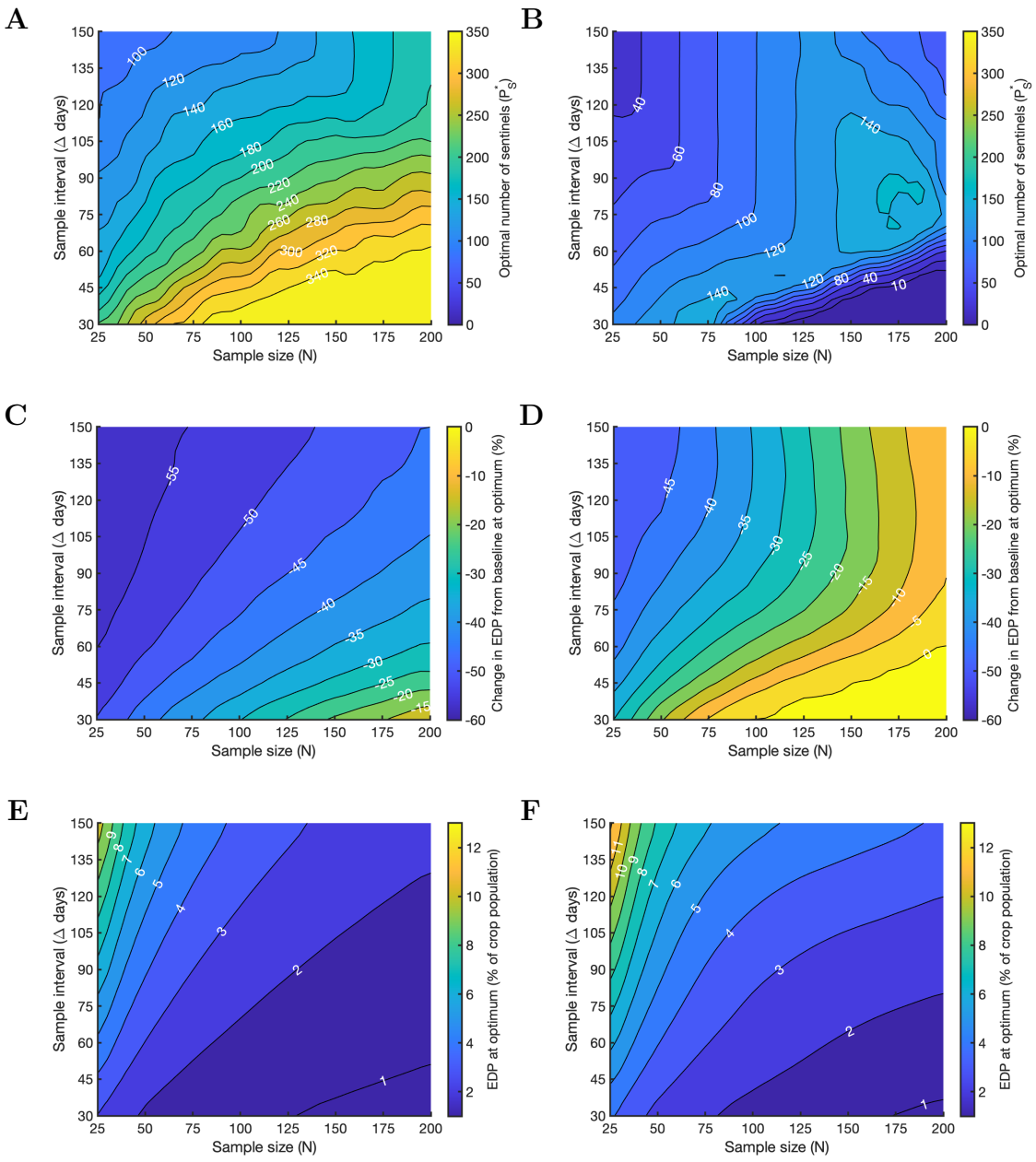


Figure C.7: The effect of varying the transmission coefficient for Detectable sentinels from $\beta_S = 5 \times 10^{-5}$ (baseline value) to $\beta_S = 2.5 \times 10^{-5}$ (A,C,E) and to $\beta_S = 1 \times 10^{-4}$ (B,D,F). Panels analogous to Figs 6.8, 6.10A, and 6.10B in the main text. A,B. The optimal number P_S^* of sentinel plants to include in the population, for which the maximal reduction in the EDP compared to the baseline level is achieved (if N_S is also chosen optimally). **C,D.** The percentage change in the EDP compared to the baseline value at the optimum, achieved when $P_S = P_S^*$ and $N_S = N_S^*$. **E,F.** The resultant value of the EDP at the optimum, expressed as a percentage of the total crop population.

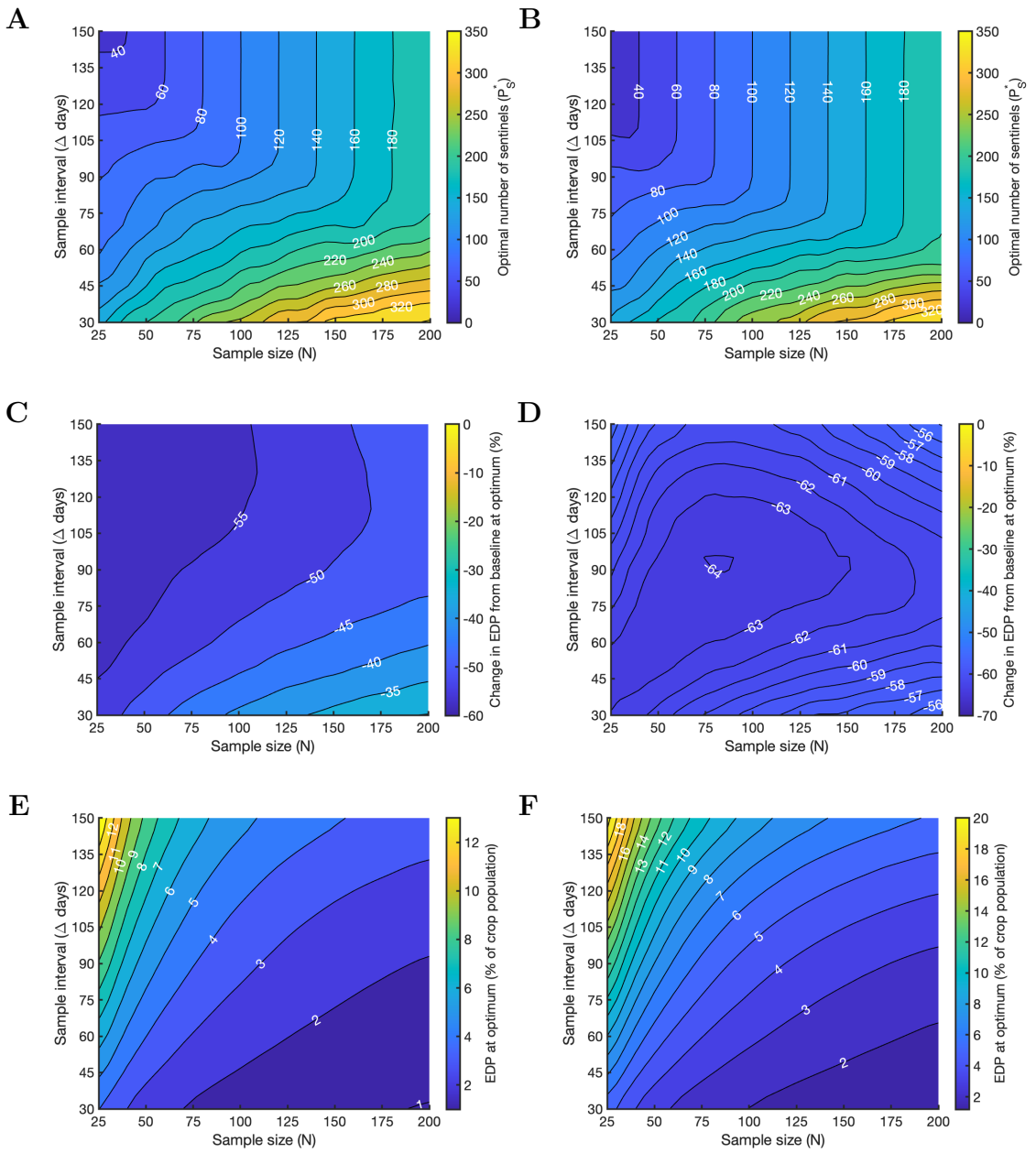


Figure C.8: The effect of varying the transmission scaling factor for Undetectable crops from $\varepsilon_C = 0.015$ (baseline value) to $\varepsilon_C = 0.1$ (A,C,E) and to $\varepsilon_C = 0.25$ (B,D,F). Panels analogous to Figs 6.8, 6.10A, and 6.10B in the main text. A,B. The optimal number P_S^* of sentinel plants to include in the population, for which the maximal reduction in the EDP compared to the baseline level is achieved (if N_S is also chosen optimally). **C,D.** The percentage change in the EDP compared to the baseline value at the optimum, achieved when $P_S = P_S^*$ and $N_S = N_S^*$. **E,F.** The resultant value of the EDP at the optimum, expressed as a percentage of the total crop population.

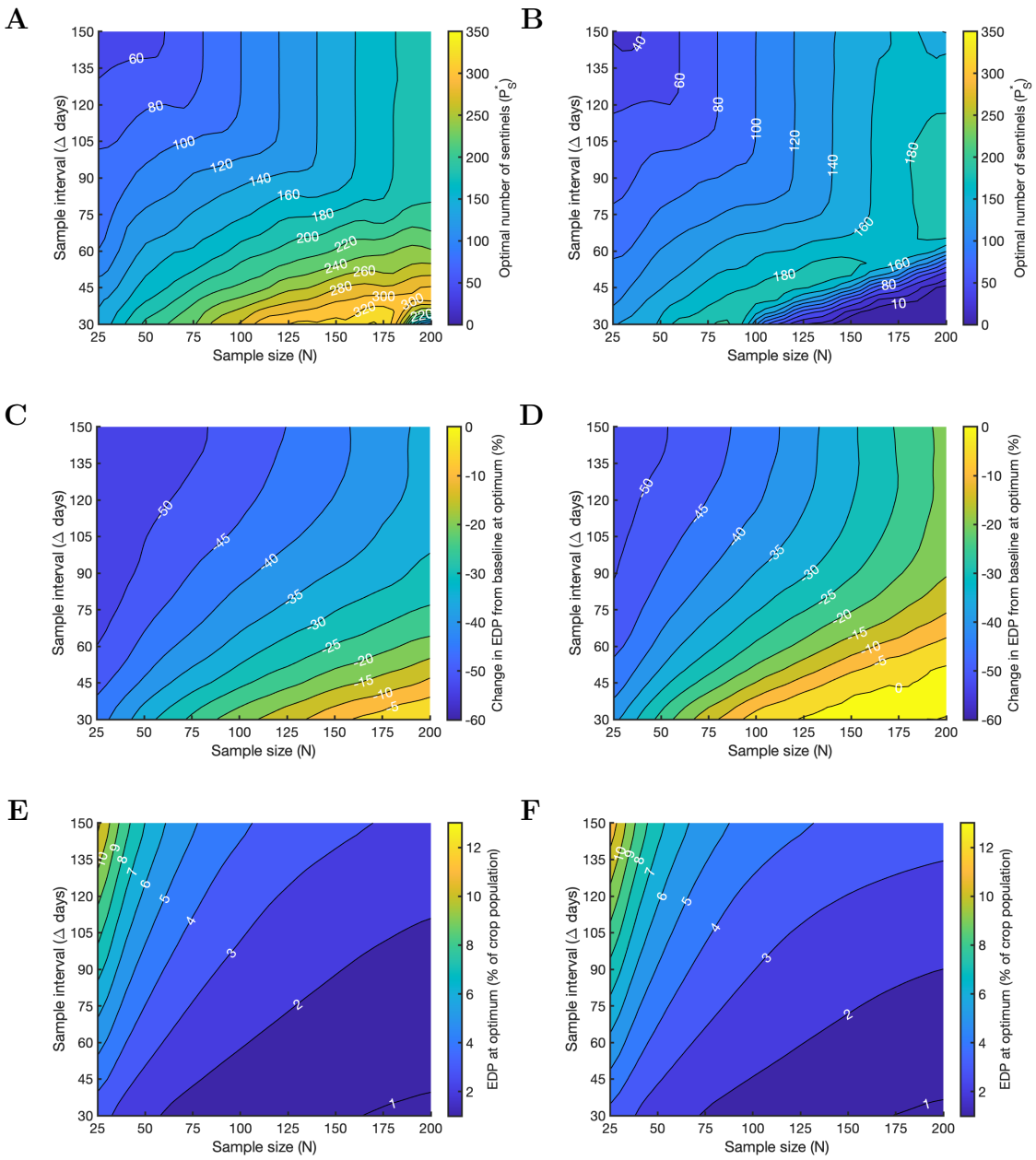


Figure C.9: The effect of varying the transmission scaling factor for Undetectable sentinels from $\varepsilon_S = 0.1$ (baseline value) to $\varepsilon_S = 0.02$ (A,C,E) and to $\varepsilon_S = 0.5$ (B,D,F). Panels analogous to Figs 6.8, 6.10A, and 6.10B in the main text. A,B. The optimal number P_S^* of sentinel plants to include in the population, for which the maximal reduction in the EDP compared to the baseline level is achieved (if N_S is also chosen optimally). **C,D.** The percentage change in the EDP compared to the baseline value at the optimum, achieved when $P_S = P_S^*$ and $N_S = N_S^*$. **E,F.** The resultant value of the EDP at the optimum, expressed as a percentage of the total crop population.

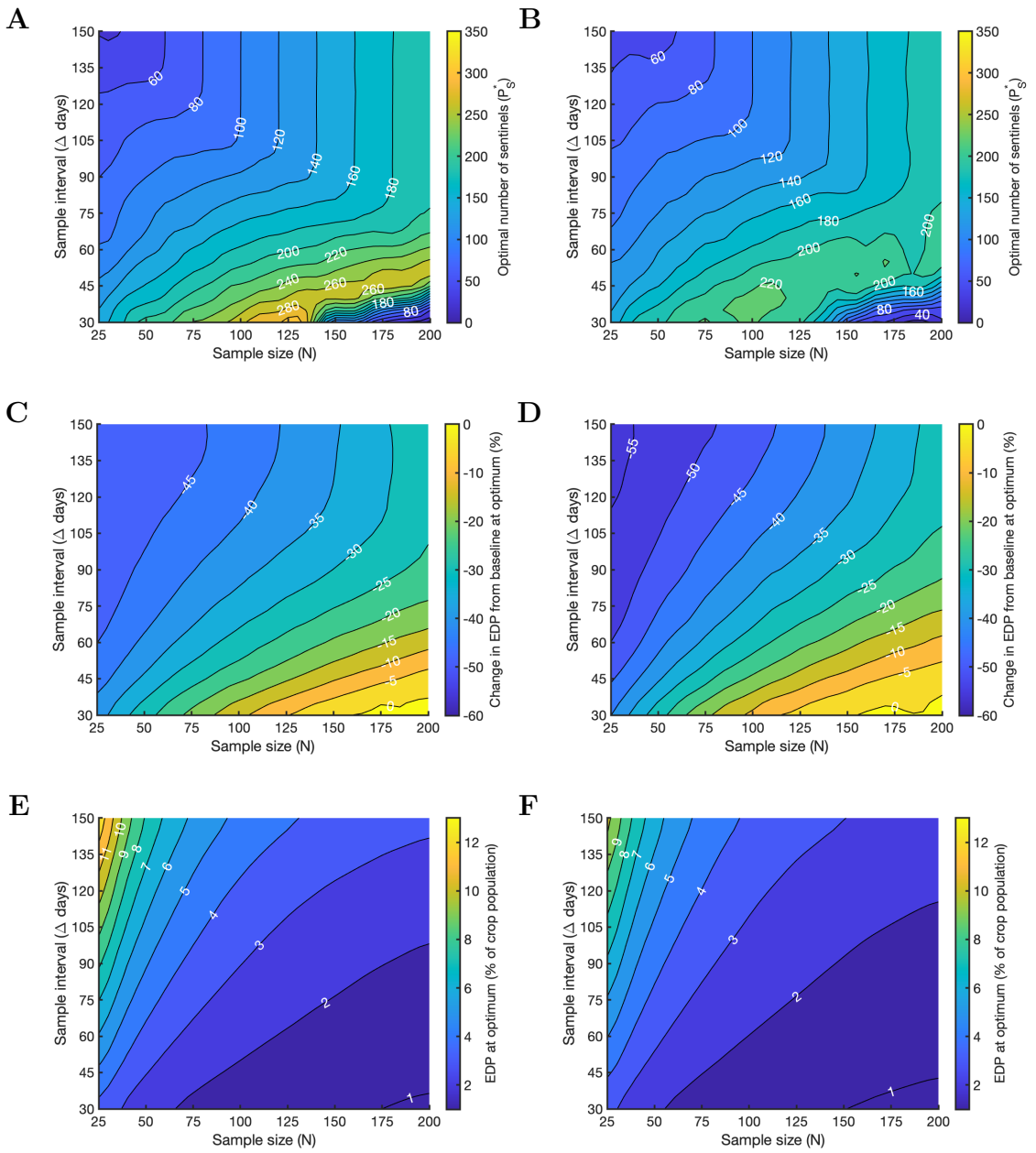


Figure C.10: The effect of varying the mean duration of the crop Undetectable period from $\gamma_C = 452$ days (baseline value) to $\gamma_C = 350$ days (A,C,E) and to $\gamma_C = 550$ days (B,D,F). Panels analogous to Figs 6.8, 6.10A, and 6.10B in the main text. A,B. The optimal number P_S^* of sentinel plants to include in the population, for which the maximal reduction in the EDP compared to the baseline level is achieved (if N_S is also chosen optimally). **C,D.** The percentage change in the EDP compared to the baseline value at the optimum, achieved when $P_S = P_S^*$ and $N_S = N_S^*$. **E,F.** The resultant value of the EDP at the optimum, expressed as a percentage of the total crop population.

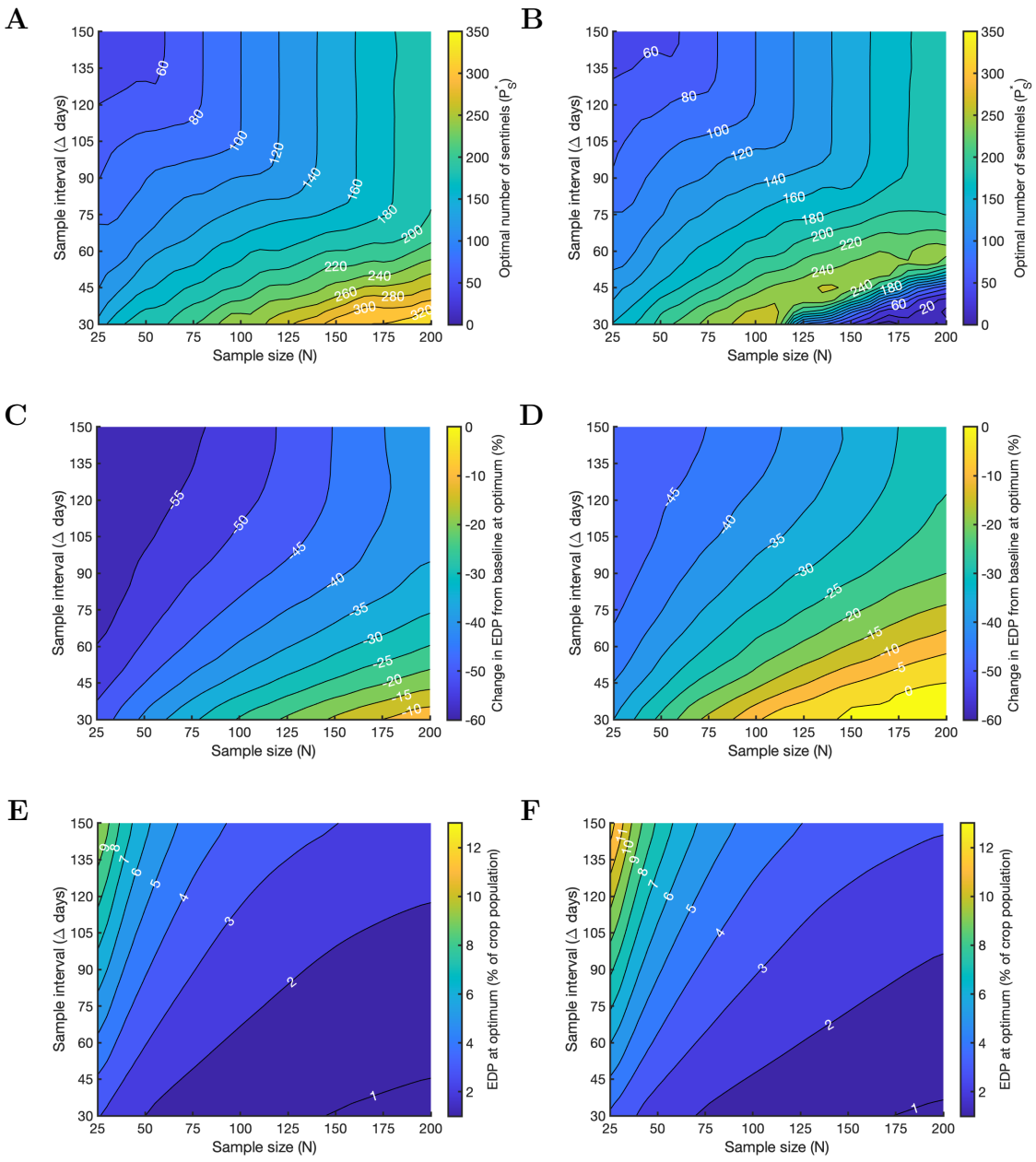


Figure C.11: The effect of varying the mean duration of the sentinel Undetectable period from $\gamma_S = 49$ days (baseline value) to $\gamma_S = 28$ days (A,C,E) and to $\gamma_S = 70$ days (B,D,F). Panels analogous to Figs 6.8, 6.10A, and 6.10B in the main text. A,B. The optimal number P_S^* of sentinel plants to include in the population, for which the maximal reduction in the EDP compared to the baseline level is achieved (if N_S is also chosen optimally). **C,D.** The percentage change in the EDP compared to the baseline value at the optimum, achieved when $P_S = P_S^*$ and $N_S = N_S^*$. **E,F.** The resultant value of the EDP at the optimum, expressed as a percentage of the total crop population.

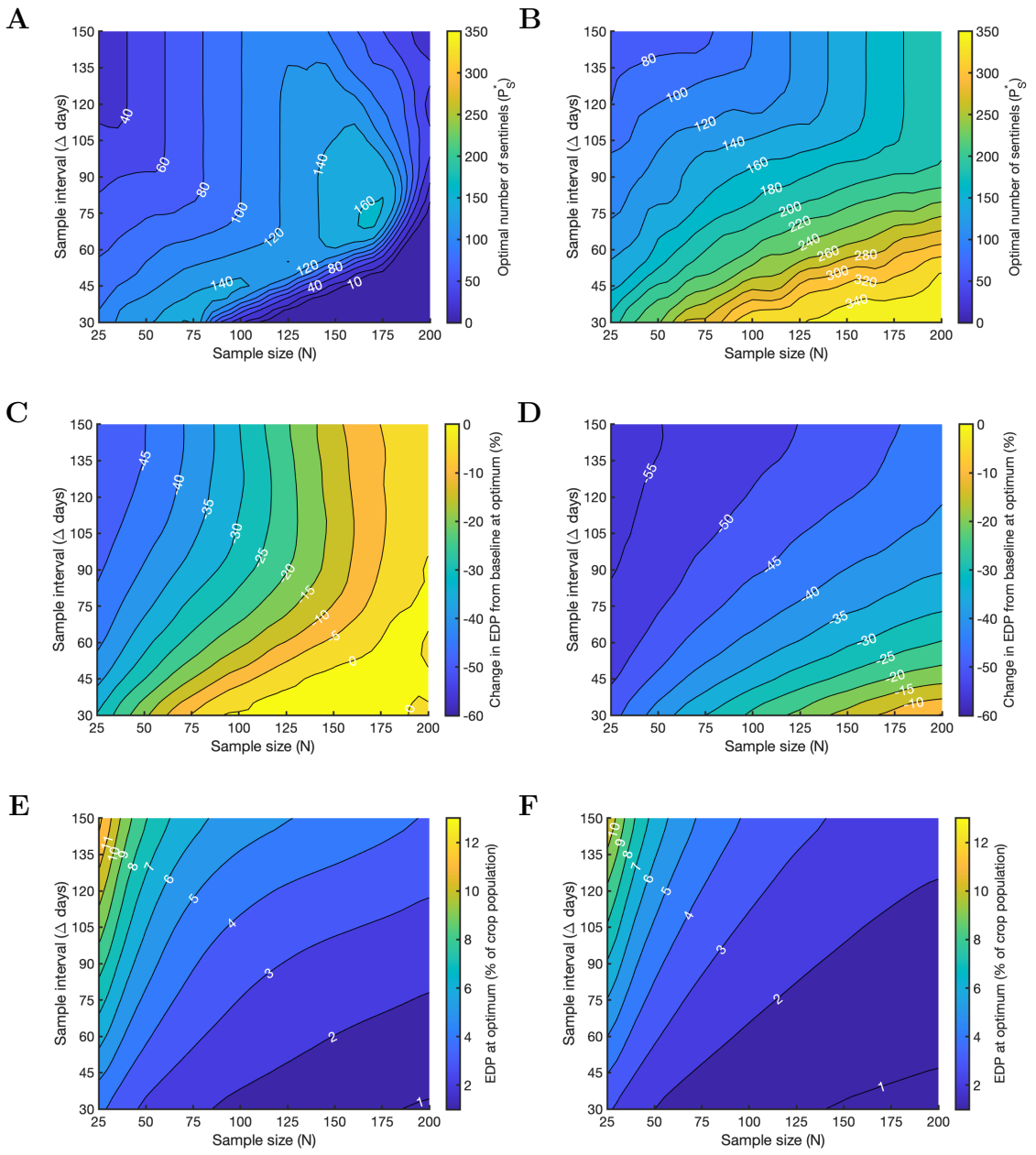


Figure C.12: The effect of varying the crop population size from $P_C = 1,000$ plants (baseline value) to $P_C = 500$ plants (A,C,E) and to $P_C = 1,500$ plants (B,D,F). Panels analogous to Figs 6.8, 6.10A, and 6.10B in the main text. A,B. The optimal number P_S^* of sentinel plants to include in the population, for which the maximal reduction in the EDP compared to the baseline level is achieved (if N_S is also chosen optimally). **C,D.** The percentage change in the EDP compared to the baseline value at the optimum, achieved when $P_S = P_S^*$ and $N_S = N_S^*$. **E,F.** The resultant value of the EDP at the optimum, expressed as a percentage of the total crop population.

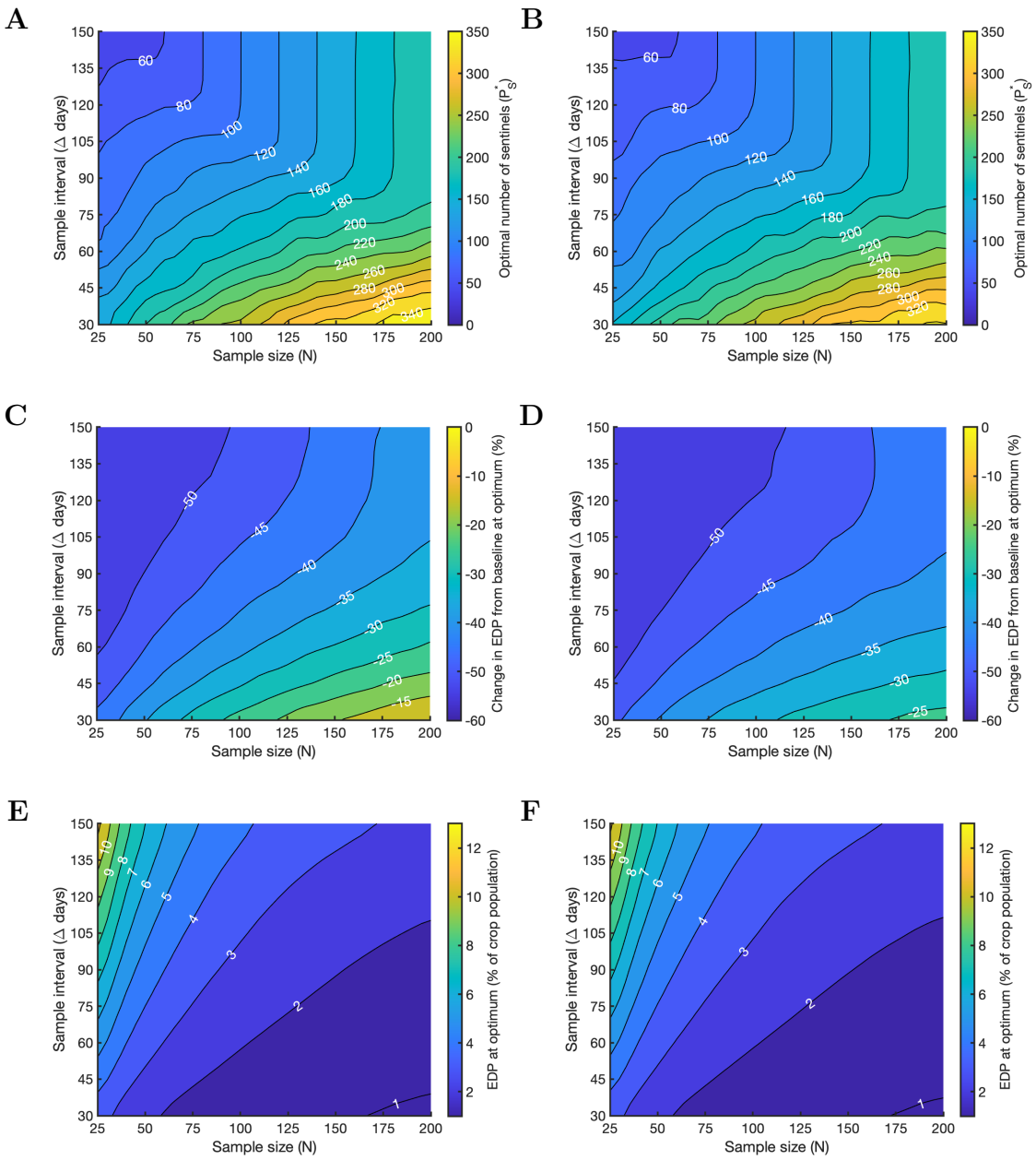


Figure C.13: The effect of varying the initial number of infected plants from $U_0 = 1$ (baseline value) to $U_0 = 2$ (A,C,E) and to $U_0 = 4$ (B,D,F). Panels analogous to Figs 6.8, 6.10A, and 6.10B in the main text. A,B. The optimal number P_S^* of sentinel plants to include in the population, for which the maximal reduction in the EDP compared to the baseline level is achieved (if N_S is also chosen optimally). **C,D.** The percentage change in the EDP compared to the baseline value at the optimum, achieved when $P_S = P_S^*$ and $N_S = N_S^*$. **E,F.** The resultant value of the EDP at the optimum, expressed as a percentage of the total crop population.

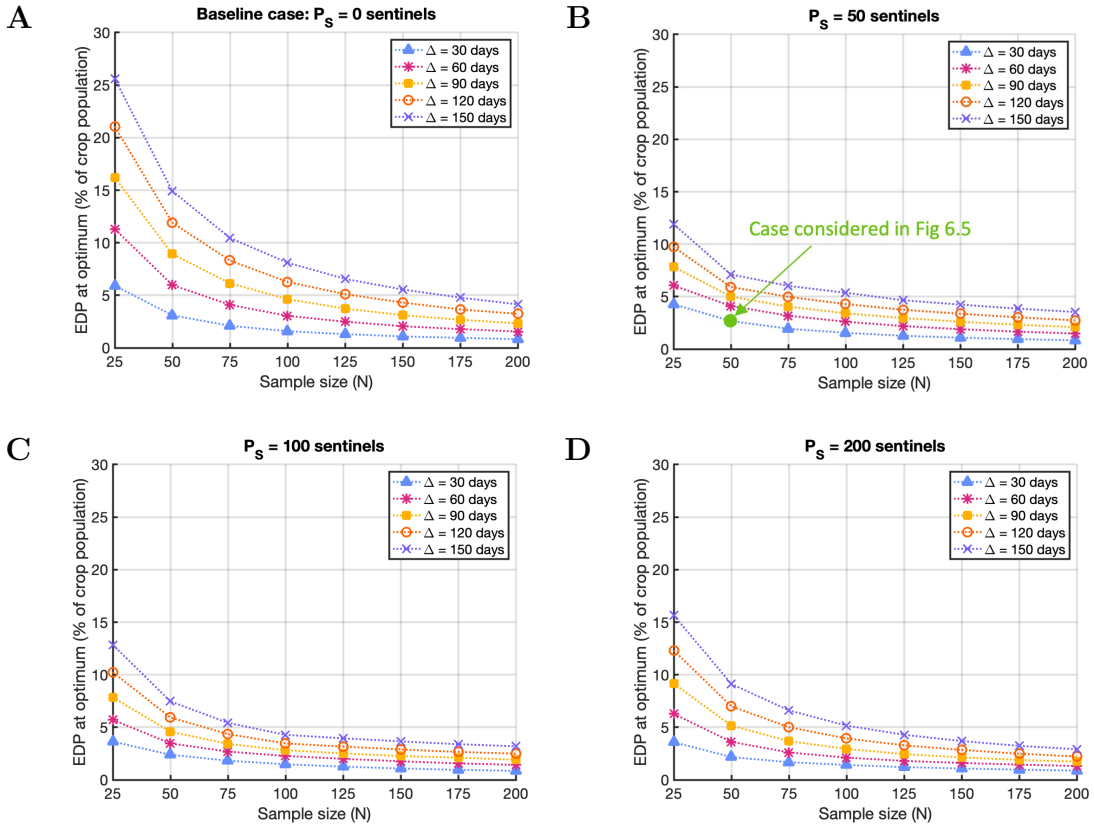


Figure C.14: Resultant EDPs in the baseline case and for the optimal strategies shown in Fig 6.6 of the main text. A. The EDP in the baseline case ($P_S = 0$), where the sample size (N) and sample interval (Δ) vary as shown. Since the baseline EDP depends on N and Δ, relative changes in the EDP compared to this baseline for different values of N and Δ (Figs 6.6B, 6.6D, 6.6F of the main text) are not a measure of the resultant EDP. **B.** The best achievable resultant EDP when $P_S = 50$, corresponding to the optimal strategies identified in Fig 6.6A in the main text. Green circle marks the case considered in Fig 6.5 in the main text ($P_S = 50, N = 50, \Delta = 30$ days). **C.** The analogous figure to B, but with $P_S = 100$ sentinels added to the population and results corresponding to the strategies identified in Fig 6.6C. **D.** The analogous figure to B, but with $P_S = 200$ sentinels added to the population and results corresponding to the strategies identified in Fig 6.6E.

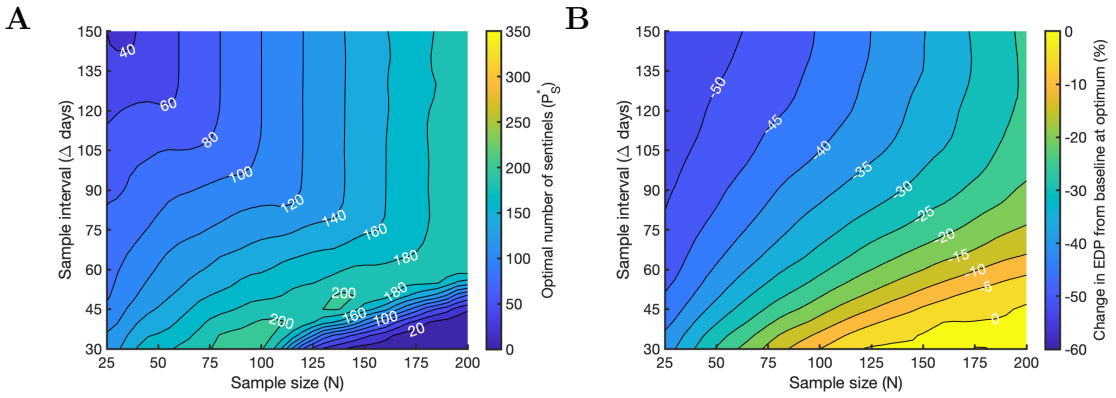


Figure C.15: Optimising the number of sentinels to include in the population and in the sample when $\Omega = \mathbf{EDP} + (0.5 \times \mathbf{EDP_{sent}})$. A. The optimal number P_S^* of sentinel plants to include in the population, for which the maximal reduction in Ω compared to the baseline level is achieved (if N_S is also chosen optimally). **B.** The percentage change in Ω compared to the baseline value at the optimum, achieved when $P_S = P_S^*$ and $N_S = N_S^*$.

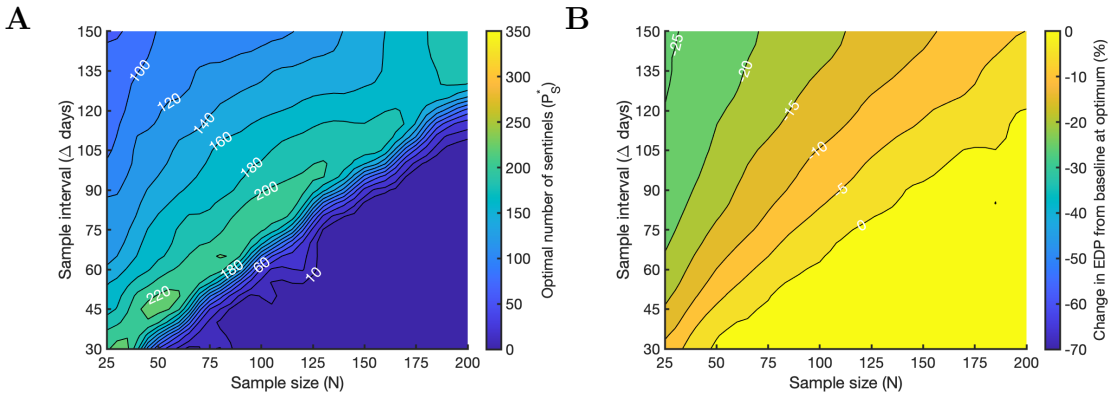


Figure C.16: Optimising the number of sentinels to include in the population and in the sample when the susceptibility of sentinels to infection is halved. A. The optimal number P_S^* of sentinel plants to include in the population, for which the maximal reduction in the EDP compared to the baseline level is achieved (if N_S is also chosen optimally). **B.** The percentage change in the EDP compared to the baseline value at the optimum, achieved when $P_S = P_S^*$ and $N_S = N_S^*$.

Chapter 7

Seasonal surveillance for *Xylella fastidiosa*: Optimising resource allocation for early detection

7.1 Introduction

In the previous chapter of this thesis, we explored the potential for sentinel plants to facilitate early detection of invading plant pathogens, using *Xylella fastidiosa* infection in *Olea europaea* (European olive) as a case study of significant current importance. With *Catharanthus roseus* (Madagascan periwinkle) as a candidate sentinel species, we demonstrated that sentinel plants can be used to achieve substantial reductions in the expected prevalence of infection in the population at the time of detection (expected detection prevalence, or EDP) compared to a non-sentinel based surveillance strategy with equivalent sampling effort. In the absence of previous quantitative studies investigating the use of sentinel plants for early detection monitoring, in that chapter our aim was to use the simplest possible model to gain insight into the potential benefits and drawbacks of sentinel-based surveillance strategies. However, as discussed in Section 6.4, this simple model did not account for a range of real-world complexities that may influence how sentinel plants should be deployed in practice. One such important realism is how the introduction, transmission, and detection of an invasive pathogen are influenced by seasonally varying factors.

For vector borne pathogens such as *X. fastidiosa*, the risk that the pathogen is introduced to a susceptible population and the subsequent rate of transmission within that population are both strongly influenced by seasonal fluctuations in the vector population size. The primary vector of *X. fastidiosa* in the olive groves of southern Italy is the the meadow spittlebug *Philaenus spumarius*, a xylem sap-feeding insect that is commonly found across Europe and in other temperate regions [310, 326–328]. *P. spumarius* exhibits highly seasonal population dynamics, following an annual life cycle with a single new generation produced per year [263]. Eggs are laid amongst the grasses and weeds that grow beneath olive trees at the end of the summer, with nymphs emerging during March and April the following year [327, 329].

Adult vectors, which are the main source of transmission between susceptible plants due to their increased mobility compared to nymphs [330], then emerge between late April and the end of May and move to the olive canopy during the dry period between May and October [327, 329, 331]. The adult vector population peaks during the summer months before decreasing again in the late autumn due to mortality, with very few adults surviving the winter [263, 329, 332]. This seasonal fluctuation in the abundance of vectors means that the risk of introduction of the pathogen and the subsequent rate of transmission are greatest during the summer months, when the adult vector population is at its largest, and decrease over the winter as the vector population declines.

As well as variation in the risk of introduction and the rate of transmission that arises due to temporal variations in the vector population size, seasonality also influences the rate at which symptoms develop in infected plants. As its taxonomy indicates, *Xylella fastidiosa* colonises the xylem vessels of infected hosts, forming a biofilm that occludes the vessels and restricts water transport within the plant [267, 280, 283, 322, 333]. As the bacterial population increases, the host plant undergoes a physiological response that is akin to water stress, resulting in visible symptoms such as leaf scorch and the desiccation of twigs and branches [267, 280, 334]. The rate at which these symptoms develop depends on the rate of bacterial multiplication within the xylem vessels, which has been shown to vary with temperature. *X. fastidiosa* grows rapidly between 25°C and 32°C, with an optimal growth temperature of approximately 28°C [335]. Outside of this range, growth is less rapid or may not occur at all; the minimum threshold temperature for multiplication of *X. fastidiosa* has been estimated to be approximately 17°C, whilst temperatures below 12°C or above 34°C may affect bacterial survival [335]. Therefore, seasonal variations in temperature result in symptoms developing at different rates throughout the year. For example, in Lecce (the province in Apulia where *X. fastidiosa* has had the most devastating impact in olive groves), the bacteria multiply most rapidly within infected hosts during July and August when the average temperature is 26°C [336]. However, between November and April bacterial growth is inhibited by cooler temperatures, reducing the likelihood that visible symptoms will develop during that period.

In addition to affecting the rate at which the bacteria multiply and colonise the xylem vessels of a host plant, seasonal weather patterns also influence the degree to which the resulting symptoms are expressed. Symptoms associated with *X. fastidiosa* infection such as leaf scorch, withering, and desiccation of twigs and branches are exacerbated by water shortages, since abiotic drought stress augments the existing water stress that is caused by blockage of the xylem vessels [334, 337–340]. Therefore, although persistent throughout the entire year, symptoms of *X. fastidiosa* infection are more strongly expressed during the summer months when there is less precipitation and the soil moisture content is low [322, 334, 337, 341].

The seasonal factors described above suggest that the effectiveness of surveillance strategies that rely on visual inspection as a means of detection may vary throughout the year. Intuitively, pathogen detection is most likely to occur during the warmer months of the year when the rates of pathogen transmission and symptom development are high, and symptoms are more strongly expressed. This motivates considering temporally varying sampling strategies, in which surveillance resources are preferentially allocated to the period of the year when the chance of detection is highest. Indeed, for this reason, EU guidelines currently recommend that visual inspection for symptoms of *X. fastidiosa* infection only takes place between 'late spring' and 'early autumn' [341]. However, quantitative support for these guidelines is lacking, and – ambiguity in the interpretation of 'late spring' and 'early autumn' aside – it is not immediately clear whether this is indeed the optimal strategy to achieve early detection. Concentrating sampling resources into a narrow window of the year increases the chance of detecting the pathogen within that window (if symptomatic plants are present in the population). However, a narrow sampling window has a potential drawback: if the pathogen is not detected during that period, there will be a long delay before the next sample is taken, during which the pathogen may spread. Distributing sampling resources across a wider portion of the year alleviates this problem, but has the disadvantage that sampling effort is diluted and may be wasted if surveillance takes place at times of year when detection is unlikely. This trade-off necessitates a quantitative assessment of the benefits of different sampling windows in order to identify the optimal balance between these conflicting factors, which has not previously been attempted.

Seasonal variation in pathogen transmission and symptom expression may also influence the potential for sentinel plants to facilitate detection at low prevalence. As discussed in Chapter 6, a key drawback of including sentinel plants in a population is the concurrent increased rate of pathogen transmission, which may counteract the benefits of achieving earlier detection if the number of sentinels is not selected carefully. This effect may be exarcerbated when seasonality is accounted for; when the transmission rate is already elevated during the summer months, a further increase may be correspondingly more detrimental. Although a seasonal surveillance strategy in which sampling resources are concentrated into the summer months may counteract this drawback, it is possible that such a strategy could also substantially reduce the EDP in its own right, leaving little scope for sentinels to facilitate further improvement. Therefore, it is not immediately clear whether the benefits of sentinel plants that we observed in Chapter 6 will persist in the seasonal setting. Assessing the optimal deployment of sentinel plants when seasonality in transmission, symptom expression, and the surveillance strategy are accounted for is therefore a critical next step for understanding how sentinels could be used in a realistic context.

In this chapter, we extend the non-seasonal compartmental transmission model that we considered in Chapter 6 (see Section 6.2) to account for seasonality in the risk of pathogen introduction, the rate of pathogen transmission, and the rate of symptom development. Additionally, we extend the sampling algorithm applied in Chapter 6 to incorporate seasonal variation in the sensitivity of visual inspection, and to allow for sampling resources to be preferentially allocated to a specified window of the year. Once again using *X. fastidiosa* infection in *O. europaea* (crop) and *C. roseus* (sentinel) as a case study, we examine whether sentinel plants are still beneficial for achieving early detection when seasonal factors are accounted for, and investigate the optimal window of the year in which to focus limited sampling resources for both sentinel-based and non-sentinel-based strategies.

Our results in this chapter demonstrate that when seasonal factors are accounted for, sentinel plants retain the potential to facilitate substantial reductions in the EDP compared to the baseline level, regardless of whether sampling takes place throughout the year or is instead restricted to the optimal window. The optimal total number of sentinels to include in the population and in the sample varies depending on the sampling effort available, supporting our conclusion from Chapter 6 that these quantities must be chosen carefully based on the precise conditions under which surveillance is taking place. Seasonally focused sampling strategies in which surveillance resources are concentrated into a particular window of the year also lead to substantial reductions in the EDP compared to year-round sampling strategies, provided that the sampling window is chosen optimally. Our results indicate that in the absence of sentinel plants, the lowest detection prevalence is achieved when surveillance resources are focused into the period between late April/early May and late August/mid September. These findings not only provide valuable quantitative support for the existing EU guidelines regarding the timing of visual inspection for *X. fastidiosa*, but also clarify how the terms 'late spring' and 'early autumn' should be interpreted within those guidelines. However, when sentinel plants are included in the population, the optimal window of the year in which to conduct surveillance widens, particularly if the time between samples is long. This result highlights once again that the effects of the individual factors that define a surveillance strategy (which in this context are the sample size, the sample interval, the sampling window, and the number of sentinels in the population and in the sample) cannot be treated as independent, but must instead be considered in the context of the strategy as a whole.

Overall, our conclusions in this chapter provide further support for the use of sentinel plants as a tool to improve early detection monitoring for *X. fastidiosa*, demonstrating that sentinel-based surveillance strategies can still lead to substantial reductions in the EDP when seasonal factors are accounted for. Sentinel plants can be highly effective when used in combination with seasonally focused surveillance strategies, provided that the interacting

effects of the sentinel strategy and the sampling window are carefully evaluated to ensure that the greatest benefits are achieved.

7.2 Methods

7.2.1 Incorporating seasonality into the transmission model

The seasonal transmission model we will use in this chapter is an extension of the non-seasonal compartmental model used in Chapter 6 (see Section 6.2.1) that allows for temporal heterogeneities in:

- the likelihood that the pathogen is introduced to the population at a given time;
- the rate of pathogen transmission within the population; and,
- the rate at which symptoms develop in infected hosts.

In this subsection, we will describe how each of these factors are accounted for in our seasonal model of pathogen transmission. The fourth seasonal factor that we will consider in this chapter – the sensitivity of visual inspection as a means of detection – does not directly affect pathogen transmission, but instead influences the effectiveness of the sampling scheme, as outlined in Subsection 7.2.2.

In the non-seasonal model used in Chapter 6, the force of infection acting on each Healthy plant at time t was given by $\phi(t) = \beta_C(\varepsilon_C U_C(t) + D_C(t)) + \beta_S(\varepsilon_S U_S(t) + D_S(t))$, where $U_C(t)$, $D_C(t)$, $U_S(t)$, and $D_S(t)$ respectively represent the numbers of Undetectable crops, Detectable crops, Undetectable sentinels, and Detectable sentinels present in the population at time t, the parameters β_C and β_S represent the transmission coefficients for Detectable crops and sentinels, and the parameters ε_C and ε_S represent the transmission scaling factors for Undetectable crops and sentinels (compared to Detectable crops and sentinels).

In the seasonal model we will use in this chapter, we allow the transmission rate to vary in time by instead taking the force of infection to be

$$\phi_S(t) = \omega(t)\phi(t),$$

where $\phi(t)$ is as given above, and $\omega(t)$ is a function describing the time-dependent variation in environmental suitability for infection. Here, we assume that the rate of pathogen transmission in the seasonal model is directly proportional to the abundance of adult vectors present in the population at any given time. We take this abundance from Mastin *et al.* [263], who modelled the change in the relative density of adult *P. spumarius* throughout the year by fitting a non-linear regression model to field data [332, 342]. To ensure that the

average transmission rate is consistent with the non-seasonal model used in Chapter 6, we choose $\omega(t)$ to be a scaled version of this fitted vector density that has a mean of one (Fig 7.1A).

We assume that the likelihood that the pathogen is introduced to the population on a particular day of the year is also proportional to the abundance of adult vectors present in the population on that day, and construct a probability mass function for the day of introduction by scaling $\omega(t)$ by the sum of its daily values. In each simulation of the seasonal transmission model, we sample the day of the year that pathogen introduction occurs from this distribution.

In addition to a time-dependent transmission rate and introduction likelihood, we also allow the rate at which symptoms develop in infected hosts to vary seasonally. We incorporate a time-dependent rate of symptom development by multiplying the constant rates of symptom development used in the previous chapter ($1/\gamma_C$, $1/\gamma_S$ for crops and sentinels, respectively) by a factor $\psi(t)$ that represents the seasonal variability in environmental suitability for symptom development. As discussed in the introduction to this chapter, symptoms of *X. fastidiosa* infection occur in an infected host as the pathogen multiplies and the xylem vessels become occluded [267, 280, 283, 322, 333, 334]. The rate at which symptoms develop thus depends on the rate of multiplication of the pathogen within the xylem vessels, which has been shown to be temperature dependent (see Introduction). Therefore, to construct the scaling factor $\psi(t)$, we considered the average monthly temperature in Lecce, which ranges from 9°C in January and February to 26°C in July and August ([336], see also Supplementary Text D.1). In months in which the average temperature is less than 17°C, we set $\psi(t) = 0$ to reflect the fact that the bacteria cannot multiply below that threshold [335]. For consistency with the non-seasonal model used in Chapter 6, we then scaled the twelve resulting monthly values to have a mean of one. Finally, we performed linear interpolation on these monthly values to obtain daily values for $\psi(t)$ (Fig 7.1B, Supplementary Table D.1). The scaled rates $\psi(t)/\gamma_C$ and $\psi(t)/\gamma_S$ then respectively represent the (instantaneous) rates at which crop and sentinel plants progress from being Undetectable to Detectable in the seasonal model.

The resulting compartmental differential equation model reflecting seasonal transmission is given by:

$$\begin{aligned}
\frac{\mathrm{d}H_C}{\mathrm{dt}} &= -\omega\phi H_C, & \frac{\mathrm{d}H_S}{\mathrm{dt}} &= -\omega\phi H_S,\\
\frac{\mathrm{d}U_C}{\mathrm{dt}} &= \omega\phi H_C - \frac{\psi}{\gamma_C}U_C, & \frac{\mathrm{d}U_S}{\mathrm{dt}} &= \omega\phi H_S - \frac{\psi}{\gamma_S}U_S,\\
\frac{\mathrm{d}D_C}{\mathrm{dt}} &= \frac{\psi}{\gamma_C}U_C, & \frac{\mathrm{d}D_S}{\mathrm{dt}} &= \frac{\psi}{\gamma_S}U_S,
\end{aligned}$$

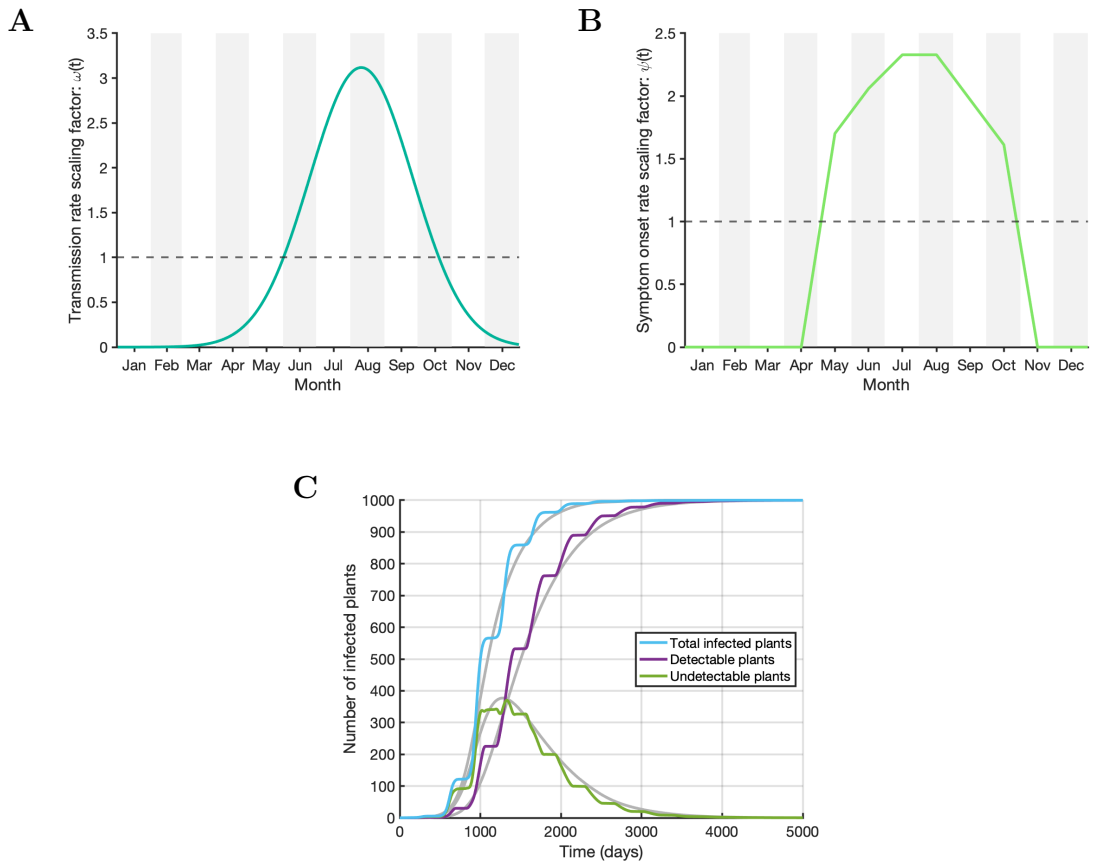


Figure 7.1: Incorporating seasonality in the rates of pathogen transmission and symptom development. A. Seasonal scaling factor for the pathogen transmission rate, $\omega(t)$ (solid blue line), taken to be proportional to the size of the adult vector population at time t [263] and scaled to have a mean of one (black dashed line) for consistency with the non-seasonal model. For each simulation, the time of pathogen introduction is selected randomly from a probability distribution of the same shape. **B.** Seasonal scaling factor for the rate at which symptoms develop in Undetectable plants, $\psi(t)$ (solid green line), constructed to reflect temperature-dependent variation in the bacterial growth rate (see Section 7.2.1 and Supplementary Text D.1) and scaled to have a mean of one (black dashed line) for consistency with the non-seasonal model. **C.** Daily numbers of infected plants (blue line), Detectable plants (purple line) and Undetectable plants (green line) in the seasonal baseline transmission model (Section 7.2.3), averaged over $1,000$ simulations with model parameters as specified in Table 6.2. Equivalent results from the non-seasonal model are plotted in grey for comparison.

where $\phi(t)$ is the non-seasonal force of infection acting on each Healthy plant, as used in Chapter 6 (see Equation (6.1)); $\omega(t)$ is the seasonal scaling factor for the rate of pathogen transmission (reflecting changes in the vector population size); and $\psi(t)$ is the seasonal scaling factor for the rate of symptom development (based on variations in precipitation throughout the year). The model parameters β_C, β_S, ε_C, ε_S, γ_C, and γ_S remain at their values used in the main analyses conducted in Chapter 6 (see Table 6.2).

Since stochasticity in the early outbreak is an important factor affecting the detection

prevalence, in our analyses we do not compute epidemic curves using this ODE model. Instead, we run simulations of the analogous stochastic model using the Modified Next Reaction Method (MNRM), a stochastic simulation algorithm that allows the reaction rates to vary in time [343, 344]. The average epidemic curve obtained by performing $1,000$ simulations using this method, with the number of sentinel plants set to $P_S = 0$ and model parameters as specified in Table 6.2, is shown in Fig 7.1C. Each of these simulations began with a single infected individual, with the time of pathogen introduction selected according to the probability distribution described earlier in this subsection.

7.2.2 Incorporating seasonality in symptom expression

In the previous chapter, we assumed that the sensitivity of visual inspection was 100% for Detectable plants and 0% for Undetectable plants, so that detection occurred at a given sample time if and only if at least one Detectable plant (crop or sentinel) was included in the sample selected at that time. In the seasonal model, we still assume that Undetectable plants are never correctly identified as being infected (0% sensitivity). However, we now allow the sensitivity of visual inspection for Detectable plants to vary seasonally. As discussed in the introduction to this chapter, although symptoms of *X. fastidiosa* infection persist throughout the year, they are more strongly expressed during the summer months [322, 334, 337, 341]. At that time of year, precipitation is reduced and the resulting increase in water stress enhances the severity of the characteristic leaf scorch symptoms [334, 337–340]. Here, we therefore assume that the sensitivity of visual inspection varies throughout the year according to the degree of water stress experienced by the host plants, and use data on the average monthly rainfall in Lecce [336] (blue dashed line in Fig 7.2) to construct a function $\chi(t)$ describing the sensitivity of visual inspection at time t (solid orange line in Fig 7.2) (see Supplementary Text D.2 for further details). We incorporate this sensitivity computationally by specifying that a Detectable plant sampled at time t is correctly identified as infected with probability $\chi(t)$.

Throughout this chapter, we will use the phrase 'full seasonal model' to refer to the stochastic version of the seasonal transmission model described in Section 7.2.1, coupled with the seasonal variation in detection sensitivity outlined above. For the reader's convenience, a summary of the terminology that is introduced in this Section and used throughout this chapter is provided in Table 7.1.

7.2.3 The seasonal baseline – model in the absence of sentinel plants

In the absence of sentinel plants ($P_S = 0$), the transmission model reduces to:

$$\frac{\mathrm{d}H_C}{\mathrm{dt}} = -\omega\phi_b H_C, \quad \frac{\mathrm{d}U_C}{\mathrm{dt}} = \omega\phi_b H_C - \frac{\psi}{\gamma_C}U_C, \quad \frac{\mathrm{d}D_C}{\mathrm{dt}} = \frac{\psi}{\gamma_C}U_C,$$

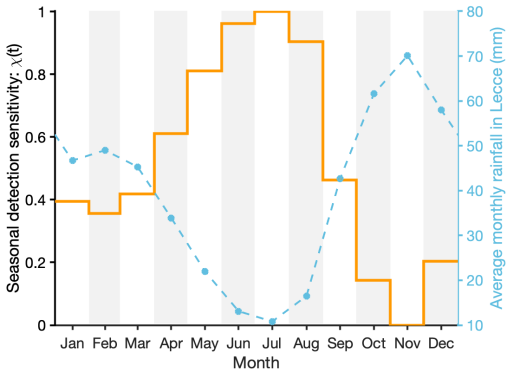


Figure 7.2: Incorporating seasonality in the strength of symptom expression. The monthly sensitivity of visual inspection as a means of pathogen detection, $\chi(t)$, constructed as described in Section 7.2.2 and Supplementary Text D.2 (solid orange line; left hand axis). A Detectable plant contained within a sample selected at time t is correctly identified as infected with probability $\chi(t)$. The average monthly rainfall in Lecce (mm) from which $\chi(t)$ is constructed is overlaid for comparison (blue dashed line; right hand axis).

where, as in Chapter 6, the force of infection ϕ_b acting on each Healthy plant is

$$\phi_b = \beta_C(\varepsilon_C U_C + D_C). \tag{7.1}$$

Throughout, this reduced system coupled with the seasonal detection sensitivity described in Section 7.2.2 is what we will refer to as the 'seasonal baseline model' for a particular choice of sampling strategy. As in Chapter 6, in this chapter the effect of including sentinel plants in a surveillance strategy will always be assessed relative to the seasonal baseline with equivalent parameterisation and sampling effort. Again, when considering the seasonal baseline model, we run simulations of the analogous stochastic model using the MNRM.

7.2.4 Allowing for 'window' sampling strategies

Previously, we have only considered non-seasonal surveillance strategies in which samples of size N are selected every Δ days throughout the year. We refer to this type of strategy as a 'uniform' sampling strategy, because sampling resources are spread uniformly in time.

In this chapter, we extend the uniform sampling algorithm used in Chapter 6 to allow for seasonal surveillance strategies, in which sampling only takes place during a fixed specified window of the year, $[\Delta_{\text{start}}, \Delta_{\text{end}}]$ (Fig 7.3). Here, Δ_{start} and Δ_{end} represent the days of the year on which the sampling window opens and closes (where day 0 corresponds to 1$^{\text{st}}$ January and day 364 corresponds to 31$^{\text{st}}$ December throughout). As in Chapter 6, monitoring begins at a random time δ relative to the time of primary infection, where δ is drawn from a $U[0, \Delta]$ distribution. Potential sampling times subsequently arise at regular time intervals

Δ, but random samples of size N are only selected from the population at a sample time T if $T \in [\Delta_{\text{start}}, \Delta_{\text{end}}]$ (Fig 7.3). The uniform sampling algorithm can be recovered by setting $\Delta_{\text{start}} = 0$, $\Delta_{\text{end}} = 364$.

Throughout, we will refer to this type of strategy as a 'window' sampling strategy. We will make use of this type of strategy in Results Sections 7.3.2 and 7.3.3, in which we aim to identify the optimal choice of sampling window $[\Delta_{\text{start}}, \Delta_{\text{end}}]$.

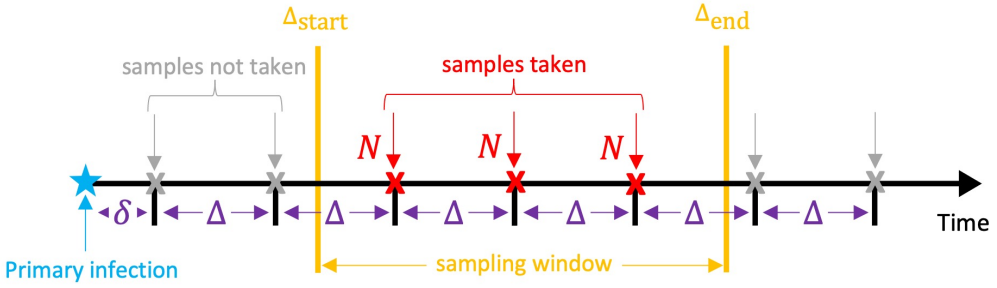


Figure 7.3: Schematic illustrating the implementation of a 'window' sampling strategy. Monitoring begins at a random time δ relative to the time of primary infection, where δ is drawn from a $U[0, \Delta]$ distribution. Potential sampling times subsequently arise at regular time intervals Δ, but random samples of size N are only selected from the population at sample times that lie within the sampling window $[\Delta_{\text{start}}, \Delta_{\text{end}}]$. As in Chapter 6, sampling continues until detection occurs, at which point the EDP is recorded.

7.2.5 Sampling effort

For a uniform surveillance strategy of the type considered in Chapter 6, in which samples of size N are selected every Δ days throughout the year, we define the sampling effort to be $E = N/\Delta$. Larger samples or more frequent samples therefore correspond to a larger sampling effort, or correspondingly, a larger required sampling resource [264, 265]. A contour plot of the sampling effort corresponding to the range of pairs of values of $N = 25, 30, 35, \cdots, 200$, $\Delta = 30, 35, 40, \cdots, 150$ days is shown in Fig 7.4.

When comparing the effects of different window strategies relative to each other and/or to a uniform sampling strategy, we fix the total annual sampling effort E at a given value to ensure that we are comparing strategies that utilise the same total amount of sampling resource. For a uniform sampling strategy, with a given sampling effort E and fixed sample interval Δ, we choose the sample size N to satisfy $N/\Delta = E$ and perform uniform sampling accordingly. For a window sampling strategy of equivalent effort and the same sample interval Δ, we concentrate the available effort E into the specified window $[\Delta_{\text{start}}, \Delta_{\text{end}}]$ by

adjusting the sample size N according to the following equation:

$$\frac{N}{\Delta} \times \frac{\Delta_{\text{end}} - \Delta_{\text{start}} + 1}{365} = E. \tag{7.2}$$

This ensures that the sampling effort employed within the window multiplied by the proportion of the year that the window represents is equal to the total sampling effort. In either case, if the calculated sample size N exceeds the total population size P, we instead set $N = P$.

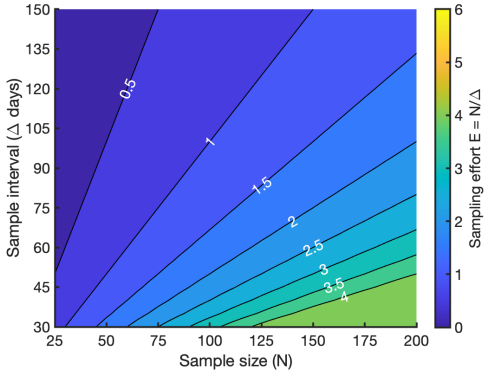


Figure 7.4: The sampling effort $E = N/\Delta$ for sample sizes $N = 25, 30, 35, \cdots, 200$ and sample intervals $\Delta = 30, 35, 40, \cdots, 150$ days. Black solid lines mark contours along which the sampling effort E is constant. Larger sampling efforts E correspond to larger sample sizes N (inspecting a larger number of plants per sample) and/or smaller sample intervals Δ (sampling more frequently).

7.2.6 Application of the seasonal model

We use the seasonal model described in the previous subsections to answer the following three questions:

- How does seasonality in pathogen transmission and symptom expression affect the conclusions drawn in Chapter 6 regarding the use of sentinel plants to facilitate early detection?
- What is the optimal calendar time window in which to perform surveillance in the absence of sentinels, based on these seasonal factors?
- Are sentinel plants still beneficial when used in combination with a seasonally focused surveillance strategy, and if so, how is the optimal sampling window affected by their inclusion?

In this subsection, we provide technical details of the analyses that we conducted in this chapter to address each of these questions in turn.

Term	Definition
Full seasonal model	The stochastic version of the two-species seasonal transmission model described in Section 7.2.1, coupled with the seasonal detection sensitivity described in Section 7.2.2.
Seasonal baseline model	The stochastic version of the reduced transmission model described in Section 7.2.3 (no sentinels), coupled with the seasonal detection sensitivity described in Section 7.2.2.
Uniform sampling strategy	A strategy in which samples of size N are selected every Δ days throughout the year (i.e. sampling resources are spread uniformly in time).
Window sampling strategy	A strategy in which sampling only takes place during a fixed specified window of the year, $[\Delta_{\text{start}}, \Delta_{\text{end}}]$, and available surveillance resources are concentrated into that period.
Uniform Seasonal Baseline EDP (USB)	The EDP achieved by applying a uniform sampling strategy to the seasonal baseline model.
Optimal Window Baseline EDP (OWB)	The best achievable EDP when a window sampling strategy is applied to the seasonal baseline model (i.e. $[\Delta_{\text{start}}, \Delta_{\text{end}}]$ is optimised).
Optimal Sentinel-based Window EDP	The best achievable EDP when a window sampling strategy is applied to the full seasonal model (i.e. $[\Delta_{\text{start}}, \Delta_{\text{end}}]$, P_S^*, and N_S^* are optimised).

Table 7.1: Summary of the terminology used to describe the seasonal models, sampling strategies and resulting EDPs considered in this chapter.

The effect of seasonal factors on the conclusions of Chapter 6

Initially, we considered how incorporating the seasonal factors described in Sections 7.2.1 and 7.2.2 affected the conclusions drawn in Chapter 6 regarding the effects of uniform sentinel-based surveillance strategies on the EDP (see Section 7.3.1). To do this, we followed an analogous procedure to that described in Section 6.2.5 to explore how the optimal number of sentinels to include in the population varied with the choice of sample size and sample interval, using the seasonal transmission model described in Section 7.2.1 and accounting for seasonal variation in detection sensitivity as outlined in Section 7.2.2. We began by computing the seasonal baseline EDP (i.e. the baseline EDP in the seasonal model without sentinels, see Section 7.2.3) by applying a uniform sampling strategy for sample sizes $N = 25, 30, 35, \cdots, 200$ and sample intervals $\Delta = 30, 35, 40, \cdots, 150$ days. Throughout, we will refer to the EDP achieved by applying a uniform sampling strategy to the seasonal baseline model as the 'uniform seasonal baseline' EDP (abbreviated to USB). Then, as in Section 6.2.5, for each of these (N, Δ) pairs we allowed the number of sentinels in the population (P_S) and in the sample (N_S) to vary simultaneously, and applied a Bayesian optimisation algorithm to identify the pair of values (P_S^*, N_S^*) that maximised the reduction in EDP compared to the seasonal baseline. As in Chapter 6, for each pair of (N, Δ) values

we allowed the total number of sentinels added to the population (P_S) to vary in the range $[0, 350]$, and the number of sentinels included in the sample (N_S) to vary in the range $[0, \min(P_S, N)]$. The results of this analysis are shown in Section 7.3.1 (Fig 7.5).

Optimising the sampling window for the seasonal baseline model

Next, we considered how the seasonal factors described in Sections 7.2.1 and 7.2.2 affected the optimal allocation of sampling resources across the year in the absence of sentinels (Section 7.3.2). Specifically, we investigated how the seasonal baseline EDP was affected by employing a window surveillance strategy in which sampling was restricted to a given window of the year, compared to a uniform surveillance strategy with the same annual sampling effort (see Sections 7.2.4 and 7.2.5).

We considered eight fixed values for the total annual sampling effort: $E = 0.5, 1, 1.5, \cdots, 4$. In each case, we allowed the sample interval Δ to take the values $\Delta = 7, 14, 30$, and 60 days. For a given pair of values (E, Δ) we first computed the seasonal baseline EDP using a uniform sampling strategy (USB) in which the sample size N was defined by $N/\Delta = E$. Then, for the same (E, Δ) pair we computed the seasonal baseline EDP using a window sampling strategy, for all sampling windows $[\Delta_{\text{start}}, \Delta_{\text{end}}]$ such that $\Delta_{\text{start}} \in \{0, 10, 20, \cdots, 160\}$ and $\Delta_{\text{end}} \in \{200, 210, \cdots, 360\}$. In every case, the sample size N was selected to satisfy Equation (7.2) (unless by this calculation N exceeded the total population size P, in which case we set $N = P$). The optimal sampling window for that (E, Δ) was then taken to be the window $[\Delta^*_{\text{start}}, \Delta^*_{\text{end}}]$ which led to the largest reduction in the seasonal baseline EDP compared to the uniform sampling strategy with equivalent effort. Throughout, we will refer to the seasonal baseline EDP achieved using the optimal window sampling strategy as the 'optimal window baseline' EDP (abbreviated to OWB). The results of this analysis are shown in Section 7.3.2 (Figs 7.6 and 7.7).

Optimising the sampling window for the full seasonal model

Finally, we combined the two approaches described above to investigate the optimal allocation of sampling resources across the year when sentinels are present in the population (Section 7.3.3). Exactly as above, we considered eight fixed values for the total annual sampling effort: $E = 0.5, 1, 1.5, \cdots, 4$, and allowed the sample interval to take the values $\Delta = 7, 14, 30$, and 60 days. Once again, for each (E, Δ) pair we considered all sampling windows $[\Delta_{\text{start}}, \Delta_{\text{end}}]$ such that $\Delta_{\text{start}} \in \{0, 10, 20, \cdots, 160\}$ and $\Delta_{\text{end}} \in \{200, 210, \cdots, 360\}$, and selected the total sample size N to satisfy Equation (7.2) (unless by this calculation N exceeded the total population size P, in which case we set $N = P$). For every (E, Δ) pair and sampling window $[\Delta_{\text{start}}, \Delta_{\text{end}}]$ that we considered, we allowed the total number of sentinels added to the population (P_S) to vary in the range $[0, 350]$ and the number of sentinels included in the sample (N_S) to vary in the range $[0, \min(P_S, N)]$, and as described

above applied a Bayesian optimisation algorithm to identify the pair of values (P_S^*, N_S^*) that minimised the EDP. For a given (E, Δ) pair, the optimal sampling window was taken to be the window $[\Delta_{\text{start}}^*, \Delta_{\text{end}}^*]$ for which this minimised EDP was the smallest.

For each (E, Δ) pair we recorded the best achievable EDP, corresponding to the optimal choice of sampling window, $[\Delta_{\text{start}}^*, \Delta_{\text{end}}^*]$, and the optimal number of sentinels to include in the population and in the sample, (P_S^*, N_S^*). Throughout, we will refer to this EDP as the 'sentinel-based window EDP' for that (E, Δ). We expressed this EDP as a percentage of the uniform seasonal baseline EDP (USB) obtained for the same choice of E and Δ. Then, we directly compared this quantity with the equivalent results achieved by applying the optimal window strategy in the absence of sentinels (from Section 7.3.2) to investigate the marginal benefit of including sentinel plants in addition to focusing surveillance resources at the optimal time of year. The results of this analysis are shown in Section 7.3.3 (Figs 7.8 and 7.9).

Since in practice there is some uncertainty in the most appropriate construction of the seasonal scaling factors $\omega(t)$, $\psi(t)$, and $\chi(t)$, we performed additional analyses to investigate the effects of varying these factors from their values described in Sections 7.2.1 and 7.2.2, as outlined in Supplementary Text D.3. The results of these additional analyses are shown in Supplementary Figs D.1, D.2, and D.3.

7.2.7 Code availability

All computing code used to implement the above methods was written in MATLAB (version R2022a), and is available at https://github.com/francescalovellread/seasonal_surveillance.

7.3 Results

7.3.1 The effect of seasonal factors on the conclusions of Chapter 6

We first considered the effects of implementing uniform sampling strategies (see Section 7.2.4) on the seasonal baseline model (i.e. the seasonal model without sentinels, see Section 7.2.3) (Fig 7.5A). These results are analogous to those shown in Fig 6.4C, but accounting for the seasonal factors described in Sections 7.2.1 and 7.2.2 that affect pathogen transmission and symptom expression. As expected, lower EDPs were achieved with larger sample sizes N and smaller sample intervals Δ. For every choice of (N, Δ), a higher EDP was observed for the seasonal baseline than for the non-seasonal baseline (Figs 7.5A and 6.4C). This can be attributed to the reduced sensitivity of detection throughout most of the year in the seasonal model compared to the original model used in Chapter 6. Similarly to before, the seasonal baseline EDP values provide a point of comparison that can be used to evaluate the relative effects of sentinel-based strategies with a given sample size and sample interval.

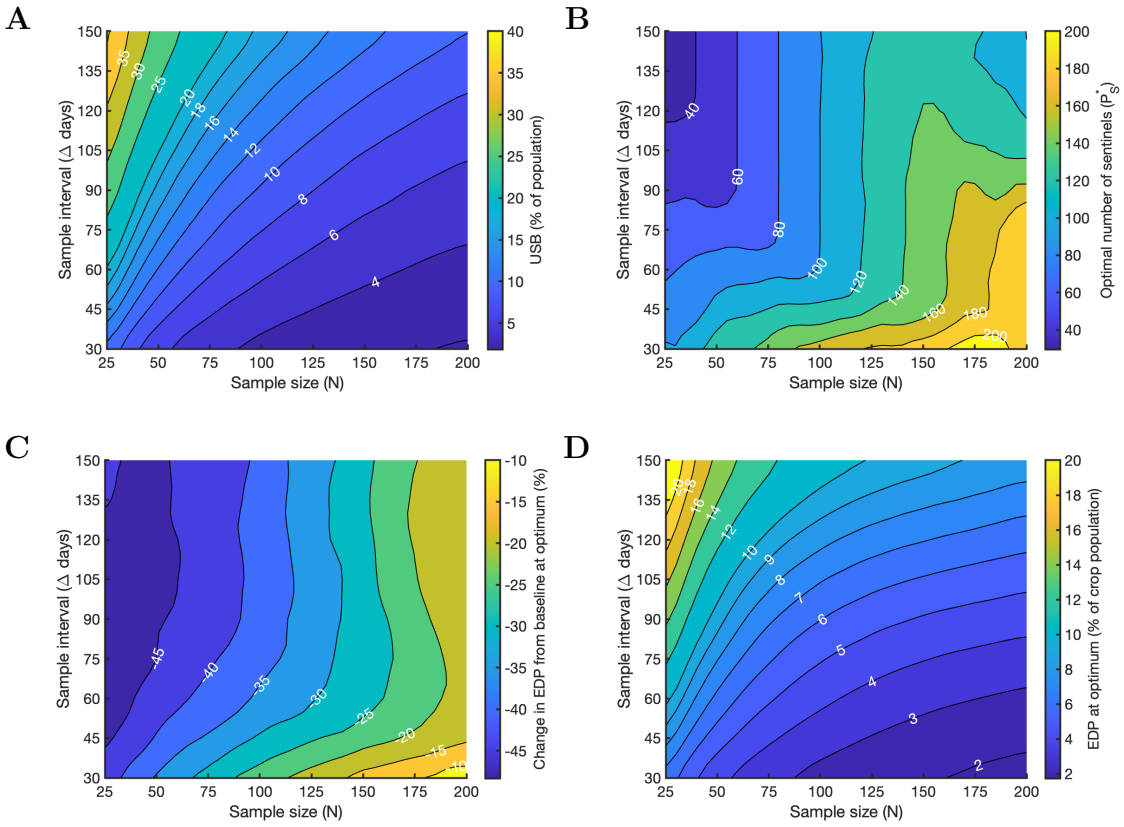


Figure 7.5: Optimising the number of sentinels to include in the population for the seasonal model with uniform sampling. A. The uniform seasonal baseline EDP (USB), expressed as a percentage of the total crop population size, as the sample size ($N = N_C$) and sample interval (Δ) vary. **B.** The optimal number of sentinel plants to include in the population, P_S^*, for which the maximal reduction in the EDP compared to the baseline level is achieved (if N_S is also chosen optimally). **C.** The percentage change in the EDP compared to the baseline value at the optimum, achieved when $P_S = P_S^*$ and $N_S = N_S^*$. **D.** The resultant value of the EDP at the optimum, expressed as a percentage of the total crop population.

Next, we investigated how the seasonal factors described in Sections 7.2.1 and 7.2.2 affected the conclusions drawn in Chapter 6 regarding the use of sentinel plants as a tool to reduce the EDP compared to the seasonal baseline. Specifically, as described in Section 7.2.6, we applied uniform sampling strategies to the full seasonal model and used a Bayesian optimisation algorithm to determine the optimal number of sentinels to include in the population (P_S^*) and in the sample (N_S^*), as well as the corresponding change in the EDP compared to the seasonal baseline and the resultant EDP at the optimum (Figs 7.5B, 7.5C, 7.5D). These results are analogous to those shown in Figs 6.8, 6.10A, and 6.10B, but accounting for seasonal variations in the probability of pathogen introduction, rate of pathogen transmission, rate of symptom development, and strength of symptom expression.

When these seasonal factors were accounted for, our results remained consistent with those

obtained from the non-seasonal model in Chapter 6. In the full seasonal model, including sentinel plants in the population was beneficial across the range of sample sizes and sample intervals considered (Figs 7.5B, 7.5C). The optimal number of sentinels added to the population (P_S^*) was greater than zero for all values of (N, Δ) (Fig 7.5B), with corresponding reductions in the EDP compared to the seasonal baseline ranging between 7% and 48% (Fig 7.5C). Therefore, when considering the seasonal model, our main conclusion from Chapter 6 remains unchanged: sentinel plants have the potential to facilitate substantial reductions in the EDP for a wide range of sampling efforts. As in all previous cases, the resultant EDP decreased with greater sampling effort: increasing the sample size (N) and/or decreasing the sample interval (Δ) always led to a smaller EDP (Fig 7.5D).

As in Chapter 6, the optimal number of sentinels to include in the population (P_S^*) varied substantially with the sample size (N) and sample interval (Δ) (Fig 7.5B). These results reinforce our previous conclusion that the potential impact of sentinel strategies should be carefully assessed in advance according to the sampling resources that are available. For all of the (N, Δ) values considered, the optimal number of sentinels to include in the sample (N_S^*) was the maximum possible (i.e. $N_S^* = \min(P_S, N)$).

7.3.2 Optimising the sampling window for the seasonal baseline model

Next, we returned to the seasonal baseline model to explore the effects of window sampling strategies compared to uniform sampling strategies on the EDP. Specifically, we wished to investigate whether focusing surveillance resources into a particular window of the year can be advantageous, and to explore whether we could identify an optimal such window for sampling to take place.

As outlined in Section 7.2.6, we considered eight fixed values for the total annual sampling effort ($E = 0.5, 1, 1.5, \cdots, 4$) and four fixed values for the sampling interval ($\Delta = 7, 14, 30,$ and 60 days). For each pair of values (E, Δ) we first computed the seasonal baseline EDP using a uniform sampling strategy (the 'uniform seasonal baseline' EDP, or 'USB') (Fig 7.6A). Then, for each (E, Δ) pair we computed the seasonal baseline EDP using a window sampling strategy, for all sampling windows $[\Delta_{\text{start}}, \Delta_{\text{end}}]$ such that $\Delta_{\text{start}} \in \{0, 10, 20, \cdots, 160\}$ and $\Delta_{\text{end}} \in \{200, 210, \cdots, 360\}$. We recorded the choice of sampling window $[\Delta^*_{\text{start}}, \Delta^*_{\text{end}}]$ that maximised the reduction in EDP compared to the USB with the same (E, Δ).

For example, in the case $E = 2$, $\Delta = 15$ days, the maximum reduction in the seasonal baseline EDP compared to the USB achieved with the same (E, Δ) occurred when $\Delta_{\text{start}} = 120$, $\Delta_{\text{end}} = 260$, representing a sampling window beginning on 1$^{\text{st}}$ May and ending on 18$^{\text{th}}$ September (Fig 7.6B). This choice of sampling window (indicated by the black square) led to a 39% reduction in the EDP compared to the USB obtained with the same E and Δ.

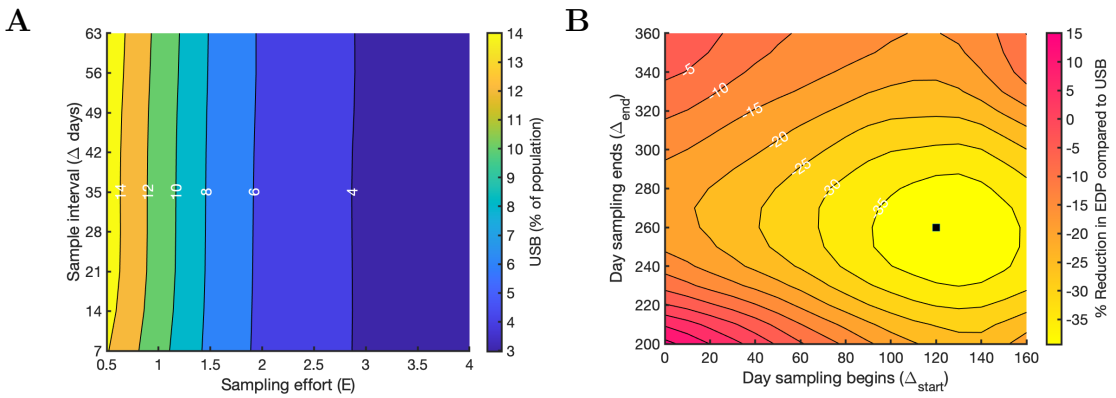


Figure 7.6: Optimising the sampling window for the seasonal baseline model. A. The seasonal baseline EDP obtained using a uniform sampling strategy (the 'uniform seasonal baseline' EDP, or 'USB') as the sampling effort (E) and the sample interval (Δ) vary, expressed as a percentage of the total population size. **B.** The effect of varying the sampling window $[\Delta_{\text{start}}, \Delta_{\text{end}}]$ on the percentage reduction in the EDP compared to the USB, in the case $E = 2$, $\Delta = 15$ days. The sampling window for which the reduction in EDP is maximised ($\Delta^*_{\text{start}} = 120$, $\Delta^*_{\text{end}} = 260$) is indicated by the black square.

For alternative values of Δ_{start} and Δ_{end}, smaller reductions (or even increases) in the EDP compared to the USB were achieved. This demonstrates that selecting the sampling window carefully is critical to ensure that the available sampling resources are utilised most efficiently.

The optimal choice of sampling window for the range of values of E and Δ considered is shown in Fig 7.7A, with the corresponding reductions in the EDP compared to the seasonal baseline for the same (E, Δ) shown in Fig 7.7B. For all values of (E, Δ) considered, the optimal sampling window $[\Delta^*_{\text{start}}, \Delta^*_{\text{end}}]$ began in late April/early May and ended in late August/mid September (Fig 7.7A). Specifically, in every case considered, Δ^*_{start} lay between day 110 (21$^{\text{st}}$ April) and day 130 (11$^{\text{th}}$ May), whilst corresponding values of Δ^*_{end} lay between day 240 (29$^{\text{th}}$ August) and day 260 (18$^{\text{th}}$ September). During that period, the sensitivity of visual inspection as a means of detection ($\chi(t)$) is at its highest (Fig 7.2), increasing the likelihood that a given sample will lead to identification of the pathogen. By restricting sampling to this time of year, sampling effort is not wasted in the winter months when the inspection sensitivity is low and detection is unlikely. The rate of pathogen transmission ($\omega(t)$) is also largest between late spring and early autumn, meaning that the number of infected plants increases most rapidly during that time (Fig 7.1A). Sampling intensively during this period of rapid increase makes sense, because earlier detection has a larger impact on the overall detection prevalence when the transmission rate is high. This is further motivated by the high rate of symptom development in the warmer months, which similarly leads to a rapid increase in the number of Detectable plants during this time (Fig 7.1B).

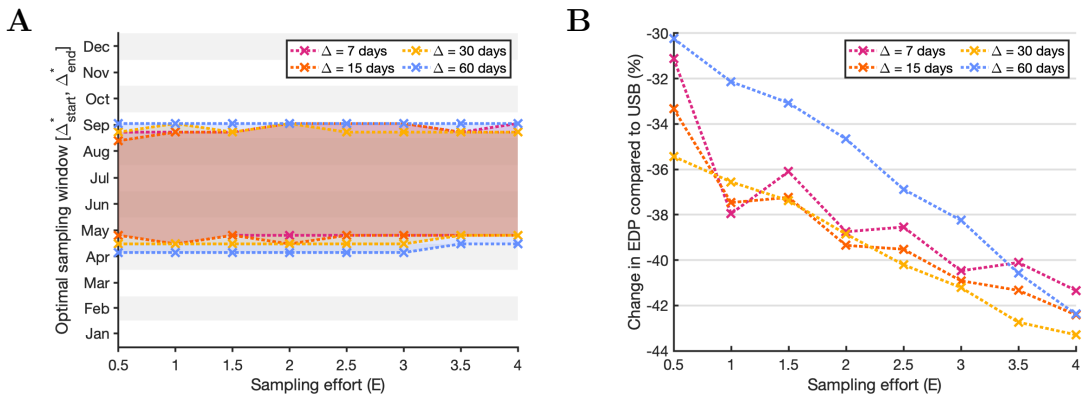


Figure 7.7: Optimising the sampling window for the seasonal baseline model. A. The optimal sampling window $[\Delta^*_{\text{start}}, \Delta^*_{\text{end}}]$ for each (E, Δ) pair considered (coloured shaded regions). Values of Δ_{start}, Δ_{end} are translated into calendar dates and plotted on a monthly scale for ease of interpretation. **B.** The best achievable percentage changes in the EDP compared to the USB for each (E, Δ) pair, corresponding to the optimal sampling windows identified in panel A. Note that the USB depends on E and Δ, so the relative changes in EDP shown here are not a measure of the resultant EDP.

For each choice of (E, Δ), when the sampling window was selected optimally (as indicated in Fig 7.7A) a reduction in the EDP compared to the USB was achieved (Fig 7.7B). Greater reductions in the EDP were achieved for larger sampling efforts E, when very intensive sampling occurred within the optimal window. In Chapter 6, we observed that larger sampling efforts (i.e. larger sample sizes N and smaller sample intervals Δ) limited the potential for sentinel plants to reduce the EDP compared to the baseline level. When the effort was high, the baseline EDP was low, and the detrimental effects of sentinel plants outweighed their benefits. However, in the absence of sentinels the same limitations do not apply: a low seasonal baseline EDP does not necessarily limit the capacity for further reduction that is achieved by narrowing the sampling window.

Here, a sample interval of $\Delta = 60$ days generally led to smaller reductions in the EDP than were achieved for shorter sample intervals. This is because in each simulation the first sample time was selected at random from a $U[0, \Delta]$ distribution, with potential sample times subsequently arising every Δ days after the first sample time (Fig 7.3). Therefore, when the sampling window opened the next sample was not necessarily taken immediately; instead, it could be taken at any time up to Δ days later. When Δ is small, this makes little difference to the resulting EDP. However, for large Δ this approach may lead to a worse performance, since the samples may not be taken at the optimal time within the window. In particular, the random timing of the sampling scheme relative to the sampling window may result in fewer samples being taken than could theoretically fit inside the window, meaning that some of the available sampling resources may be wasted.

7.3.3 Optimising the sampling window for the full seasonal model

Finally, we considered the effects of window sampling strategies in the full seasonal model, i.e. with sentinel plants present in the population. Specifically, for a given choice of sampling effort (E) and sample interval (Δ) we sought to optimise the sampling window $[\Delta_{\text{start}}, \Delta_{\text{end}}]$ and the sentinel strategy (choice of P_S and N_S) simultaneously. Considering the effects of window strategies used in combination with sentinel plants raises two important questions that we address here:

i) **Are sentinel plants still beneficial when surveillance resources are focused on a particular time of year?** In Chapter 6, we demonstrated that including sentinel plants in the population could lead to substantial reductions in the EDP compared to the baseline case, but that sentinels were less beneficial when the sampling effort was high and the corresponding baseline EDP was low. In that case, the earlier detection facilitated by sentinel plants was not enough to offset the increase in the transmission rate they caused. In Section 7.3.2, we showed that in the absence of sentinels, window sampling strategies also have the potential to facilitate substantial reductions in the EDP compared to uniform strategies. Therefore, employing a window strategy may similarly limit the scope for sentinels to further reduce the EDP, and it is not immediately clear whether sentinels will still have a beneficial effect.

ii) **Is the optimal choice of sampling window altered when sentinels are present in the population?** In Section 7.3.2, we demonstrated that in the absence of sentinels the optimal sampling window lies between late spring and early autumn, which is consistent with the current EU recommendations regarding the timing of visual surveys for *X. fastidiosa*. However, since sentinel plants display visible symptoms more rapidly than crop plants, the time of year when detection is more likely to occur may vary, which could alter the optimal timing of surveillance.

As in Section 7.3.2, we considered eight fixed values for the total annual sampling effort ($E = 0.5, 1, 1.5, \cdots, 4$), four fixed values for the sample interval ($\Delta = 7, 14, 30$, and 60 days), and all sampling windows $[\Delta_{\text{start}}, \Delta_{\text{end}}]$ such that $\Delta_{\text{start}} \in \{0, 10, 20, \cdots, 160\}$ and $\Delta_{\text{end}} \in \{200, 210, \cdots, 360\}$. For every (E, Δ) pair and sampling window $[\Delta_{\text{start}}, \Delta_{\text{end}}]$ that we considered, we optimised the total number of sentinels included in the population (P_S) and in the sample (N_S), and identified the sampling window $[\Delta^*_{\text{start}}, \Delta^*_{\text{end}}]$ and corresponding sentinel strategy (P^*_S, N^*_S) that minimised the EDP.

For the previously used example in which $E = 2$, $\Delta = 15$ days, the maximum reduction in the EDP in the sentinel-based window model compared to the uniform seasonal baseline with the same (E, Δ) occurred for $[\Delta^*_{\text{start}}, \Delta^*_{\text{end}}] = [130, 280]$, representing a sampling window beginning on 11$^{\text{th}}$ May and ending on 8$^{\text{th}}$ October (Fig 7.8A). The corresponding optimal number of sentinels to include in the population was $P^*_S = 194$ (Fig 7.8B), all of which were

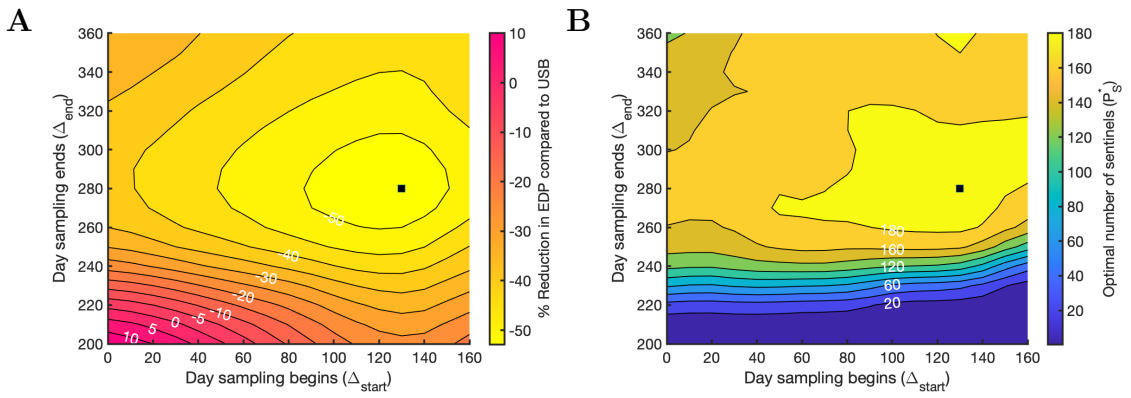


Figure 7.8: Optimising the sampling window and the number of sentinels present in the population simultaneously, for the example case $E = 2$, $\Delta = 15$ days. A. The effect of varying the sampling window $[\Delta_{\text{start}}, \Delta_{\text{end}}]$ on the EDP obtained by applying the optimal sentinel-based strategy. Values are expressed as a percentage change relative to the uniform seasonal baseline EDP (USB). The sampling window for which the reduction in EDP compared to the USB is maximised ($\Delta^*_{\text{start}} = 130$, $\Delta^*_{\text{end}} = 280$) is indicated by the black square. **B.** The effect of varying the sampling window $[\Delta_{\text{start}}, \Delta_{\text{end}}]$ on the optimal number of sentinels to include in the population (P_S^*). As in panel A, the black square indicates the sampling window for which the reduction in EDP is maximised ($\Delta^*_{\text{start}} = 130$, $\Delta^*_{\text{end}} = 280$).

included in the sample (i.e. $N_S^* = P_S^*$). This choice of $[\Delta_{\text{start}}, \Delta_{\text{end}}]$, P_S, and N_S led to a 53% reduction in the EDP compared to the uniform seasonal baseline obtained with the same E and Δ.

The optimal choice of sampling window ($[\Delta^*_{\text{start}}, \Delta^*_{\text{end}}]$) for the range of values of E and Δ considered is shown in Fig 7.9A, with the corresponding optimal number of sentinels to include in the population (P_S^*) shown in Fig 7.9B. We also plotted the value of N_S^* expressed as a proportion of its maximum possible value, $\min(P_S^*, N)$ (where the sample size N is calculated by setting $[\Delta_{\text{start}}, \Delta_{\text{end}}] = [\Delta^*_{\text{start}}, \Delta^*_{\text{end}}]$ in Equation (7.2)) (Fig 7.9C). For each (E, Δ) pair, we then computed the marginal benefit of utilising the optimal sentinel strategy in addition to optimising the sampling window (Fig 7.9D). Specifically, we plotted the difference between the EDP obtained when the optimal sentinel-based window strategy was applied (as in this Section) and the EDP obtained when the optimal window strategy in the absence of sentinels was applied (as in the previous Section), expressed as a percentage of the uniform seasonal baseline EDP (USB). The values shown in Fig 7.9D therefore represent the additional reductions in the EDP that are achieved when the sentinel strategy and sampling window are optimised simultaneously, compared to the reductions in EDP shown in Fig 7.7B that are achieved solely by optimising the sampling window.

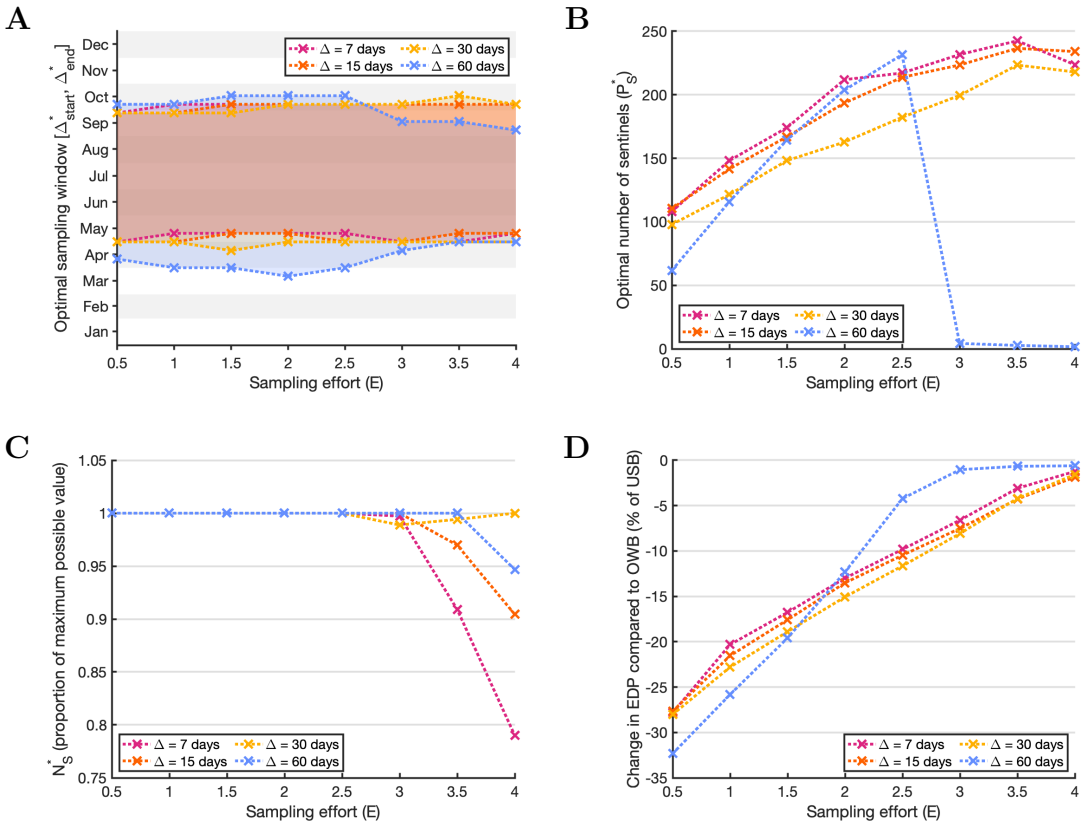


Figure 7.9: Optimising the sampling window and the number of sentinels present in the population simultaneously, for the range of (E, Δ) values considered. A. The optimal sampling window $[\Delta^*_{\text{start}}, \Delta^*_{\text{end}}]$ when the numbers of sentinels present in the population and sample are also optimised, for each (E, Δ) pair considered (coloured shaded regions). Values of Δ_{start}, Δ_{end} are translated into calendar dates and plotted on a monthly scale for ease of interpretation. **B.** The optimal number of sentinels to include in the population (P^*_S) when the sampling window $[\Delta^*_{\text{start}}, \Delta^*_{\text{end}}]$ is also optimised, for each (E, Δ) pair considered. **C.** The optimal number of sentinels to include in the sample (N^*_S) for each (E, Δ) pair considered, expressed as a proportion of its maximum possible value, $\min(P^*_S, N)$ (where the sample size N is calculated by setting $[\Delta_{\text{start}}, \Delta_{\text{end}}] = [\Delta^*_{\text{start}}, \Delta^*_{\text{end}}]$ in Equation (7.2)). **D.** The difference between the EDP obtained when the optimal sentinel-based window strategy is applied and when the optimal window strategy in the absence of sentinels is applied, for each (E, Δ) pair. Values are expressed as a percentage of the uniform seasonal baseline EDP (USB).

First, we considered the answer to question i) above: **Are sentinel plants still beneficial when surveillance resources are focused on a particular time of year?**

Including sentinel plants in the population in addition to employing a window sampling strategy was beneficial across a wide range of sampling efforts (E) and sample intervals (Δ), with the optimal number of sentinels to include in the population (P^*_S) substantially greater than zero for almost every pair of (E, Δ) values that we considered (Fig 7.9B). In most cases, including sentinel plants in the population in combination with window sampling led

to a further reduction in the EDP relative to the reduction that was already achieved with a window strategy in the absence of sentinels (Fig 7.9D). Exceptions to this occurred when the sample interval was $\Delta = 60$ days (blue line in Figs 7.9B, 7.9D) and the sampling effort E was three or more. Since sampling occurred infrequently in those cases, very large sample sizes were needed in order to focus the high sampling effort into the window of the year when detection was most likely (Supplementary Fig D.4A). When the sample size was very large, the EDP in the absence of sentinels was already very low, so sentinels did not have a beneficial effect (Supplementary Fig D.4B) (analogous to region 2 in Fig 6.8 in Chapter 6). Therefore, in those cases the optimal number of sentinels dropped very close to zero (Fig 7.9B) and correspondingly negligible changes in the EDP compared to a window strategy without sentinels were observed (Fig 7.9D).

In general, the marginal benefits of including sentinel plants in the population in addition to utilising the optimal window sampling strategy were less when the sampling effort was high (Fig 7.9D). For large values of E, the existing reductions in the EDP achieved using the optimal window strategy alone were greater (Fig 7.7B) and the scope for sentinel plants to facilitate further reductions was reduced.

For almost all of the (E, Δ) values considered, the optimal number of sentinels to include in the sample (N_S^*) was the maximum possible (i.e. $N_S^* = \min(P_S^*, N)$) (Fig 7.9C). As observed in Chapter 6, N_S^* only fell below $\min(P_S, N)$ when the sampling effort was very high and the sample size was large relative to the number of sentinels in the population. In those cases, always sampling the maximum possible number of sentinels would lead to repeated sampling of the same plants, resulting in less information gained per sample (particularly when the sample interval was short).

Second, we considered the answer to question ii) above: **Is the optimal choice of sampling window altered when sentinels are present in the population?**

For all values of (E, Δ) for which including sentinels in the population was beneficial (i.e. all values excepting $\Delta = 60$ days, $E \geqslant 3$), the optimal sampling window was wider than when sentinel plants were absent.

When the sample interval was $\Delta = 7, 15$, or 30 days, the optimal sampling window $[\Delta^*_{\text{start}}, \Delta^*_{\text{end}}]$ once again began in late April/early May. For all values of the sampling effort E, Δ^*_{start} lay between day 110 (21$^{\text{st}}$ April) and day 130 (11$^{\text{th}}$ May) (Fig 7.9A), exactly as we observed in the case where sentinels were not present (Fig 7.7A). However, the optimal sampling window $[\Delta^*_{\text{start}}, \Delta^*_{\text{end}}]$ ended in late September/mid October, with Δ^*_{end} lying between day 270 (28$^{\text{th}}$ September) and day 290 (18$^{\text{th}}$ October) for all values of E (Fig 7.9A).

This end date is approximately one month later than in the case where sentinels were not present, for which the optimal sampling window ended in late August/mid September (Fig 7.7A). This extension of the optimal sampling window into mid autumn reflects the fact that in our model sentinel plants develop symptoms after an average of 49 days. Therefore, sentinel plants infected at the peak of transmission during July and August may already begin to display visible symptoms of infection in September and October (unlike crop plants, which are highly unlikely to become Detectable until the following spring/summer). Sampling later into the year may therefore be rewarded by detecting recently infected sentinel plants.

When the sample interval was $\Delta = 60$ days and the sampling effort E was less than three, the optimal sampling window $[\Delta^*_{\text{start}}, \Delta^*_{\text{end}}]$ widened further, with Δ^*_{start} lying between day 80 (22^{nd} March) and day 100 (11^{th} April), and Δ^*_{end} lying between day 280 (8^{th} October) and day 290 (18^{th} October) for all values of E (Fig 7.9A). These results can be explained by considering that for a given sampling effort E, increasing the sample interval Δ requires a corresponding increase in the sample size N and/or the width of the sampling window $[\Delta_{\text{start}}, \Delta_{\text{end}}]$ to ensure that Equation (7.2) is satisfied. Although in the absence of sentinels, fixing the sampling window and increasing the sample size N within that window was the optimal approach (Fig 7.7A), this is not necessarily the case when sentinels are present. Since the number of sentinels in the population is limited by the need to balance their advantages and their drawbacks, large increases in N due to a very large sample interval may lead to a sample size in excess of the number of sentinels available. This does not make the best use of available sampling resources when the sample interval is large, since in that case sentinel plants can be repeatedly resampled without any loss of information (see Supplementary Text C.1) and it is preferable to sample entirely from the sentinel population if possible. By increasing the width of the sampling window, slightly smaller samples that consist entirely of sentinels can be taken.

When $\Delta = 60$ days and the sampling effort was three or more, including sentinels in the population was no longer beneficial (Fig 7.9B) and the optimal sampling window returned to the date range computed in the absence of sentinels in Section 7.3.2 (Fig 7.9A).

7.4 Discussion

Transmission of plant pathogens and the expression of the associated symptoms of infection are affected by seasonally varying factors, including vector population sizes, temperature, and precipitation. As a result, pathogen and disease dynamics change during the year, with implications for plant disease management strategies. In this chapter, we have extended the non-seasonal model constructed in Chapter 6 to account for seasonality in the rate of pathogen transmission, the rate of symptom development, and the sensitivity of visual

inspection as a means of detection, and explored how limited surveillance resources may be allocated optimally in this context.

One key objective of this chapter was to determine whether sentinel plants are still beneficial when seasonal factors are incorporated into the model. Overall, our results indicate that when seasonal factors are accounted for, sentinel plants retain the potential to achieve the substantial reductions in the EDP that we observed in Chapter 6 (Fig 7.5). This conclusion remains unchanged when 'window' sampling strategies are considered, in which sampling resources are focused into the window of the year when detection is most likely (Fig 7.9). These results therefore provide further support for the key conclusion reached in Chapter 6: that sentinel plants can be a useful tool to facilitate the early detection of invasive plant pathogens.

A second key objective of this chapter was to identify the optimal window of the year in which to focus limited sampling resources. Our results demonstrate that in the absence of sentinel plants, the smallest EDPs are achieved when the available sampling resources are concentrated between late April/early May and late August/mid September (Fig 7.7A). This provides quantitative support for the current EU guidelines regarding the timing of visual surveys for *X. fastidiosa*, which recommend that inspection should take place between late spring and early autumn when symptoms are more strongly expressed [341]. To the best of our knowledge, this is the first study to attempt a quantitative assessment of the optimal time of year for visual inspection of *X. fastidiosa*. These results therefore provide valuable evidence in support of current recommendations, as well as lending additional clarity to the interpretation of 'late spring' and 'early autumn' within those guidelines.

We found that, when sentinel plants are included in the population, it is beneficial to extend the sampling window later into the year (compared to when sentinel plants are not deployed), continuing surveillance until late September/mid October (Fig 7.9A). Widening the sampling window allows a grower to take full advantage of the available sentinel plants, particularly if sampling can only be performed infrequently due to logistical constraints. Of course, the quantitative results shown in this chapter are influenced by our choice of *C. roseus* as a candidate sentinel plant, and may vary if an alternative sentinel species is selected. However, this serves to emphasise the key qualitative message of our results: that the existing guidelines regarding the timing of visual inspection for *X. fastidiosa* should be carefully reviewed when surveillance strategies that make use of sentinel plants are implemented, to ensure that the available sampling resources are deployed in the most efficient way. As discussed in Chapter 6, the use of sentinel plants would also need to be considered in a broader economic context before sentinel-based strategies can be deployed in practice.

In this chapter, our aim was to investigate the effects of seasonality on surveillance pro-

grammes that rely upon visual inspection of potential infected hosts, with a particular focus on the use of sentinel plants to facilitate earlier detection. However, our seasonal model provides a framework that could be adapted to assess the impacts of seasonality on alternative surveillance methods and to explore the optimal sampling window for their deployment. For example, in addition to performing visual inspection for symptoms of *X. fastidiosa* infection, growers are also advised to select samples of symptomatic and nonsymptomatic plant material to undergo laboratory testing [341]. However, the ability of a molecular or serological test to correctly identify the presence of the pathogen in infected plant material depends on the bacterial titre within the sample [317, 345]. Recently infected presymptomatic hosts will not be correctly identified until the bacterial concentration exceeds the threshold sensitivity of the test being used. As described in the introduction to this chapter, the growth rate of *X. fastidiosa* is highly temperature dependent, with the bacteria unable to multiply below around 17°C [335]. This suggests that molecular testing of presymptomatic hosts is unlikely to be as effective during the winter months. Indeed, for plants that are not displaying visible symptoms, current guidelines recommend that sampling for laboratory testing should take place between late summer and early autumn, when warmer temperatures lead to an increase in the bacterial growth rate [341]. The high cost of performing large scale laboratory testing means that targeting sampling at the time of year when detection is most likely is particularly desirable. Our model could be adapted to consider the effects of sampling for laboratory testing by introducing an additional stage between the Undetectable and Detectable compartments, in which infected hosts do not display visible symptoms but are identifiable when a laboratory test is applied. Analogous methodology to the approach used in this chapter could then be applied to investigate the optimal window of the year in which to collect samples for laboratory testing, and establish whether the current guidelines are supported.

As discussed in Section 6.4, an important limitation of our model is that we did not model vector population dynamics explicitly. Instead, we modelled transmission as occurring directly from infected (Undetectable or Detectable) plants to susceptible (Healthy) plants, and assumed that vector behaviour was captured implicitly through the choice of transmission rate parameters (β_C and β_S). In this chapter, we additionally assumed that the seasonal variation in transmission rate that arises due to fluctuations in the vector population could be considered proportional to the average density of adult vectors at any given time (see Section 7.2.1 and Fig 7.1A). The vector density therefore acts as a seasonal scaling factor for the transmission rate parameters, altering the likelihood of transmission between infected and susceptible plants. In reality, however, transmission does not occur directly between plants, but rather from infected vectors to susceptible plants. Our model involves an implicit assumption that the prevalence of infection in the vector population at any given time is equal to the prevalence of infection in the plant population, which is not necessarily the

case [263]. Including vectors in the model explicitly and modelling transmission directly between infected vectors and susceptible plants would allow us to make use of observed data on the seasonal prevalence of infection within the vector population [263, 332, 342], which may alter the time of year when the transmission rate is at its highest and consequently shift the optimal window for sampling. However, although constructing a full host-vector model would provide a more accurate representation of transmission dynamics within the model system, this would come at the cost of substantially increased model complexity.

In this chapter, we made the standard modelling assumption that the time that each plant spends in each compartment is drawn from an exponential distribution. Although this is a common simplification in compartmental modelling, it results in a high probability that the time spent in a given compartment is in fact substantially less than the mean value. In the context of this chapter, this means that there is a high likelihood that an Undetectable plant will develop visible symptoms much sooner than the specified average amount of time (γ_C or γ_S for crops and sentinels, respectively), influencing the time of year that they are most likely to become Detectable. This may have a knock-on effect on the timing of the optimal sampling window. To test the impact of this simplification on our results, our model could be extended to include a series of n identical sequential Undetectable compartments which are each occupied for an exponentially distributed amount of time. By setting the mean time spent in each of these compartments to be γ_C/n (or equivalently γ_S/n for sentinels), the resulting total time spent Undetectable would be gamma distributed with the same overall mean γ_C (or γ_S for sentinels). The total duration of the Undetectable period would then be more likely to take values close to the mean, reducing the likelihood of very rapid progression to the Detectable compartment. If desired, we could also use this approach to consider a scenario in which the detection sensitivity increases between successive model compartments, representing a more gradual development of visible symptoms than is currently accounted for in our model (see, for example, Hart *et al.* [211]).

In summary, the results of this chapter provide further support for using sentinel plants to facilitate early detection of invasive plant pathogens when the rates of transmission, symptom development, and symptom expression follow seasonal patterns. As we have demonstrated, sentinel plants are still effective at reducing the expected incidence of disease upon pathogen discovery when seasonal factors are accounted for and when sampling resources are correspondingly focused on the window of the year when detection is most likely to occur. Using sentinel plants in combination with temporally focused surveillance has the potential to substantially reduce the negative impacts of invading plant pathogens that exhibit strong seasonal characteristics.

Appendix D

Supplementary material for Chapter 7

D.1 Construction of the seasonal scaling factor for the rate of symptom development, $\psi(t)$

In the seasonal model of pathogen transmission used in Chapter 7, we incorporated a time-dependent rate of symptom development by multiplying the constant rates of symptom development used in Chapter 6 by a seasonally varying factor $\psi(t)$. As described in Section 7.2.1, $\psi(t)$ represents the seasonal variability in environmental suitability for symptom development, and is constructed to reflect the temperature-dependent rate at which *X. fastidiosa* multiplies. Here, we provide the values on which the construction of $\psi(t)$ is based. We began by considering the average monthly temperature in Lecce [336], values of which are shown in the second column of Supplementary Table D.1. We replaced any of these values that were less than 17°C with zero, to reflect the fact that *X. fastidiosa* cannot multiply below that threshold [335]. These adjusted values are shown in the third column of Supplementary Table D.1. For consistency with the non-seasonal model, we then scaled the twelve resulting monthly values to have a mean of one, as shown in the fourth column of Supplementary Table D.1. Finally, we performed linear interpolation on these monthly values to obtain daily values for $\psi(t)$ (Fig 7.1B).

D.2 Construction of the seasonal detection sensitivity, $\chi(t)$

As described in Section 7.2.2, in the seasonal model used in Chapter 7 we allowed the sensitivity of visual inspection for Detectable plants to vary throughout the year. Specifically, a Detectable plant sampled at time t is correctly identified as infected with probability $\chi(t)$, which represents the degree of water stress experienced by the plants at time t. Here, we provide further details on how the function $\chi(t)$ is constructed.

We began by defining the vector $\mathbf{P} = [P_{\text{Jan}}, P_{\text{Feb}}, \cdots, P_{\text{Dec}}] \in \mathbb{R}_{1\times 12}$, whose values represent

Month	Average monthly temperature (°C)	Adjusted temperature values (°C)	Monthly values of $\psi(t)$
Jan	9	0	0
Feb	9	0	0
Mar	11	0	0
Apr	14	0	0
May	19	19	1.7
Jun	23	23	2.1
Jul	26	26	2.3
Aug	26	26	2.3
Sep	22	22	2.0
Oct	18	18	1.6
Nov	13	0	0
Dec	10	0	0

Table D.1: The values used to construct the seasonal scaling factor for the rate of symptom development, $\psi(t)$. The values in the second column represent the average monthly temperature (°C) in Lecce [336]. Values that are below 17°C are replaced by zero to reflect the fact that the bacteria cannot multiply below that threshold [335], as shown in the third column. The resulting monthly values are then scaled to have a mean of one for consistency with the non-seasonal model used in Chapter 6, as shown in the fourth column. Finally, daily values of $\psi(t)$ are obtained by performing linear interpolation on these scaled monthly values.

the average monthly rainfall (mm) in Lecce [336] (values shown in the second column of Supplementary Table D.2). Then, we scaled these values to lie between zero and one according to the following equation:

$$\hat{\mathbf{P}} = [\hat{P}_{\text{Jan}}, \hat{P}_{\text{Feb}}, \cdots, \hat{P}_{\text{Dec}}] := \frac{\mathbf{P} - \min \mathbf{P}}{\max \mathbf{P}}. \tag{D.1}$$

Values of the vector $\hat{\mathbf{P}}$ are shown in the third column of Supplementary Table D.2. To calculate the detection sensitivity for a sample taken at time t, we note that t represents the time elapsed in days since the simulation was initialised on 1$^{\text{st}}$ January, and therefore determine the corresponding month of the year, m_t, as follows:

$$m_t = \left\lceil \text{mod}(t, 365) \times \frac{12}{365} \right\rceil. \tag{D.2}$$

Here, $m_t = 1$ corresponds to time t falling in January, $m_t = 2$ corresponds to time t falling

in February, and so on. Then, the sensitivity of detection at time t is given by

$$\chi(t) = 1 - \hat{\mathbf{P}}_{m_t} \tag{D.3}$$

(where $\hat{\mathbf{P}}_{m_t}$ represents the ${m_t}^{th}$ entry of the vector $\hat{\mathbf{P}}$). Values of $\chi(t)$ are shown in the fourth column of Supplementary Table D.2.

Month	**Average monthly rainfall (mm)**	**Scaled values**	**Seasonal detection sensitivity** $\chi(t)$
Jan	46.7	0.61	0.39
Feb	49.0	0.64	0.36
Mar	45.3	0.58	0.42
Apr	33.9	0.39	0.61
May	22.0	0.19	0.81
Jun	13.1	0.04	0.96
Jul	10.8	0.00	1.00
Aug	16.5	0.10	0.90
Sep	42.7	0.54	0.46
Oct	61.6	0.86	0.14
Nov	70.1	1.00	0.00
Dec	58.0	0.80	0.20

Table D.2: The values used to construct the seasonal detection sensitivity, $\chi(t)$. The values in the second column represent the average monthly rainfall (mm) in Lecce [336], which form the entries of the vector $\mathbf{P}$. These values are then scaled to lie between zero and one to form the entries of the vector $\hat{\mathbf{P}}$, as shown in the third column. Monthly values of $\chi(t)$ are then obtained by subtracting these from one, shown in the fourth column.

D.3 Sensitivity to changes in the seasonal scaling factors $\omega(t)$, $\psi(t)$, and $\chi(t)$

Throughout the analyses conducted in Chapter 7, we used the values for the scaling factors $\omega(t)$, $\psi(t)$, and $\chi(t)$ that were described in Sections 7.2.1 and 7.2.2, and Supplementary Texts D.1 and D.2. That is, we assumed that the seasonal risk of pathogen introduction and rate of pathogen transmission could be considered proportional to the size of the vector population at any given time, that the rate at which symptoms develop in infected plants was governed by the ambient temperature, and that the dryness of the soil (as constructed from seasonal precipitation data) could be used as a proxy for the degree of water stress experienced by

the plants, which in turn determined the strength of symptom expression and corresponding chance of detection. However, in practice these assumptions may not be entirely accurate. For example, as discussed in Section 7.4, our choice of $\omega(t)$ involves an implicit assumption that the prevalence of infection in the vector population at any given time is equal to the prevalence of infection in the plant population, which is not necessarily the case. Similarly, soil moisture levels may not correspond directly to the amount of precipitation at any given time; there may be a delay effect, such that the degree of water stress experienced by the plants is disproportionately greater after a long period of little precipitation.

Changes to these seasonal factors could influence both the optimal number of sentinels to include in the population and the optimal window of the year in which to focus surveillance resources. Therefore, we conducted a preliminary analysis to investigate how variation in these factors from their assumed values used in Chapter 7 could affect the results described in Section 7.3. For each of the three seasonal factors considered – the seasonal transmission rate scaling factor, $\omega(t)$, the seasonal scaling factor for the rate at which symptoms develop, $\psi(t)$, and the sensitivity of visual inspection as a means of detection, $\chi(t)$ – we explored the effects of removing seasonality in that factor entirely, whilst the other two factors remained unchanged. In doing so, we aimed to assess the degree to which our results depended on each choice of seasonal factor, and to identify whether or not any one of these choices was solely responsible for driving our conclusions. Thus, we considered the following three scenarios:

- The seasonal transmission rate scaling factor, $\omega(t)$, was set to be $\omega(t) = \text{constant} = 1$ (with a correspondingly uniform risk of pathogen introduction throughout the year), whilst $\psi(t)$ and $\chi(t)$ remained as in the main text;
- The seasonal scaling factor for the rate at which symptoms develop, $\psi(t)$, was set to be $\psi(t) = \text{constant} = 1$, whilst $\omega(t)$ and $\chi(t)$ remained as in the main text;
- The sensitivity of visual inspection as a means of detection, $\chi(t)$, was set to be $\chi(t) = \text{constant} = 0.5221$ (the average of the monthly values used in the main text), whilst $\omega(t)$ and $\psi(t)$ remained as in the main text.

For each of these scenarios, we repeated the analyses described in Sections 7.3.1, 7.3.2, and 7.3.3. We produced plots analogous to those displayed in Figs 7.5B, 7.7A, 7.9A, and 7.9B, respectively showing i) the optimal total number of sentinels to include in the population under a temporally uniform sampling strategy; ii) the optimal window of the year in which to perform sampling in the absence of sentinel plants; iii) the optimal window of the year in which to perform sampling when sentinel plants are present in the population, and; iv) the corresponding optimal number of sentinels to include in the population when that window-based strategy is used. These results are shown in Supplementary Figs D.1 (for the case in which $\omega(t) = \text{constant}$), D.2 (for the case in which $\psi(t) = \text{constant}$), and D.3 (for

the case in which $\chi(t) = \text{constant}$).

In each scenario that we considered, our qualitative conclusions remained unchanged. When a uniform sampling strategy was used, including sentinel plants in the population was beneficial across most of the range of (N, Δ) values considered (Panel A in each of Supplementary Figs D.1, D.2, and D.3). As expected, the optimal total number of sentinels to include varied with the choice of N and Δ, in the same pattern as previously observed. When a window sampling strategy was used instead, including sentinel plants in the population remained beneficial for a range of choices of the sampling effort E and the sample interval Δ (panel D in each of Supplementary Figs D.1, D.2, and D.3). As observed in Fig 7.9B, sentinels were typically more useful when the sampling effort was low and/or sampling was performed frequently, with optimal sentinel numbers dropping to near zero when the sampling effort was high and/or the sample interval was large. Exactly as before, the optimal window of the year in which to perform sampling widened when sentinels were present in the population, compared to the optimal sampling window for the same choice of E and Δ in the absence of sentinels (panels B and C in each of Supplementary Figs D.1, D.2, and D.3).

Of course, the precise quantitative results that we observed varied between the three scenarios considered. Further research into the effects of seasonally varying environmental factors on pathogen transmission and symptom development is therefore needed to ensure that limited surveillance resources are deployed throughout the year in the most efficient way. However, despite these quantitative variations in the optimal choice of surveillance strategy, our results clearly indicate that none of the three individual seasonal factors that we considered were solely responsible for driving the results that we observed in Chapter 7. This suggests that our general conclusions are likely to remain relevant, even if our choices of $\omega(t)$, $\psi(t)$, and $\chi(t)$ are not entirely accurate.

D.4 Supplementary Figures

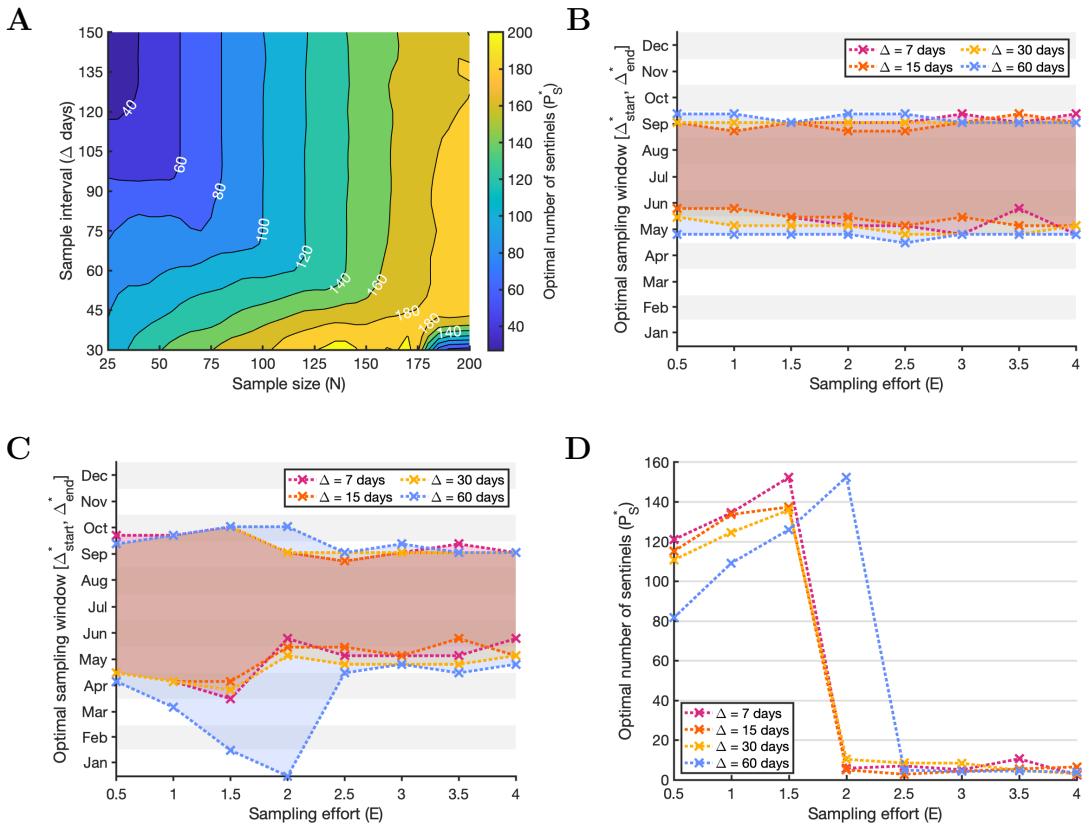


Figure D.1: The effect of discounting seasonality in the vector population on the key results obtained from the seasonal model in Chapter 7. Panels are analogous to Figs 7.5B, 7.7A, 7.9A, and 7.9B in the main text, but in the case where the seasonal transmission rate scaling factor is set to be $\omega(t) =$ constant $= 1$ (and the risk of pathogen introduction throughout the year is correspondingly uniform). A. The optimal number of sentinel plants to include in the population (P_S^*) when a uniform sampling strategy is applied, as the sample size N and sample interval Δ vary. **B.** The optimal sampling window ($[\Delta^*_{\text{start}}, \Delta^*_{\text{end}}]$) when a window sampling strategy without any sentinel plants is applied, as the total annual sampling effort E and the sample interval Δ vary. **C.** The optimal sampling window ($[\Delta^*_{\text{start}}, \Delta^*_{\text{end}}]$), and **D.** The optimal number of sentinels to include in the population (P_S^*), when a sentinel-based window sampling strategy is applied, as the total annual sampling effort E and the sample interval Δ vary. In panels B and C, values of Δ_{start}, Δ_{end} are translated into calendar dates and plotted on a monthly scale for ease of interpretation.

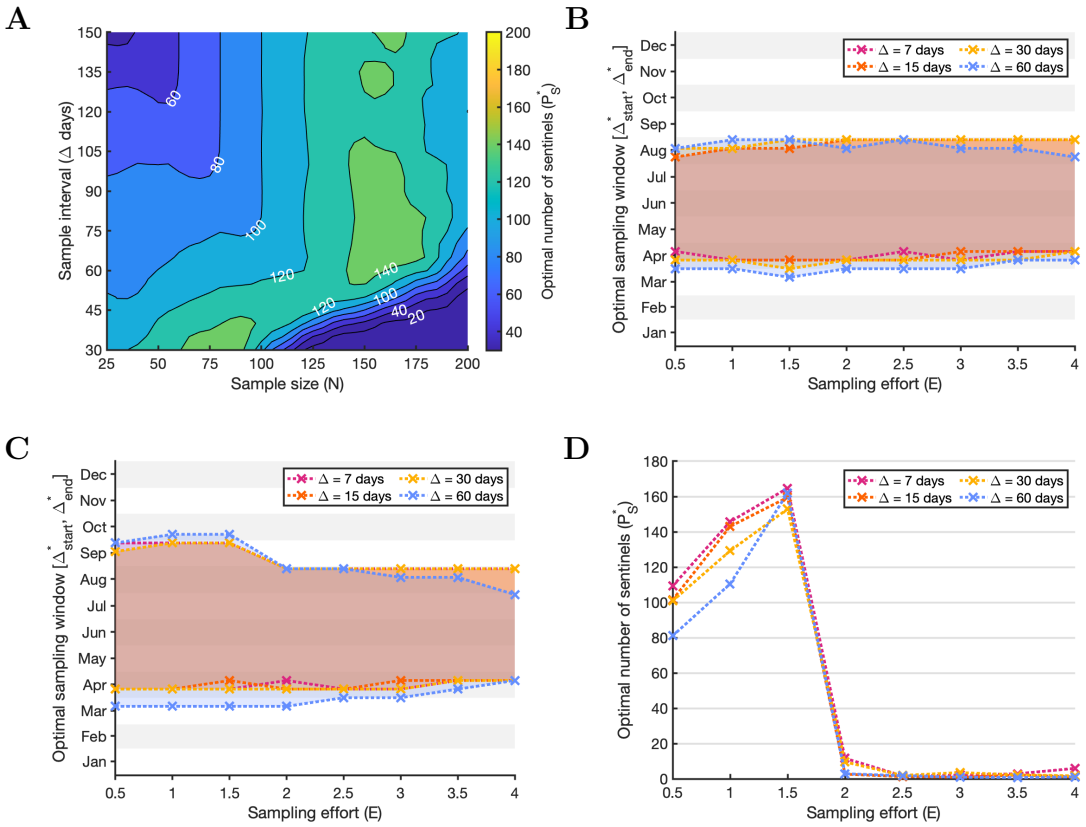


Figure D.2: The effect of discounting seasonal variations in temperature on the key results obtained from the seasonal model in Chapter 7. Panels are analogous to Figs 7.5B, 7.7A, 7.9A, and 7.9B in the main text, but in the case where the seasonal scaling factor for the rate at which symptoms develop is set to be $\psi(t) = \text{constant} = 1$. A. The optimal number of sentinel plants to include in the population (P_S^*) when a uniform sampling strategy is applied, as the sample size N and sample interval Δ vary. **B.** The optimal sampling window ($[\Delta^*_{\text{start}}, \Delta^*_{\text{end}}]$) when a window sampling strategy without any sentinel plants is applied, as the total annual sampling effort E and the sample interval Δ vary. **C.** The optimal sampling window ($[\Delta^*_{\text{start}}, \Delta^*_{\text{end}}]$), and **D.** The optimal number of sentinels to include in the population (P_S^*), when a sentinel-based window sampling strategy is applied, as the total annual sampling effort E and the sample interval Δ vary. In panels B and C, values of Δ_{start}, Δ_{end} are translated into calendar dates and plotted on a monthly scale for ease of interpretation.

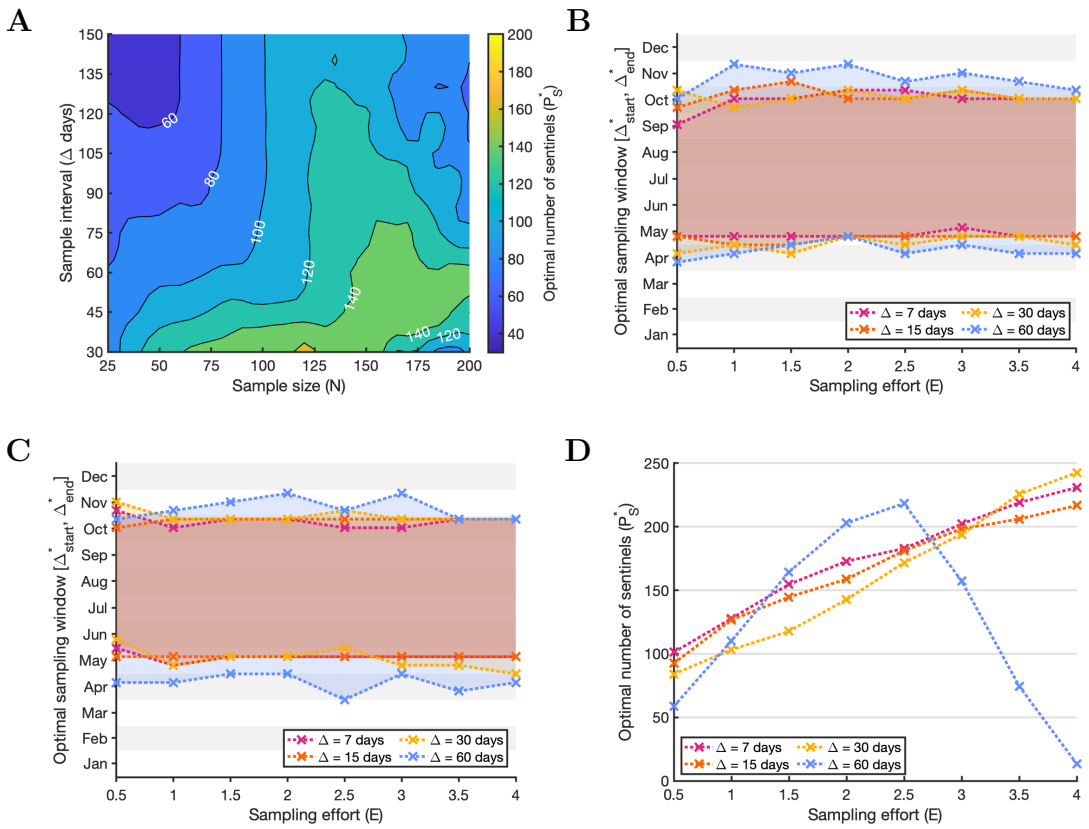


Figure D.3: The effect of discounting seasonal variations in precipitation on the key results obtained from the seasonal model in Chapter 7. Panels are analogous to Figs 7.5B, 7.7A, 7.9A, and 7.9B in the main text, but in the case where the sensitivity of visual inspection as a means of detection is set to be $\chi(t)$ = constant = 0.5221 (the average of the monthly values used in the main text). A. The optimal number of sentinel plants to include in the population (P_S^*) when a uniform sampling strategy is applied, as the sample size N and sample interval Δ vary. **B.** The optimal sampling window ($[\Delta^*_{\text{start}}, \Delta^*_{\text{end}}]$) when a window sampling strategy without any sentinel plants is applied, as the total annual sampling effort E and the sample interval Δ vary. **C.** The optimal sampling window ($[\Delta^*_{\text{start}}, \Delta^*_{\text{end}}]$), and **D.** The optimal number of sentinels to include in the population (P_S^*), when a sentinel-based window sampling strategy is applied, as the total annual sampling effort E and the sample interval Δ vary. In panels B and C, values of Δ_{start}, Δ_{end} are translated into calendar dates and plotted on a monthly scale for ease of interpretation.

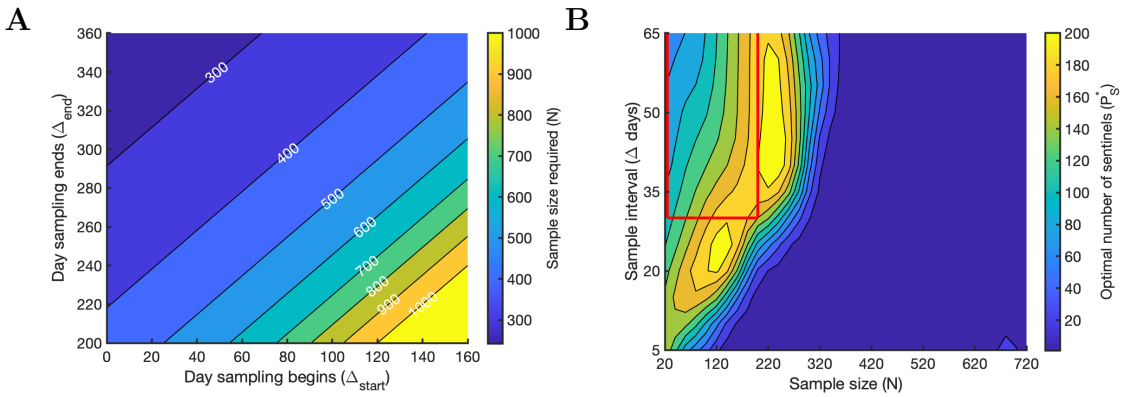


Figure D.4: A. The sample size N that is required in order to concentrate the annual sampling effort $E = 4$ into the sampling window $[\Delta_{\text{start}}, \Delta_{\text{end}}]$ when the sample interval is $\Delta = 60$ days, as Δ_{start} and Δ_{end} vary. For each pair of values $(\Delta_{\text{start}}, \Delta_{\text{end}})$ that we considered, the necessary sample size N was calculated according to Equation (7.2) (unless by that calculation N exceeded the total population size $P = 1,000$, in which case $N = P$). Since the sampling effort is high and samples are taken infrequently, very large sample sizes are required when sampling is restricted to a narrow window of the year (as in the bottom right hand side of the figure). **B.** The optimal number of sentinel plants to include in the population (P_S^*) when a uniform sampling strategy is applied, as the sample size varies between $N = 20$ and $N = 720$, and the sample interval varies between $\Delta = 5$ days and $\Delta = 65$ days. The results displayed in this panel are analogous to those shown in Fig 7.5B, but for a different range of N and Δ that correspond more closely to the values taken by these parameters in Sections 7.3.2 and 7.3.3. The red box outlines the region of this figure that overlaps with the results shown in Fig 7.5B. For very large sample sizes (corresponding to cases in which a large sampling effort E must be concentrated into a narrow sampling window) the baseline EDP is already low and therefore sentinels did not have a beneficial effect.

Part II: Summary and concluding remarks

Invasive plant pathogens represent a serious threat to global plant health, with potentially severe implications for food security, biodiversity, and people's livelihoods. *X. fastidiosa* is a key example of a dangerous plant pathogen that could negatively affect global agriculture on a large scale if effective control measures are not rapidly devised. For plant pathogens such as *X. fastidiosa* for which no treatment is available, early detection of the pathogen soon after it has arrived in a susceptible population is crucial if containment measures such as roguing are to be successful at preventing further spread. However, in many cases early detection monitoring typically relies on visual inspection of potential infected hosts for visible symptoms of infection, which can be substantially hindered if the pathogen has a long period of asymptomatic infection in the host plant. One potential strategy that could be used to combat the problem of asymptomatic infection is to consider the use of 'sentinel' plants – alternative plant species that are known hosts of the pathogen in question, but that have a much shorter incubation period than the main species of interest. If sentinel plants are intercropped in the population and included in visual surveillance programmes, they could act as early warning beacons for infection. However, although the use of sentinel plants to improve early detection monitoring has previously been suggested at the G20 Meetings of Agricultural Chief Scientists, a quantitative assessment of the potential of this approach has never previously been attempted.

In this thesis, we filled this knowledge gap by constructing and analysing a mathematical model of pathogen transmission to investigate the potential of sentinel plants to facilitate early detection of an invading pathogen. Using *X. fastidiosa* infection in *O. europaea* as a case study, with *C. roseus* as a candidate sentinel plant, we have investigated the effects of including sentinel plants in a monitoring programme on the expected prevalence of infection at the time of detection, and demonstrated how the choice of sentinel strategy can be optimised to facilitate the greatest reductions in this prevalence compared to a non-sentinel based strategy that uses an equivalent surveillance effort. Additionally, we have investigated how the optimal sentinel strategy is affected by seasonal factors that influence pathogen transmission and symptom expression, and identified the optimal window of the year in which to focus limited surveillance resources.

The results obtained in Chapters 6 and 7 of this thesis have highlighted the following key messages regarding the use of sentinel plants for early detection of *X. fastidiosa* infection in *O. europaea*.

- Sentinel plants have the potential to facilitate substantial reductions in the prevalence of infection in the population at the time of detection, compared to what would be achieved using a non-sentinel based surveillance strategy of equivalent sampling effort. Therefore, for the same total amount of surveillance resource, the number of plants that must be removed and the risk that the pathogen will be exported to a new location before control measures are implemented can be reduced.

- The optimal sentinel strategy (i.e. the number of sentinel plants to include in the population and in the sample) varies substantially with the total sample size and sample interval being used – there is no 'one size fits all' approach. Therefore, the sentinel strategy must be carefully selected on a case by case basis depending on the sampling resources available and any logistical constraints on the sample size or frequency of sampling. If the sentinel strategy is selected sub-optimally, smaller reductions – or even increases – in the detection prevalence may occur, due to the increased transmission rate induced by including sentinels in the population.

- Our general conclusions regarding the benefits of including sentinel plants in a monitoring programme remained unchanged when we accounted for seasonal factors that affect pathogen transmission and symptom expression in our model. That is, including sentinel plants in the population remained beneficial across the range of sample sizes and sample intervals considered, but as before the optimal sentinel strategy to deploy was strongly influenced by the sampling resources available.

- In the absence of sentinel plants, the optimal window of the year in which to perform surveillance was between late April/early May and late August/mid September. This result provides valuable quantitative support for the existing EU guidelines on the timing of visual inspection for *X. fastidiosa*, which recommend that surveillance should occur between late spring and early autumn. However, when sentinel plants were included in the population the optimal sampling window extended later into the year, typically lasting until late September/mid October. Additionally, when sampling only occurred infrequently, the optimal sampling window began earlier in the year than it did in the absence of sentinels. This suggests that the precise timing of surveillance for a given pathogen should be carefully evaluated for specific crop-sentinel systems to ensure that the interacting effects of sentinel strategy and sampling window are accounted for.

The work contained within Part II of this thesis represents the first quantitative investigation

into the use of sentinel plants to facilitate early detection of invasive plant pathogens, and as such provides a valuable contribution to the existing literature on early detection monitoring for preventive control. Our results suggest that sentinel-based surveillance strategies could lead to substantial reductions in the prevalence of infection at the time of detection compared to non-sentinel based strategies that require an equivalent surveillance effort, providing support for incorporating sentinel plants into disease monitoring programmes. Although we have used *X. fastidiosa* infection in *O. europaea* as a case study, sentinel plants may also have the potential to be beneficial in other pathosystems for which early detection is crucial for control, but relies on visual inspection for symptomatic hosts. Of course, further work is needed to explore the practicalities of deploying sentinel plants in a real-world pathosystem, and to evaluate the potential economic benefits and costs of sentinel-based surveillance strategies. However, we hope that the work presented in this thesis will provide motivation for future researchers to continue investigating the potential for sentinel plants to assist in the ongoing global fight against invasive plant pathogens.

Chapter 8

Conclusion

Infectious disease outbreaks in human, animal, and plant populations are responsible for severe economic, environmental, and societal consequences [1–6, 52, 224, 225, 228]. Developing effective methods for containing the spread of infectious pathogens is crucial if these negative effects are to be minimised. The emergence of novel pathogens and the introduction of non-native pathogens to previously uninfected locations represent particular challenges for global disease control efforts, since such pathogens can appear unexpectedly in susceptible populations and spread widely before being identified [1, 3, 6, 224]. Additionally, novel or non-native pathogens have the potential to cause severe disease in infected hosts due to immunological naievety in the population (for humans and other animals) or a lack of co-evolved resistance (for plants) [4, 52, 346]. Therefore, combating the spread of infectious pathogens that have appeared for the first time, reappeared after a period of decline, or are transported to new locations continues to be a key focus of epidemiological research [347].

A pathogen may be introduced to a susceptible population in a variety of ways. Approximately 60% of novel emerging pathogens that have been observed in humans have been introduced as a result of zoonotic spillover events, including HIV, Ebola viruses, and the coronaviruses MERS-CoV, SARS-CoV, and SARS-CoV-2 [3, 56–60, 348]. As well as being the dominant source of previously unseen pathogens, zoonotic spillover may also lead to the re-emergence of known pathogens that have previously caused one or more disease outbreaks in humans. For example, since the Ebola virus was first identified in Zaire (now the Democratic Republic of the Congo) in 1976, multiple outbreaks of Ebola in Africa have been linked to separate spillover events [58]. In most cases, pathogens that enter the human population as a result of zoonotic spillover are not well adapted for human-to-human transmission, particularly if the reservoir host is phylogenetically distant from humans [349, 350]. However, the minority of zoonotic pathogens that are able to adapt to human-to-human transmission have the potential to trigger major outbreaks of disease [346]. Indeed, every viral pandemic that has occurred within the last century can be attributed to a pathogen of zoonotic origin, and the risk of spillover events continues to increase as human activities

encroach further into the natural landscape [11].

A novel pathogen may alternatively arise in a population when a known pathogen that is already circulating in that population mutates to form a new variant. If the mutated pathogen is well adapted for infection, and its transmission is not substantially hindered by existing immunity in the population (e.g. if the mutation allows the pathogen to evade existing immunity, or the previously infected proportion of the population is small), the new variant may have the potential to spread widely. For example, in May-June 2005, a small outbreak of Chikungunya virus infections occurred on La Réunion Island in the Indian Ocean, which led to around $6,000$ reported cases in total [351]. However, at the end of that year a new strain of the virus emerged that differed from the first by two point mutations, which improved the ability of the virus to infect the mosquito vector [352]. The new variant was therefore more strongly adapted for infection than the previous strain and led to a much larger outbreak on the island in January-February 2006, with over $200,000$ cases reported [351]. Influenza viruses are also well known for their ability to mutate and recombine to form new strains, resulting in the frequent emergence of novel variants that are able to evade existing population immunity [353].

The introduction of a novel pathogen to a population also commonly occurs when a known pathogen that has already emerged in one region is transported to a new location where it was not previously present. For example, at the beginning of the COVID-19 pandemic cases were limited to China, but the pathogen was soon introduced to many other countries worldwide as a result of international travel [13, 50]. Similarly, the migration of wild birds is a key route through which avian influenza viruses are introduced to new locations [354, 355], whilst plant pathogens commonly invade susceptible populations following the importation and planting of infected nursery stock [51–54].

Regardless of the means of introduction, when an infectious pathogen arrives in a susceptible population it may lead to a major disease outbreak with potentially severe consequences. However, if a pathogen can be identified soon after it has arrived in the population, and the outbreak can be controlled whilst it is still in its very early stages, many of the potential costs of the outbreak can be averted [31, 178, 264, 356]. For example, following the 2001 outbreak of foot-and-mouth disease in the United Kingdom, Ferguson *et al.* [31] demonstrated that if culling of animals on infected premises had been achieved within 24 hours of case detection from the start of the epidemic, the total number of cases observed and the number of farms culled would have been reduced by over 40%. For this reason, designing effective strategies for the surveillance of invasive pathogens, and intervention packages that aim to interrupt transmission occurring from the early infected cases, is a key focus for policy makers.

Unfortunately, achieving early detection of invading pathogens and subsequently interrupting their transmission before a large number of infections have occurred can be extremely challenging. Surveillance programmes that aim to detect a pathogen soon after it has arrived in the population are often hindered by periods of asymptomatic infection, making the rapid identification of infected individuals and their subsequent isolation, culling, or roguing (for humans, animals, and plants, respectively) difficult [12, 51, 119, 251, 270]. Additionally, the scale, intensity, and effectiveness of surveillance programmes are typically limited by economic and logistical constraints. For example, as discussed in Part II of this thesis, laboratory testing of potential hosts for *Xylella fastidiosa* infection is too expensive to perform on a large scale, meaning that monitoring programmes instead typically rely on visual inspection as a means of detection, which is far less sensitive [263, 280]. Expense and practical constraints also hinder the rapid deployment of control measures such as vaccination programmes, which require an efficient supply chain and a coordinated large scale delivery plan [357].

The rapid deployment of control interventions early in an outbreak may also be hindered by any potential negative consequences associated with the measures being considered. For example, the 'lockdown' strategies that were widely used during the COVID-19 pandemic resulted in severe economic repurcussions, as well as negatively affecting people's education and their non-disease health [185, 187, 188, 358]. Interventions that rely on culling of infected animals or roguing of infected plants will also result in economic losses for the farmer or grower, meaning that intensive large-scale culling/roguing strategies must be carefully evaluated before being deployed [20, 34]. In such cases, a policy maker must determine not only whether a given intervention package is sufficient to achieve pathogen control, but also whether it is necessary and its use justified, ensuring that the desired goal is achieved without measures being unnecessarily stringent, or valuable resources being wasted.

Mathematical modelling is a critical component of planning for disease management, particularly when resources are limited and the advantages and disadvantages of control strategies must be evaluated. Modelling allows a variety of complex factors that influence pathogen transmission and control to be integrated into an analytical framework, which can be used to project the course of the outbreak and to predict the effects of intervention strategies.

8.1 Key results of this thesis and areas for future research

In this thesis, we have developed mathematical modelling techniques to explore the optimal targeting of surveillance and control interventions to achieve early outbreak detection and containment. In particular, we have considered the following two key questions that arise for policy makers at the very beginning of an infectious disease outbreak:

1. Will the initial case(s) that arrive in the population lead to a major disease outbreak driven by sustained local transmission, and how can control interventions be optimally deployed to reduce the risk of this occuring?
2. If pathogen transmission is sustained in the population, how many individuals will be infected before the outbreak is first identified, and how can we design effective surveillance programmes that facilitate detection at low infection prevalence?

8.1.1 Summary of Part I

In Part I of this thesis, we addressed the first of these key questions, using SARS-CoV-2 as a case study.

If an infectious disease outbreak is to be contained whilst still in its early stages, it is essential to ensure that the initial cases in the population do not go on to generate chains of sustained transmission. To prevent a large number of individuals becoming infected, a policy maker may choose to deploy control interventions that aim to reduce the number of secondary infections an infected host generates. A key metric that can be used at the beginning of an outbreak to assess the potential effects of such interventions is the probability that an initial infected host triggers a major outbreak – that is, an outbreak in which sustained transmission occurs and a large number of individuals become infected, as opposed to an outbreak in which only a few individuals are ever infected and the pathogen dies out quickly (Section 2.1) [40]. As discussed in Chapter 2, this probability can be estimated by performing repeated stochastic simulations of a pathosystem, and observing the proportion of those simulations in which a major outbreak occurs (Section 2.2). However, an alternative approach for calculating this probability without needing large numbers of model simulations is to use a method known as the 'first-step analysis' technique, which has previously been used to assess outbreak risks in a variety of settings (Section 2.3) [40–42, 64, 67, 77, 78]. In Part I of this thesis, we applied this technique to compartmental models of SARS-CoV-2 transmission to investigate the effectiveness of early outbreak control interventions for COVID-19.

Our initial focus was on understanding the relative contribution to transmission of symptomatic and nonsymptomatic individuals, where 'nonsymptomatic' is used as an umbrella term to describe both 'presymptomatic' individuals (infected hosts who have not yet developed symptoms, but will go on to do so) and 'asymptomatic' individuals (infected hosts who do not develop symptoms at any point during their infection). In particular, we sought to understand whether the rapid isolation of symptomatic hosts or the testing and isolation of nonsymptomatic hosts would have a greater effect on the overall probability that a major outbreak occurs, with the aim of determining whether surveillance should be targeted at symptomatic hosts, nonsymptomatic hosts, or a combination of the two. This was a particularly relevant question during the COVID-19 pandemic, since both presymptomatic

and asymptomatic transmission were shown to occur and presented a substantial challenge for disease control efforts – yet testing of nonsymptomatic hosts was not initially prioritised (see Section 3.1) [93–105, 128–130]. In Chapter 3 of this thesis, we constructed a compartmental model of SARS-CoV-2 transmission that accounted for both presymptomatic and asymptomatic transmission explicitly, and used the first-step analysis technique to calculate the probability that a single infected host at the beginning of their infection triggers a major outbreak (Section 3.2). Using this analytic estimate, we explored how the outbreak probability would be affected by increasing the rate at which symptomatic and nonsymptomatic hosts were isolated, corresponding to increasing surveillance targeted at either group (Section 3.3.1). Then, we considered how a limited amount of surveillance resource could be optimally divided between symptomatic and nonsymptomatic hosts to maximally reduce the outbreak probability, taking into account the fact that successfully identifying and isolating nonsymptomatic infected hosts is more challenging than finding symptomatic hosts (Section 3.3.2). Our results in this chapter clearly demonstrate that surveillance for and isolation of nonsymptomatic infected hosts can play a crucial role in reducing the risk of an outbreak occurring, even when increasing the isolation rate of nonsymptomatic hosts requires more surveillance resources than the equivalent increase in isolation rate for symptomatic hosts (Fig 3.3). Targeting and isolating infected hosts who are not showing symptoms not only results in reduced transmission from those individuals, but also indirectly reduces transmission from symptomatic hosts, since many of those individuals will be isolated before developing symptoms. Overall, mixed strategies in which surveillance was targeted at both symptomatic and nonsymptomatic individuals proved the most successful at mitigating the outbreak risk (Fig 3.4). Our results indicate that rapidly assessing the potential contribution of nonsymptomatic hosts to transmission should be a key focus of research for future invading pathogens, since if substantial nonsymptomatic transmission can occur, surveillance for and isolation of infected hosts who are not yet showing symptoms can be critical for containment of the early outbreak.

As the COVID-19 pandemic continued, it became increasingly clear that individuals of different ages responded differently to infection. In particular, it was suggested that younger individuals may be less susceptible to infection than older individuals [106–110], and more likely to experience a fully asymptomatic course of infection [107, 108, 111–115]. Additionally, the expected numbers of contacts that people have with other individuals depend heavily on age, which can influence the transmission dynamics of viral respiratory pathogens such as SARS-CoV-2 that are primarily transmitted through person-to-person contact [168, 189]. Motivated by the potential influence of these age-related factors on the probability of observing a major outbreak, in Chapter 4 of this thesis we extended the compartmental transmission model developed in Chapter 3 to account for the age structure of the population and the potential for associated variations in susceptibility to SARS-CoV-2 infection, chance of developing clinical symptoms (assumed to lead to a greater infectiousness), and expected

daily patterns of social contacts (see Section 4.2) [107, 168]. Our focus here was to investigate how the probability of a single infected host triggering a major outbreak depended on their age, and to explore how this in turn influenced the effectiveness of different types of control interventions that target different age groups within the population: specifically, school closures, workplace closures, and population-wide social distancing policies.

Using the first-step analysis technique described in Section 2.3, we calculated the probability of a major outbreak beginning from a single infected host of a specified age at the beginning of their infection, and demonstrated that younger individuals were less likely to trigger a major outbreak than older individuals (Section 4.3.1, Fig 4.3). In our model, the much larger expected numbers of daily contacts between young people compared to older people were outweighed by the reduced susceptibility to infection of younger individuals and their increased chance of experiencing a fully asymptomatic infection (which was assumed to make them less infectious than those showing severe clinical symptoms). With this in mind, we then investigated how control interventions that reduce the number of contacts between individuals in the population influence the probability of a major outbreak occurring (Section 4.3.2). Using location-specific breakdowns of expected daily contact patterns, we modelled the effects of school closures, workplace closures, and population-wide social distancing policies on the outbreak risk, and observed that of these three interventions, school closures had the least impact on the probability of observing a major outbreak (Fig 4.4). This can be attributed to the fact that school-aged individuals were amongst the lowest risk groups for triggering an outbreak, and also have many contacts that occur outside the school environment. In contrast, reducing contacts that occurred outside school or the workplace had the greatest effect on the outbreak probability, affecting a greater proportion of the population – including those at highest risk of infection. Naturally, contact-reducing interventions were more successful when used in combination with one another than when used individually (Fig 4.5), and our model provides a framework for exploring how combinations of interventions can be used to reduce the probability of a major outbreak below a specified threshold (Section 4.3.3). Additionally, we considered how effective surveillance strategies could be used in combination with these contact-reducing interventions to further reduce the probability of a major outbreak occurring (Section 4.3.4). Our results demonstrated that if efficient strategies for the surveillance and isolation of infected hosts can be deployed, the stringency of other types of contact-reducing interventions required to suppress the risk of an outbreak can be substantially reduced (Fig 4.6). As in Chapter 3, the value of effective surveillance targeting nonsymptomatic hosts was clearly highlighted – if all nonsymptomatic individuals could be successfully isolated within one day of becoming infected, the probability of a major outbreak would be entirely mitigated by this measure alone (Figs 4.6A, 4.6B).

Overall, in Part I of this thesis we have demonstrated that when a novel pathogen arises in

a population there is clear motivation to rapidly assess the potential for nonsymptomatic transmission to occur. If infected individuals who are not showing symptoms are able to contribute to transmission, surveillance that specifically targets those individuals can play a critical role in disease containment efforts. Additionally, accounting for the age structure of the population and any related heterogeneities in pathogen transmission and symptom expression is crucial for understanding which individuals in the population are likely to contribute most to transmission, which could aid the design of targeted control interventions.

8.1.2 Part I: potential areas for future research

The research presented in Part I of this thesis could be extended in several ways to address additional key questions relating to the optimal deployment of non-pharmaceutical interventions for outbreak risk reduction. One natural direction for future research would be to use the age-structured SARS-CoV-2 transmission model constructed in Chapter 4 to consider the effects of an age-targeted surveillance programme on the probability of a major outbreak. So far, we have demonstrated that effective surveillance for infected hosts and their subsequent rapid isolation can play a crucial role in outbreak containment, with surveillance for nonsymptomatic hosts being of particular benefit, but we have not explored the effects of varying the intensity of surveillance that is aimed at different age groups. Since we have shown that the probability of a major outbreak occurring depends strongly on the age of the index case (Fig 4.3), it would make sense to consider surveillance strategies in which more resources are allocated to age groups that have a higher risk of triggering an outbreak. Additionally, since the proportion of individuals who experience a fully asymptomatic course of infection varies substantially with age [107], a further consideration would be to vary the relative resources that are focused on surveillance for symptomatic and nonsymptomatic hosts within each age group. Plausibly, it may be beneficial to focus surveillance more strongly on nonsymptomatic hosts in age groups that have a high proportion of asymptomatic infection, and prioritise surveillance for symptomatic hosts in age groups where asymptomatic infection is less likely to occur.

Age-dependent surveillance strategies could be implemented in our model by allowing the isolation rates due to surveillance for symptomatic hosts (ρ_k) and nonsymptomatic hosts (σ_k) to vary with the age group k (rather than being assumed equal across all age groups, as in Chapter 4 – see Section 4.2). Of course, to investigate the optimal allocation of surveillance resources between groups, the total amount of available resource would need to be specified, as well as the relationship between the amount of resource allocated to a particular group and the resulting increase in the isolation rate. This could be done analogously to the approach used in Chapter 3, in which we defined a function $f(\rho)$ representing the proportional reduction in the time to isolation for a given surveillance effort ρ and scaled the isolation rates accordingly, allowing for differences in the ease of identifying symptomatic and

nonsymptomatic infected hosts (see Section 3.2.1). For a given fixed total surveillance effort, different allocations of that resource between symptomatic and nonsymptomatic hosts in different age groups could then be compared by evaluating the corresponding age-dependent isolation rates ρ_k and σ_k, and calculating the resulting impact on the probability of a major outbreak.

Another interesting area of investigation would be to consider how the effectiveness of NPIs at reducing the outbreak risk is influenced by the age demographic and typical contact patterns of the specific location being considered. In Chapter 4 of this thesis, we calculated the age-dependent outbreak risk profile using population and contact data from the United Kingdom, but these factors are known to vary substantially between countries [168]. High income countries such as the UK tend to have an older average age, with the population distributed relatively evenly across the age range (apart from at the extreme upper end); in contrast, lower income countries tend to have a much younger population on average, with a clearly triangular age demographic [359, 360]. Additionally, patterns of contacts differ greatly between countries – the contact patterns in higher income countries tend to be more highly assortative with respect to age, whilst lower income countries often have a larger number of intergenerational contacts [168]. The interacting effects of these differences in population structure and behaviour on the age-dependent risk profile are not immediately obvious without conducting further analyses. A younger population may correspond to an increased effectiveness of control interventions that specifically target younger individuals, but more age-diverse mixing patterns may increase the importance of interventions that target the population as a whole. Our model could be used to explore the differing effects of interventions in any of the 177 countries for which contact matrix data are available, taking these interacting factors into account [361].

Throughout this thesis, we have parameterised our models using existing parameter estimates from the literature, where available. However, as noted previously, such estimates often vary considerably between studies – and even within studies, parameter values are commonly provided as distributional estimates rather than point estimates. Since varying the parameter values used will naturally affect the outbreak probability estimates obtained from our models (See Supplementary Figs A.1-A.10, and Supplementary Figs B.1-B.12), accounting for parameter uncertainty in our model predictions is another key area for future work. Uncertainty in parameter values has previously been incorporated explicitly into outbreak probability estimates by Thompson [64], who constructed a branching process model to estimate the probability that a major outbreak of COVID-19 occurs based on the time between symptom onset and hospitalisation (denoted γ). The author first demonstrated how a distributional estimate of γ could be obtained from available line list data, before calculating the outbreak probability under the range of different possible γ values, and

weighting these probability estimates based on the likelihood of the corresponding γ. Thus, they obtained a distributional estimate of the outbreak probability, from which they calculated key quantities including the mean and the 95^{th} percentile. By propagating uncertainty in parameter estimates through to the resulting model predictions in this way, modellers can explore the range of outcomes that may be likely to occur in a given scenario. Policy makers may then use these prediction ranges to design a control strategy based on a specified level of risk aversion. For example, if their goal is to suppress potential outbreaks even in the worst case scenario, they may design an intervention package with the aim of reducing the 95^{th} percentile of the outbreak probability distribution below a specified threshold. In that case, the result would be a more stringent control strategy than would be required if the goal was instead to reduce the mean outbreak probability below the same threshold.

In Part I of this thesis, we have focused on assessing and reducing the probability that a major disease outbreak occurs in the population as a whole. We have not specifically considered the effects of control interventions on the size and duration of any outbreaks that do occur, or the age distribution of cases in the population. In practice, of course, these are also important considerations for a policy maker designing an intervention package, who may have specific goals other than reducing the probability that an outbreak occurs. As well as seeking to reduce the total number of infections that occur, or the peak incidence of the outbreak, there may be particular motivation to minimise infections in the most vulnerable individuals – with older individuals being an obvious example. For many infectious pathogens, including SARS-CoV-2, the severity of infection and risk of death increases with age, and a policy maker may therefore wish to focus particularly on minimising infections in older individuals [68]. Indeed, during the COVID-19 pandemic, older individuals were prioritised for vaccination over younger individuals (excepting those at particularly high risk of infection or severe disease, such as healthcare workers and those with underlying comorbidities)[128–130]. Performing stochastic simulations of an age-structured compartmental model would allow us to assess the effects of NPIs not only on the risk of an outbreak occurring in the first place, but also on the size of any outbreaks that do occur, as well as the number of infections occurring in different age groups within the population. Using that approach, we could explore how combinations of age-targeted NPIs could be used to achieve specific goals such as minimising the number of infections that occur in individuals above a certain age. It would also be interesting to explore the relationship between the effects an intervention has on the outbreak risk, and on the size of outbreaks that do occur, to assess whether outbreak risk estimates could be useful indicators of the broader effectiveness of a control intervention.

Here, it is important to note that defining the optimal outcome of an intervention package is both complex and highly subjective, as it requires quantifying the value of protecting

someone from infection. In practice, assessments of the benefits of interventions are often performed using health outcome measures such as Quality-Adjusted Life Years (QALYs) and Disability-Adjusted Life Years (DALYs) [362, 363]. The QALY is an indicator that combines duration and quality of life, and is calculated by multiplying the number of years of life gained from an intervention with the associated quality of life that will be experienced during those years (on a scale from 0 to 1, where 0 corresponds to death and 1 corresponds to perfect health). The DALY is an alternative measure that instead represents the number of years of life that will be influenced by morbidity or lost due to premature mortality. The number of QALYs gained or DALYs averted can be used to evaluate the cost effectiveness of control interventions and to determine priorities for resource allocation [362–364]. However, the use of these indicators to assess the value of control interventions necessarily requires difficult ethical judgements to be made, with one key example being the question of whether a QALY gained, or DALY averted, is of equal value regardless of the age of the individual affected [363, 365].

Although the outbreak risk estimates generated as part of the research presented in this thesis do not account for these various complexities, they can nonetheless provide useful insights into the likely effectiveness of control interventions that aim to contain infectious disease outbreaks without large numbers of cases occurring. The probability that a major outbreak occurs can be used as a simple metric for comparing the effects of different interventions, with the key advantages of being straightforward to calculate and easy to interpret. Whilst other modelling techniques that tell us about the characteristics of an outbreak once sustained transmission is underway are also important for designing intervention packages, outbreak risk estimates can be a highly useful tool for epidemic preparedness and planning.

8.1.3 Summary of Part II

In Part II of this thesis, we turned our attention to the second of the two key questions outlined above: how to achieve early detection of an invading pathogen.

Rapid identification of an invading pathogen typically requires the deployment of an effective surveillance programme that intentionally and systematically monitors potential hosts for infection. Surveillance programmes are used in the context of human, animal, and plant health, but can be particularly important for pathogens affecting animals and plants, since those hosts are unable to self-report infection. An important challenge for policy makers when designing a surveillance programme is determining how intensively to perform surveillance to ensure that the pathogen is detected while it is still at a low prevalence in the population. A very intensive monitoring programme is likely to be costly and logistically challenging to implement, but if surveillance is not intensive enough, the pathogen is more likely to spread widely in the population before it is detected. In large populations where it is not feasible to

sample every potential host, there may also be a trade off between the scale and frequency of surveillance – a policy maker must determine whether to use resources to perform infrequent surveillance of a large number of individuals, or regular sampling of a smaller number.

Designing effective surveillance programmes becomes particularly challenging when nonsymptomatic transmission can occur, which can present a substantial hindrance to early outbreak control [12, 51, 119, 251, 270]. In Part I of this thesis, we demonstrated that surveillance targeting nonsymptomatic hosts can be crucial for minimising the risk of major outbreaks occurring – but rapid identification of nonsymptomatic hosts can be both challenging and costly. Whilst molecular tests can typically be performed in laboratories to detect infection in samples taken from nonsymptomatic hosts, large scale laboratory testing can be extremely expensive, and requires appropriate infrastructure to be in place. Therefore, for plant pathogens in particular, surveillance programmes typically rely instead on visual inspection as a means of detection – despite this resulting in nonsymptomatic infected hosts going undetected [262, 263, 280]. If the incubation period of the pathogen is long, this can pose a substantial hindrance to early detection. A possible approach explored in this thesis that could be used to improve the effectiveness of visual inspection for invading plant pathogens is to intercrop 'sentinel' plants of an alternative host species with a shorter incubation period amongst the susceptible population [273, 278]. The rapid development of visible symptoms in the sentinel plants could alert growers to the presence of a pathogen more quickly, combatting the problem of nonsymptomatic infection. In Part II of this thesis, we focused on exploring the potential for sentinel plants to be incorporated into monitoring programmes to act as early warning beacons for infection, using the dangerous bacterial pathogen *Xylella fastidiosa*, and the Quick Decline Syndrome it causes in European olive trees (*Olea europaea*), as a case study.

In Chapter 6, we focused on performing a preliminary investigation to assess whether sentinel plants have the potential to facilitate detection of an invading pathogen at a lower prevalence than standard, non-sentinel based, monitoring programmes. In particular, our goal in that chapter was to consider the crucial trade-off between the more rapid detection of an invading pathogen that could be achieved by using sentinel plants, and the higher rate of pathogen transmission that may arise when the size and density of the population is increased. Using *X. fastidiosa* infection in *O. europaea* as a case study, with *C. roseus* as a candidate sentinel species, we constructed a compartmental model of pathogen transmission that accounted for these two different types of plant and their differing incubation periods (Section 6.2.1). In our model, transmission was allowed to occur from any infected host to any susceptible plant, with the rate of transmission depending on the plant type and symptom status of the infector (Fig 6.1). We performed stochastic simulations of pathogen transmission in our model system using the direct method version of the Gillespie SSA (Algorithm 1.1), and simulated a surveillance programme by applying a random sampling scheme to the resulting

epidemic curves (Section 6.2.3). Specifically, samples of a specified size N were selected at random from the population every Δ days, with detection occurring if and only if the sample contained at least one symptomatic plant. When detection occurred, we recorded the total proportion of infected plants present in the population at that time, and repeated this many times to obtain a distribution on the detection prevalence for a given surveillance programme. Throughout, we used the mean of this distribution – the expected detection prevalence, or EDP – as a metric to evaluate and compare the effectiveness of different monitoring programmes (choices of N and Δ).

Initially, we considered a system without any sentinel plants, termed the 'baseline case', and showed how the EDP varied with the choice of N and Δ (Section 6.3.1). Subsequently, when we considered sentinel-based surveillance strategies, we compared the results obtained using those strategies to the baseline case with the same N and Δ, to demonstrate how including sentinel plants in a surveillance programme could reduce the EDP attainable for a given amount of sampling effort. Allowing both the number of sentinels added to the population (P_S) and the number of sentinels included in each sample (N_S) to vary, we applied a Bayesian optimisation algorithm to determine the optimal sentinel-based sampling approach (choice of P_S and N_S) for a given choice of sample size N and sample interval Δ (Sections 6.3.2, 6.3.3). Our results clearly demonstrated that sentinel plants could substantially reduce the EDP obtained with a given surveillance effort, indicating that sentinels could be a promising strategy for improving early detection monitoring (Fig 6.10). Crucially, however, the optimal sentinel strategy was shown to vary substantially depending on the surveillance effort applied (Fig 6.8), with suboptimal strategies (e.g. strategies in which too many or too few sentinels are planted) resulting in smaller decreases or even increases in the EDP (Figs 6.5, 6.6). Therefore, our results caution that potential sentinel-based surveillance strategies must be carefully evaluated in the specific context being considered, accounting for the transmission characteristics of the pathosystem in question and the amount of resource available for sampling.

Motivated by the encouraging results in support of sentinel-based surveillance strategies that we observed in Chapter 6, in Chapter 7 of this thesis we extended our model to incorporate additional realism, accounting for the effects of seasonality on pathogen transmission and detection. In particular, we allowed for seasonal fluctuations in the size of the insect vector population, which influences the risk of pathogen introduction at different times of year and the subsequent rate of pathogen transmission, as well as seasonal variation in temperature and precipitation, which respectively affect the rate at which symptoms of *X. fastidiosa* infection develop and the degree to which they are expressed (Sections 7.2.1, 7.2.2) [263, 317, 322, 334, 335]. We began by investigating whether our previous conclusions regarding the potential usefulness of sentinel plants were influenced by the inclusion of these

seasonal factors, and demonstrated that our key results from Chapter 6 were qualitatively unchanged – that is, when seasonality was accounted for, including sentinel plants in the population was still beneficial for achieving early detection (Section 7.3.1). With this confirmed, we then began to explore how surveillance resources could be optimally allocated throughout the year, based on the seasonal factors included in our model. Specifically, we considered a scenario in which surveillance was restricted to a specified window of the year, with the available annual sampling resources focused into that period (Section 7.2.4). By considering a range of different possible windows, we determined that in the absence of sentinel plants, the optimal time to perform surveillance was between late April/early May and late August/mid September, which is in line with current EU guidelines regarding the timing of visual inspection for *X. fastidiosa* infection (Section 7.3.2) [341]. By focusing surveillance resources into that window, the EDP was substantially reduced compared to a scenario in which surveillance resources were spread uniformly throughout the year (Fig 7.7). Finally, we considered including sentinel plants in the population in combination with a seasonally focused surveillance strategy (Section 7.3.3). When the sentinel strategy and sampling window were optimised simultaneously, we observed that including sentinel plants in the population was still beneficial – that is, sentinel plants were able to achieve additional reductions in the EDP compared to those already achieved by focusing surveillance into the optimal window of the year (Fig 7.9D). Interestingly, the optimal window in which to perform surveillance widened when sentinel plants were included in the population, with the best results obtained when sampling took place until late September/mid October (Fig 7.9A). This highlights the need to consider the interacting effects of the various components of a sampling scheme, including the choice of sample size, sample interval, timing of sampling, and the inclusion of sentinels, rather than treating these as independent factors that can be chosen separately from one another.

Overall, in Part II of this thesis we have shown that sentinel plants show great promise as a tool for improving early detection monitoring for invasive plant pathogens, provided that the choice of sentinel based strategy and the time of year when sampling takes place are evaluated carefully in line with the surveillance resources that are available. If designed optimally, sentinel based strategies have the potential to lead to substantial reductions in the prevalence of infection in the population at the time of detection compared to standard, non-sentinel based approaches, which could promote the effectiveness of subsequent control interventions and reduce the risk of the pathogen spreading to new locations.

8.1.4 Part II: potential areas for future research

The work completed in Part II of this thesis strongly encourages further research into the potential to use sentinel plants as a means to facilitate earlier detection for invasive plant pathogens. One natural direction for further investigation would be to extend our

transmission models to account for the spatial structure of the population. Throughout Part II of this thesis, we considered well-mixed transmission models that were not spatially explicit – that is, we assumed that the risk of any particular plant infecting any other plant was governed solely by the type (crop or sentinel) and model class (Healthy, Undetectable, or Detectable) of either plant, and was unaffected by physical proximity. Although well-mixed models are commonly used in mathematical epidemiology, and this has formed a useful starting point for gaining insight into the potential benefits of incorporating sentinel plants into a surveillance strategy, in reality spatial factors play an important role in shaping transmission dynamics [47]. As well as incorporating additional realism into our framework, the extension to a spatial model would allow us to consider important questions that could not be addressed using the well-mixed model, such as where sentinel plants should be placed within the population, and which specific plants (both crops and sentinels) should be included in a sampling scheme. These questions become particularly pertinent when we consider the possibility of spatial variation in the probability of primary infection. That is, a pathogen may be more likely to make its initial appearance in certain plants than others, based on their physical locations within the population [245, 283]. If this is the case, the strategic placement of sentinel plants near the locations of highest risk and a corresponding spatially heterogeneous sampling strategy may be beneficial. For example, an olive grove situated just north of a known infected area may experience a greater risk of invasion from the south than from any other direction; in this case, there may be an advantage to placing sentinels along the south border of the grove and sampling preferentially from that area. However, since the degree of variation in the risk of primary infection would depend on the size of the grove in question and its distance from the source of inoculum, mathematical modelling is important for understanding the conditions under which spatially heterogeneous sampling is worthwhile.

The spatial structure of the population could be incorporated into our model by allowing the rate of transmission from an infected (Undetectable or Detectable) plant to a susceptible (Healthy) plant to depend not only on the type and symptom status of the infector, which our models currently account for, but also on the physical distance between the two plants. In spatial models of pathogen transmission, the effect of distance on infection pressure is typically captured by a dispersal kernel $K(d_{ij};\alpha)$, a non-negative function that describes how the infection challenge posed by an infectious plant i to a susceptible plant j varies with the distance d_{ij} between them, and the mean dispersal distance of the pathogen in a given time period, α [366]. The dispersal kernel acts as a scaling factor for the force of infection exerted by an infected plant i on a susceptible plant j, and typically decreases with the distance between the plants, d_{ij}, to reflect the fact that the risk of infection becomes lower as you move further from the source of inoculum [249]. For example, the exponential kernel, $K(d_{ij};\alpha) = \exp(-d_{ij}/\alpha)$, is a common choice for simulating epidemics that are characterised by wave-like spread [34], and has previously been used to represent X.

fastidiosa transmission in Apulia [260]. To allow for spatial transmission patterns in our model, we would need to specify coordinates for the crop plants in our population so that the distances between pairs of plants could be calculated, and select an appropriate dispersal kernel K with which to scale the transmission coefficients. Then, we would need to adapt our simulation algorithm to explicitly keep track of the infection status of every plant in the population, and to recalculate the distance-dependent force of infection acting on each susceptible plant whenever a new infection occurs.

As discussed in Chapters 6 and 7, another natural extension to our analyses would be to account explicitly for the dynamics of the vector population. Doing so would allow us to explore the effects of various factors that we have not yet considered, such as how the infection status of a plant influences its attractiveness to vectors [320, 367]. Additionally, as mentioned in Section 7.4, this extension would allow for a more nuanced representation of how the prevalence of infection in the vector population varies over time and with the seasonal fluctuations in the population size, rather than assuming that the infection prevalence is proportional to the population size. Compartmental models of host-vector transmission are commonly used to understand the transmission dynamics of a variety of pathogens, and it would be conceptually straightforward to adapt our model to include additional vector compartments and account for host-vector interactions [77, 368]. However, as is often the case when using a mathematical model to represent a biological system, there is a trade off between increasing the complexity of the model, which has the advantage of being able to capture more realism, and the corresponding increased difficulty of parameterising the model appropriately and the time taken to perform model simulations. Constructing a full host-vector model of *X. fastidiosa* transmission in *O. europaea* and *C. roseus* would require additional data on vector preferences for plants of different species and infection statuses, the relative rates at which vectors acquire the pathogen from hosts of different species and sizes, and so on, which are currently unavailable.

In our work so far, we have considered sentinel plants solely in the context of how they could be used to improve the effectiveness of surveillance strategies that rely on visual inspection. It would be interesting to consider whether sentinel plants could have additional benefits when linked with other approaches for control – for example, in the context of sampling for laboratory testing. The overall sensitivity of a laboratory-based surveillance method depends not only on the diagnostic sensitivity of the laboratory test applied (i.e. the probability that a truly positive sample tests positive), but also on the probability that a sample taken from an infected host actually contains infected material [369]. For a large potential host plant such as *O. europaea*, only a small proportion of the plant can be sampled and tested at any given time. Therefore, if a host is infected but not yet displaying visible symptoms, the chance of selecting a sample that contains infected material may be small. In contrast,

for smaller potential hosts such as our candidate sentinel plant *C. roseus*, a sample of the same size selected from an asymptomatic infected plant may have a higher probability of containing infected material, increasing the overall sensitivity of the surveillance method. In addition to facilitating early detection through visual inspection, sentinels may therefore also have the potential to enhance the effectiveness of diagnostic testing. Again, the high cost of laboratory testing compared to visual inspection would motivate both seasonal and spatial targeting of sampling and testing resources to maximise the probability of correctly identifying an infected plant in a given sample [263]. More generally, investigating the effects of sentinel-based surveillance strategies when used in combination with other control methods (e.g. insecticides) will be important for the design of integrated control strategies going forwards.

Throughout Part II of this thesis, we have assessed the effectiveness of sentinel-based surveillance strategies by considering their effects on the EDP compared to non-sentinel-based strategies that use an equivalent amount of surveillance resource. However, we have not attempted to quantify the economic costs and benefits of using sentinel plants in this way, which in practice is a fundamental consideration for a policy maker seeking to design a surveillance programme. Achieving a reduced EDP could have substantial economic advantages. If an invading pathogen can be detected when only a small number of plants in the population have been infected, the negative effects of subsequently implemented control measures could be reduced [253, 264]. For example, if the control strategy being applied is the removal of infected hosts and any other plants within a certain distance, a smaller EDP could result in fewer plants being destroyed overall, thus reducing the cost of implementing the control strategy and minimising the loss of valuable plant material. In addition to localised cost reductions of this kind, a smaller EDP may also reduce the opportunity for the pathogen to be exported to other susceptible regions, which could have a knock-on effect on the cost of control on a larger scale [34]. However, the cost of deploying a sentinel-based strategy must also be considered. Although the per-plant cost of performing visual inspection on sentinels is plausibly no greater that it is for the main crop species, their initial sourcing, planting, and subsequent maintenance are likely to incur expense, especially if the optimal surveillance strategy requires deploying a large number of sentinel plants. In particular, sentinel plants should undergo laboratory testing for infection before being introduced to a population to ensure that they themselves do not act as a source of inoculum, which will also come at a cost. Therefore, before a sentinel-based strategy is deployed, it is essential to investigate whether the reduction in EDP and the correspondingly smaller costs of control that the strategy can achieve are sufficient to outweigh the costs of deploying it. This may vary depending on the spatial scale being considered – an expensive monitoring programme may not be cost effective at the local scale, but may be beneficial when the cost of the pathogen generating widespread infection elsewhere is also considered.

As in Part I of this thesis, throughout Part II we parameterised our transmission models using existing parameter estimates from the literature, where available. However, as noted previously, such estimates are inevitably subject to some uncertainty – and for some parameters estimates were not available, so we were required to make informed assumptions about their likely values. Although we have demonstrated that our results remain qualitatively sound under parameter variation (see Supplementary Figs C.5-C.13, and Supplementary Figs D.1-D.3), our quantitative conclusions regarding the optimal choice of surveillance strategy are naturally affected by changes in the model parameters. Since we have also shown that incorporating sentinel plants into a monitoring programme can be detrimental if a highly sub-optimal surveillance strategy is selected, for a specific sentinel-based approach to be recommended in practice it must be demonstrably beneficial over a range of likely parameter sets, with a minimal risk of accidentally increasing the EDP. To identify robust strategies of this kind, our model could be extended to determine not only the optimal number of sentinels to include in the population for a given parameterisation and choice of sampling strategy, but also a range of strategies near to that optimum which perform acceptably well (for example, all strategies that are able to reduce the EDP below a specified level). By identifying an acceptable set of potential strategies for a variety of different model parameterisations, one could select an approach that provides a reasonably good outcome in as many cases as possible (see Cunniffe *et al.* [34] for a similar approach). The result would be a strategy that, although not necessarily optimal, performs well across a range of plausible parameter values, and is therefore likely to be reliably beneficial in practice.

Prior to the investigation conducted in this thesis, research into the potential for sentinel plants to facilitate the early detection of invasive plant pathogens was lacking. Here, we have conducted the first quantitative analysis of the benefits and drawbacks of using sentinel plants for surveillance, using one of the current most dangerous plant pathogens worldwide as a case study. The results presented in this thesis suggest that sentinel plants may indeed be a highly useful tool for improving plant disease monitoring programmes, provided that the choice of sentinel strategy is carefully evaluated with reference to the specific pathosystem in question and the resources that are available for surveillance. Although there is still much work to be done before sentinel plants can realistically be deployed as early warning systems for invasive pathogens, our results strongly encourage further research into their use, and we hope that other researchers will continue to investigate this promising approach for early outbreak detection.

8.2 Concluding remarks

The importance of continued research into effective methods for the surveillance and control of infectious pathogens will only increase in the coming years [347]. In the time frame of writing this thesis alone, we have witnessed the emergence of the novel coronavirus SARS-CoV-2 and the resulting COVID-19 pandemic, a global outbreak of mpox which – as of the end of August 2023 – has resulted in almost $90,000$ infections across 114 countries [370], and three successive years of particularly severe avian influenza outbreaks [371], to name but a few high-profile examples. With the ever-increasing risk of novel zoonotic pathogens emerging and non-native pathogens being transported to previously uninfected locations, the early identification of new infectious disease threats and the rapid response to outbreaks is likely to remain a key target for global disease control organisations. Indeed, there is increasing interest in the simple '7-1-7' principle for rapid response to an infectious disease outbreak in humans, proposed by Frieden *et al.* in 2021 [372]. This rule suggests that appropriate frameworks should be put in place to ensure that: 1) any infectious disease outbreak is detected within seven days of emergence; 2) the outbreak is then reported to public health authorities within one more day; and 3) a response to the outbreak is subsequently deployed within a further seven days. Designed to provide a clearly understandable target for stakeholders that also acts as a quantifiable metric for the timeliness of an outbreak response, the 7-1-7 principle has since been adopted by the WHO Regional Committee for Africa in their 2022-2030 Regional Strategy for Health Security and Emergencies, and has been piloted in six countries [373, 374]. However, despite this endorsement of the benefits of early outbreak detection and containment, the challenges associated with achieving this remain substantial [375]. As we have emphasised throughout this thesis, there are many complex and interacting factors affecting pathogen transmission that need to be accounted for when responding to a disease threat. Mathematical modelling will therefore continue to be a crucial tool for outbreak preparedness and planning, allowing policy makers to integrate these complex factors into analytical frameworks that can be used to generate projections during outbreaks, explore possibilities for control, and evaluate the benefits and drawbacks of different interventions.

Bibliography

[1] D. M. Morens and A. S. Fauci. Emerging Infectious Diseases: Threats to Human Health and Global Stability. PLoS Pathog., 9(7):e1003467, 2013. doi: 10.1371/journal.ppat.1003467.

[2] D. E. Bloom and D. Cadarette. Infectious Disease Threats in the Twenty-First Century: Strengthening the Global Response. Front. Immunol., 10:549, 2019. doi: 10.3389/fimmu.2019.00549.

[3] K. E. Jones, N. G. Patel, M. A. Levy, A. Storeygard, D. Balk, J. L. Gittleman, et al. Global trends in emerging infectious diseases. Nature, 451(7181):990–993, 2008. doi: 10.1038/nature06536.

[4] R. N. Strange and P. R. Scott. Plant disease: a threat to global food security. Annu Rev Phytopathol, 43:83–116, 2005. doi: 10.1146/annurev.phyto.43.113004.133839.

[5] N. Spence, L. Hill, and J. Morris. How the global threat of pests and diseases impacts plants, people, and the planet. Plants people planet, 2(1):5–13, 2020. doi: 10.1002/ppp3.10088.

[6] P. Daszak, A. A. Cunningham, and A. D. Hyatt. Emerging Infectious Diseases of Wildlife - Threats to Biodiversity and Human Health. Science, 287(5452):443, 2000. doi: 10.1126/science.287.5452.443.

[7] S. Riedel. Edward Jenner and the history of smallpox and vaccination. Proc. Bayl. Univ. Med. Cent., 18(1):21–5, 2005. doi: 10.1080/08998280.2005.11928028.

[8] R. I. Aminov. A brief history of the antibiotic era: lessons learned and challenges for the future. Front. Microbiol., 1(134), 2010. doi: 10.3389/fmicb.2010.00134.

[9] M. A. Strassburg. The global eradication of smallpox. Am. J. Infect. Control, 10(2): 53–59, 1982. doi: 10.1016/0196-6553(82)90003-7.

[10] P. Roeder, J. Mariner, and R. Kock. Rinderpest: the veterinary perspective on eradication. Philos. Trans. R. Soc. Lond., B, Biol. Sci., 368(1623):20120139, 2013. doi: 10.1098/rstb.2012.0139.

[11] Y. A. Tajudeen, H. J. Oladipo, R. O. Yusuf, I. O. Oladunjoye, A. O. Adebayo, A. F. Ahmed, et al. The Need to Prioritize Prevention of Viral Spillover in the Anthropopandemicene: A Message to Global Health Researchers and Policymakers. Challenges, 13(2):35, 2022. doi: 10.3390/challe13020035.

[12] C. Fraser, S. Riley, R. M. Anderson, and N. M. Ferguson. Factors that make an infectious disease outbreak controllable. Proc. Natl. Acad. Sci. U.S.A., 101(16):6146–6151, 2004. doi: 10.1073/pnas.0307506101.

[13] Z. Allam. The Second 50 days: A Detailed Chronological Timeline and Extensive Review of Literature Documenting the COVID-19 Pandemic From Day 50 to Day 100. In: Surveying the Covid-19 Pandemic and its Implications, pages 9–39, 2020. doi: 10.1016/B978-0-12-824313-8.00002-4.

[14] Z. Allam. The Third 50 Days: A Detailed Chronological Timeline and Extensive Review of Literature Documenting the COVID-19 Pandemic From Day 100 to Day 150. In: Surveying the Covid-19 Pandemic and its Implications, pages 41–69, 2020. doi: 10.1016/B978-0-12-824313-8.00003-6.

[15] D. M. Bell. Public health interventions and SARS spread, 2003. Emerging Infect. Dis., 10(11):1900–6, 2004. doi: 10.3201/eid1011.040729.

[16] A. Bouba, K. B. Helle, and K. A. Schneider. Predicting the combined effects of case isolation, safe funeral practices, and contact tracing during Ebola virus disease outbreaks. PLoS ONE, 18(1):e0276351, 2023. doi: 10.1371/journal.pone.0276351.

[17] Y. Wang, L. Ling, Z. Zhang, and A. Marin-Lopez. Current Advances in Zika Vaccine Development. Vaccines, 10(11):1816, 2022. doi: 10.3390/vaccines10111816.

[18] T. Y. V. de Lima Cavalcanti, M. R. Pereira, S. O. de Paula, and R. F. de Oliveira Franca. A Review on Chikungunya Virus Epidemiology, Pathogenesis and Current Vaccine Development. Viruses, 14(5):969, 2022. doi: 10.3390/v14050969.

[19] J. Pintado Silva and A. Fernandez-Sesma. Challenges on the development of a dengue vaccine: a comprehensive review of the state of the art. J. Gen. Virol., 104(3), 2023. doi: 10.1099/jgv.0.001831.

[20] D. E. te Beest, T. J. Hagenaars, J. A. Stegeman, M. P. G. Koopmans, and M. van Boven. Risk based culling for highly infectious diseases of livestock. Vet. Res., 42(1): 81, 2011. doi: 10.1186/1297-9716-42-81.

[21] N. J. Cunniffe, F. F. Laranjeira, F. M. Neri, R. E. DeSimone, and C. A. Gilligan. Cost-Effective Control of Plant Disease When Epidemiological Knowledge is Incomplete: Modelling Bahia Bark Scaling of Citrus. PLoS Comput. Biol., 10(8):e1003753, 2014. doi: 10.1371/journal.pcbi.1003753.

[22] M. A. Pandit, J. Kumar, S. Gulati, N. Bhandari, P. Mehta, R. Katyal, et al. Major Biological Control Strategies for Plant Pathogens. Pathogens, 11(2):273, 2022. doi: 10.3390/pathogens11020273.

[23] L. Steele, E. Orefuwa, S. Bino, S. R. Singer, J. Lutwama, and P. Dickmann. Earlier Outbreak Detection - A Generic Model and Novel Methodology to Guide Earlier Detection Supported by Data From Low- and Mid-Income Countries. Front. Public Health, 8(452), 2020. doi: 10.3389/fpubh.2020.00452.

[24] I. Buja, E. Sabella, A. G. Monteduro, M. S. Chiriacò, L. De Bellis, A. Luvisi, et al. Advances in Plant Disease Detection and Monitoring: From Traditional Assays to In-Field Diagnostics. Sensors, 21(6):2129, 2021. doi: 10.3390/s21062129.

[25] N. Eyal, A. Gheaus, A. Gosseries, M. Magalhaes, T. Ngosso, B. Steuwer, et al. Coronavirus Disease 2019 (COVID-19) Vaccine Prioritization in Low- and Middle-Income Countries May Justifiably Depart From High-Income Countries' Age Priorities. Clin. Infect. Dis., 75(Suppl 1):s93–s97, 2022. doi: 10.1093/cid/ciac398.

[26] D. Bernoulli. Essai d'une nouvelle analyse de la mortalité causée par la petite vérole et desavantages de l'inoculation pour la prévenir. Mémoires de Mathématiques et de Physique, Académie Royale des Sciences, Paris, pages 1–45, 1760.

[27] A. Boylston. The origins of inoculation. J. R. Soc. Med., 105(7):309–13, 2012. doi: 10.1258/jrsm.2012.12k044.

[28] D. Bernoulli and S. Blower. An attempt at a new analysis of the mortality caused by smallpox and of the advantages of inoculation to prevent it. Rev. Med. Virol., 14(5): 275–288, 2004. doi: 10.1002/rmv.443.

[29] M. Kretzschmar. Disease modeling for public health: added value, challenges, and institutional constraints. J. Public Health Policy, 41(1):39–51, 2020. doi: 10.1057/s41271-019-00206-0.

[30] E. S. McBryde, M. T. Meehan, O. A. Adegboye, A. I. Adekunle, J. M. Caldwell, A. Pak, et al. Role of modelling in COVID-19 policy development. Paediatr. Respir. Rev., 35:57–60, 2020. doi: 10.1016/j.prrv.2020.06.013.

[31] N. M. Ferguson, C. A. Donnelly, and R. M. Anderson. The Foot-and-Mouth Epidemic in Great Britain: Pattern of Spread and Impact of Interventions. Science, 292(5519): 1155–1160, 2001. doi: 10.1126/science.1061020.

[32] M. B. Barongo, R. P. Bishop, E. M. Fèvre, D. L. Knobel, and A. Ssematimba. A Mathematical Model that Simulates Control Options for African Swine Fever Virus (ASFV). PLoS ONE, 11(7):e0158658, 2016. doi: 10.1371/journal.pone.0158658.

[33] A. Malek and A. Hoque. Mathematical modeling of bird flu with vaccination and treatment for the poultry farms. Comp. Immunol. Microbiol. Infect. Dis., 80:101721, 2022. doi: 10.1016/j.cimid.2021.101721.

[34] N. J. Cunniffe, R. O. J. H. Stutt, R. E. DeSimone, T. R. Gottwald, and C. A. Gilligan. Optimising and Communicating Options for the Control of Invasive Plant Disease When There is Epidemiological Uncertainty. PLoS Comput. Biol., 11(4):e1004211, 2015. doi: 10.1371/journal.pcbi.1004211.

[35] S. Parnell, T. R. Gottwald, T. Riley, and F. van den Bosch. A generic risk-based surveying method for invading plant pathogens. Ecol. Appl., 24(4):779–790, 2014. doi: 10.1890/13-0704.1.

[36] A. G. McKendrick. Studies on the Theory of Continuous Probabilities, with Special Reference to its Bearing on Natural Phenomena of a Progressive Nature. Proc. London Math., s2-13(1):401–416, 1914. doi: 10.1112/plms/s2-13.1.401.

[37] K. Prem, Y. Liu, T. W. Russell, A. J. Kucharski, R. M. Eggo, N. Davies, et al. The effect of control strategies to reduce social mixing on outcomes of the COVID-19 epidemic in Wuhan, China: a modelling study. Lancet Public Health, 5(5):e261–e270, 2020. doi: 10.1016/S2468-2667(20)30073-6.

[38] G. Chowell, N. W. Hengartner, C. Castillo-Chavez, P. W. Fenimore, and J. M. Hyman. The basic reproductive number of Ebola and the effects of public health measures: the cases of Congo and Uganda. J. Theor. Biol., 229(1):119–126, 2004. doi: 10.1016/j.jtbi.2004.03.006.

[39] S. Khailaie, T. Mitra, A. Bandyopadhyay, M. Schips, P. Mascheroni, P. Vanella, et al. Development of the reproduction number from coronavirus SARS-CoV-2 case data in Germany and implications for political measures. BMC Med., 19(1):32, 2021. doi: 10.1186/s12916-020-01884-4.

[40] E. Southall, Z. Ogi-Gittins, A. R. Kaye, W. S. Hart, F. A. Lovell-Read, and R. N. Thompson. A practical guide to mathematical methods for estimating infectious disease outbreak risks. J. Theor. Biol., 562:111417, 2023. doi: 10.1016/j.jtbi.2023.111417.

[41] R. N. Thompson, E. Southall, Y. Daon, F. A. Lovell-Read, S. Iwami, C. P. Thompson, et al. The impact of cross-reactive immunity on the emergence of SARS-CoV-2 variants. Front. Immunol., 13:1049458, 2022. doi: 10.3389/fimmu.2022.1049458.

[42] R. N. Thompson, C. A. Gilligan, and N. J. Cunniffe. Will an outbreak exceed available resources for control? Estimating the risk from invading pathogens using practical definitions of a severe epidemic. J. R. Soc. Interface, 17(172):20200690, 2020. doi: 10.1098/rsif.2020.0690.

[43] P. Zhang, K. Feng, Y. Gong, J. Lee, S. Lomonaco, and L. Zhao. Usage of Compartmental Models in Predicting COVID-19 Outbreaks. AAPS J., 24(5):98, 2022. doi: 10.1208/s12248-022-00743-9.

[44] J. Asher. Forecasting Ebola with a regression transmission model. Epidemics, 22: 50–55, 2018. doi: 10.1016/j.epidem.2017.02.009.

[45] B. Ridenhour, J. M. Kowalik, and D. K. Shay. Unraveling R_0: considerations for public health applications. Am. J. Public. Health., 104(2):e32–e41, 2014. doi: 10.2105/AJPH.2013.301704.

[46] N. G. Davies, A. J. Kucharski, R. M. Eggo, A. Gimma, W. J. Edmunds, T. Jombart, et al. Effects of non-pharmaceutical interventions on COVID-19 cases, deaths, and demand for hospital services in the UK: a modelling study. Lancet Public Health, 5 (7):E375–E385, 2020. doi: 10.1016/S2468-2667(20)30133-X.

[47] A. Kleczkowski, A. Hoyle, and P. McMenemy. One model to rule them all? Modelling approaches across OneHealth for human, animal and plant epidemics. Philos. Trans. R. Soc. Lond., B, Biol. Sci., 374(1775):20180255, 2019. doi: 10.1098/rstb.2018.0255.

[48] D. T. Gillespie. Exact stochastic simulation of coupled chemical reactions. J. Phys. Chem., 81(25):2340–2361, 1977. doi: 10.1021/j100540a008.

[49] R. Erban and S. J. Chapman. Stochastic Modelling of Reaction–Diffusion Processes. Cambridge Texts in Applied Mathematics. Cambridge University Press, Cambridge, 2020. doi: 10.1017/9781108628389.

[50] Z. Allam. The First 50 days of COVID-19: A Detailed Chronological Timeline and Extensive Review of Literature Documenting the Pandemic. In: Surveying the

Covid-19 Pandemic and its Implications, pages 1–7, 2020. doi: 10.1016/B978-0-12-824313-8.00001-2.

[51] V. Alonso-Chavez, S. Parnell, and F. van den Bosch. Monitoring invasive pathogens in plant nurseries for early-detection and to minimise the probability of escape. J. Theor. Biol., 407:290–302, 2016. doi: 10.1016/j.jtbi.2016.07.041.

[52] C. M. Brasier. The biosecurity threat to the UK and global environment from international trade in plants. Plant Pathol., 57(5):792–808, 2008. doi: 10.1111/j.1365-3059.2008.01886.x.

[53] P. E. Hulme. Trade, transport and trouble: managing invasive species pathways in an era of globalization. J. Appl. Ecol., 46(1):10–18, 2009. doi: 10.1111/j.1365-2664.2008.01600.x.

[54] A. M. Liebhold, E. G. Brockerhoff, L. J. Garrett, J. L. Parke, and K. O. Britton. Live plant imports: the major pathway for forest insect and pathogen invasions of the US. Front. Ecol. Environ., 10(3):135–143, 2012. doi: 10.1890/110198.

[55] B. J. Quilty, S. Clifford, CMMID nCoV working group, S. Flasche, and R. M. Eggo. Effectiveness of airport screening at detecting travellers infected with novel coronavirus (2019-nCoV). Euro Surveill., 25(5):2000080, 2020. doi: 10.2807/1560-7917.ES.2020.25.5.2000080.

[56] M. T. Rahman, M. A. Sobur, M. S. Islam, S. Ievy, M. J. Hossain, M. E. El Zowalaty, et al. Zoonotic Diseases: Etiology, Impact, and Control. Microorganisms, 8(9):1405, 2020. doi: 10.3390/microorganisms8091405.

[57] P. M. Sharp and B. H. Hahn. Origins of HIV and the AIDS pandemic. Cold Spring Harb. Perspect. Med., 1(1):a006841, 2011. doi: 10.1101/cshperspect.a006841.

[58] D. Quammen. Ebola: The natural and human history of a deadly virus. W. W. Norton, 2014.

[59] E. C. Holmes, S. A. Goldstein, A. L. Rasmussen, D. L. Robertson, A. Crits-Christoph, J. O. Wertheim, et al. The origins of SARS-CoV-2: A critical review. Cell, 184(19):4848–4856, 2021. doi: 10.1016/j.cell.2021.08.017.

[60] M. N. Sparrer, N. F. Hodges, T. Sherman, S. VandeWoude, A. M. Bosco-Lauth, and C. E. Mayo. Role of Spillover and Spillback in SARS-CoV-2 Transmission and the Importance of One Health in Understanding the Dynamics of the COVID-19 Pandemic. J. Clin. Microbiol., 61(7):e01610–22, 2023. doi: 10.1128/jcm.01610-22.

[61] A. J. Mastin, F. van den Bosch, T. R. Gottwald, V. Alonso Chavez, and S. R. Parnell. A method of determining where to target surveillance efforts in heterogeneous epidemiological systems. PLoS Comput. Biol., 13(8):e1005712, 2017. doi: 10.1371/journal.pcbi.1005712.

[62] Centers for Disease Control and Prevention. Ebola Disease Distribution Map: Cases of Ebola Disease in Africa Since 1976, Accessed 24 May 2023. URL `https://www.cdc.gov/vhf/ebola/history/distribution-map.html`.

[63] M. E. Craft, H. L. Beyer, and D. T. Haydon. Estimating the Probability of a Major Outbreak from the Timing of Early Cases: An Indeterminate Problem? PLoS ONE, 8(3):e57878, 2013. doi: 10.1371/journal.pone.0057878.

[64] R. N. Thompson. Novel Coronavirus Outbreak in Wuhan, China, 2020: Intense Surveillance Is Vital for Preventing Sustained Transmission in New Locations. J. Clin. Med., 9(2):498, 2020. doi: 10.3390/jcm9020498.

[65] F. A. Lovell-Read, S. Funk, U. Obolski, C. A. Donnelly, and R. N. Thompson. Interventions targeting nonsymptomatic cases can be important to prevent local outbreaks: SARS-CoV-2 as a case-study. J. R. Soc. Interface, 18(178):20201014, 2021. doi: 10.1098/rsif.2020.1014.

[66] F. A. Lovell-Read, S. Shen, and R. N. Thompson. Estimating local outbreak risks and the effects of non-pharmaceutical interventions in age-structured populations: SARS-CoV-2 as a case study. J. Theor. Biol., 535:110983, 2022. doi: 10.1016/j.jtbi.2021.110983.

[67] R. Sachak-Patwa, H. M. Byrne, L. Dyson, and R. N. Thompson. The risk of SARS-CoV-2 outbreaks in low prevalence settings following the removal of travel restrictions. Commun. Med., 1(1):39, 2021. doi: 10.1038/s43856-021-00038-8.

[68] J. R. Glynn and P. A. H. Moss. Systematic analysis of infectious disease outcomes by age shows lowest severity in school-age children. Sci. Data, 7(1):329, 2020. doi: 10.1038/s41597-020-00668-y.

[69] G. Sorci and B. Faivre. Age-dependent virulence of human pathogens. PLoS Pathog., 18(9):e1010866, 2022. doi: 10.1371/journal.ppat.1010866.

[70] M. Porta. A Dictionary of Epidemiology. Oxford University Press, 6th edition, 2014.

[71] Centers for Disease Control and Prevention. 2014-2016 Ebola Outbreak in West Africa, Accessed 24 May 2023. URL `https://www.cdc.gov/vhf/ebola/history/2014-2016-outbreak/index.html`.

[72] World Health Organisation. WHO Director-General's opening remarks at the media briefing on COVID-19 - 11 March 2020. URL `https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020`.

[73] I. Nasell. The Threshold Concept in Stochastic Epidemic and Endemic Models. In: Epidemic Models: Their Structure and Relation to Data. Cambridge University Press, 1995.

[74] T. Britton. Stochastic epidemic models: A survey. Math. Biosci., 225(1):24–35, 2010. doi: 10.1016/j.mbs.2010.01.006.

[75] J. C. Miller. A primer on the use of probability generating functions in infectious disease modeling. Infect. Dis. Model., 3:192–248, 2018. doi: 10.1016/j.idm.2018.08.001.

[76] J. R. Norris. Markov Chains (Cambridge Series in Statistical and Probabilistic Mathematics). Cambridge University Press, Cambridge, 1997. doi: 10.1017/CBO9780511810633.

[77] A. R. Kaye, W. S. Hart, J. Bromiley, S. Iwami, and R. N. Thompson. A direct comparison of methods for assessing the threat from emerging infectious diseases in seasonally varying environments. J. Theor. Biol., 548:111195, 2022. doi: 10.1016/j.jtbi.2022.111195.

[78] M. Hartfield and S. Alizon. Epidemiological feedbacks affect evolutionary emergence of pathogens. Am. Nat., 183(4):E105–17, 2014. doi: 10.1086/674795.

[79] B. Hu, H. Guo, P. Zhou, and Z.-L. Shi. Characteristics of SARS-CoV-2 and COVID-19. Nat. Rev. Microbiol., 19(3):141–154, 2021. doi: 10.1038/s41579-020-00459-7.

[80] World Health Organisation. Naming the coronavirus disease (COVID-19) and the virus that causes it, Accessed 29 May 2023. URL https://www.who.int/emergencies/diseases/novel-coronavirus-2019/technical-guidance/naming-the-coronavirus-disease-(covid-2019)-and-the-virus-that-causes-it.

[81] A. L. Mueller, M. S. McNamara, and D. A. Sinclair. Why does COVID-19 disproportionately affect older people? Aging, 12(10):9959–9981, 2020. doi: 10.18632/aging.103344.

[82] P. Adab, S. Haroon, M. E. O'Hara, and R. E. Jordan. Comorbidities and COVID-19. BMJ, 377:o1431, 2022. doi: 10.1136/bmj.o1431.

[83] H. Tian, Y. Liu, Y. Li, C. H. Wu, B. Chen, M. U. G. Kraemer, et al. An investigation of transmission control measures during the first 50 days of the COVID-19 epidemic in China. Science, 368(6491):638–642, 2020. doi: 10.1126/science.abb6105.

[84] World Health Organisation. WHO Coronavirus (COVID-19) Dashboard, Accessed 29 May 2023. URL https://covid19.who.int.

[85] N. Wang, Y. Fu, H. Zhang, and H. Shi. An evaluation of mathematical models for the outbreak of COVID-19. Precis. Clin. Med., 3(2):85–93, 2020. doi: 10.1093/pcmedi/pbaa016.

[86] R. Padmanabhan, H. S. Abed, N. Meskin, T. Khattab, M. Shraim, and M. A. Al-Hitmi. A review of mathematical model-based scenario analysis and interventions for COVID-19. Comput. Methods Programs Biomed., 209:106301, 2021. doi: 10.1016/j.cmpb.2021.106301.

[87] J. Hellewell, S. Abbott, A. Gimma, N. I. Bosse, C. I. Jarvis, T. W. Russell, et al. Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts. Lancet Glob. Health, 8(4):e488–e496, 2020. doi: 10.1016/S2214-109X(20)30074-7.

[88] Y. Zhao, J. Huang, L. Zhang, S. Chen, J. Gao, and H. Jiao. The global transmission of new coronavirus variants. Environ. Res., 206:112240, 2022. doi: 10.1016/j.envres.2021.112240.

[89] B. J. Willett, J. Grove, O. A. MacLean, C. Wilkie, G. De Lorenzo, W. Furnon, et al. SARS-CoV-2 Omicron is an immune escape variant with an altered cell entry pathway. Nat. Microbiol., 7(8):1161–1179, 2022. doi: 10.1038/s41564-022-01143-7.

[90] Y. Araf, F. Akter, Y.-d. Tang, R. Fatemi, M. S. A. Parvez, C. Zheng, et al. Omicron variant of SARS-CoV-2: Genomics, transmissibility, and responses to current COVID-19 vaccines. J. Med. Virol., 94(5):1825–1832, 2022. doi: 10.1002/jmv.27588.

[91] T. Nyberg, N. M. Ferguson, S. G. Nash, H. H. Webster, S. Flaxman, N. Andrews, et al. Comparative analysis of the risks of hospitalisation and death associated with SARS-CoV-2 omicron (B.1.1.529) and delta (B.1.617.2) variants in England: a cohort study. Lancet, 399(10332):1303–1312, 2022. doi: 10.1016/S0140-6736(22)00462-7.

[92] C. P. Bown and T. J. Bollyky. How COVID-19 vaccine supply chains emerged in the midst of a pandemic. World. Econ., 45(2):468–522, 2022. doi: 10.1111/twec.13183.

[93] R. Huang, X. Juan, Y. chen, C. Shan, and C. Wu. A family cluster of SARS-CoV-2 infection involving 11 patients in Nanjing, China. Lancet Infect. Dis., 20(5):534–535, 2020. doi: 10.1016/S1473-3099(20)30147-X.

[94] P. Yu, J. Zhu, Z. Zhang, and Y. Han. A Familial Cluster of Infection Associated With the 2019 Novel Coronavirus Indicating Possible Person-to-Person Transmission During the Incubation Period. J. Infect. Dis., 221(11):1757–1761, 2020. doi: 10.1093/infdis/jiaa077.

[95] Y.-C. Liu, C.-H. Liao, C.-F. Chang, C.-C. Chou, and Y.-R. Lin. A Locally Transmitted Case of SARS-CoV-2 Infection in Taiwan. N. Engl. J. Med., 382(11):1070–1072, 2020. doi: 10.1056/NEJMc2001573.

[96] C. Rothe, M. Schunk, P. Sothmann, G. Bretzel, G. Froeschl, C. Wallrauch, et al. Transmission of 2019-nCoV Infection from an Asymptomatic Contact in Germany. N. Engl. J. Med., 382(10):970–971, 2020. doi: 10.1056/NEJMc2001468.

[97] W. E. Wei, Z. Li, C. J. Chiew, S. E. Yong, M. P. Toh, and V. J. Lee. Presymptomatic Transmission of SARS-CoV-2 - Singapore, January 23-March 16, 2020. Morb. Mortal. Wkly. Rep., 69(14):411–415, 2020. doi: 10.15585/mmwr.mm6914e1.

[98] X. He, E. H. Y. Lau, P. Wu, X. Deng, J. Wang, X. Hao, et al. Temporal dynamics in viral shedding and transmissibility of COVID-19. Nat. Med., 26(5):672–675, 2020. doi: 10.1038/s41591-020-0869-5.

[99] Z.-D. Tong, A. Tang, K.-F. Li, P. Li, H.-L. Wang, J.-P. Yi, et al. Potential Presymptomatic Transmission of SARS-CoV-2, Zhejiang Province, China, 2020. Emerging Infect. Dis., 26(5):1052–1054, 2020. doi: 10.3201/eid2605.200198.

[100] G. Qian, N. Yang, A. H. Y. Ma, L. Wang, G. Li, X. Chen, et al. Covid-19 Transmission Within a Family Cluster by Presymptomatic Carriers in China. Clin. Infect. Dis., 71 (15):861–862, 2020. doi: 10.1093/cid/ciaa316.

[101] L. Zou, F. Ruan, M. Huang, L. Liang, H. Huang, Z. Hong, et al. SARS-CoV-2 Viral Load in Upper Respiratory Specimens of Infected Patients. N. Engl. J. Med., 382(12): 1177–1179, 2020. doi: 10.1056/NEJMc2001737.

[102] Y. Bai, L. Yao, T. Wei, F. Tian, D.-Y. Jin, L. Chen, et al. Presumed Asymptomatic Carrier Transmission of COVID-19. JAMA, 323(14):1406–1407, 2020. doi: 10.1001/jama.2020.2565.

[103] Z. Hu, C. Song, C. Xu, G. Jin, Y. Chen, X. Xu, et al. Clinical characteristics of 24 asymptomatic infections with COVID-19 screened among close contacts in Nanjing, China. Sci. China Life Sci., 63(5):706–711, 2020. doi: 10.1007/s11427-020-1661-4.

[104] C.-Q. Zhu, S.-D. Gao, Y. Xu, X.-H. Yang, F.-Q. Ye, L.-L. Ai, et al. A COVID-19 case report from asymptomatic contact: implication for contact isolation and incubation management. Infect. Dis. Poverty., 9(1):70, 2020. doi: 10.1186/s40249-020-00686-3.

[105] C. Qiu, Z. Deng, Q. Xiao, Y. Shu, Y. Deng, H. Wang, et al. Transmission and clinical characteristics of coronavirus disease 2019 in 104 outside-Wuhan patients, China. J. Med. Virol., 92(10):2027–2035, 2020. doi: 10.1002/jmv.25975.

[106] J. Zhang, M. Litvinova, Y. Liang, Y. Wang, W. Wang, S. Zhao, et al. Changes in contact patterns shape the dynamics of the COVID-19 outbreak in China. Science, 368(6498):1481–1486, 2020. doi: 10.1126/science.abb8001.

[107] N. G. Davies, P. Klepac, Y. Liu, K. Prem, M. Jit, C. A. B. Pearson, et al. Age-dependent effects in the transmission and control of COVID-19 epidemics. Nat. Med., 26(8):1205–1211, 2020. doi: 10.1038/s41591-020-0962-9.

[108] A. P. Munro and S. N. Faust. COVID-19 in children: current evidence and key questions. Curr. Opin. Infect. Dis., 33(6):540–547, 2020. doi: 10.1097/QCO.0000000000000690.

[109] R. M. Viner, O. T. Mytton, C. Bonell, G. J. Melendez-Torres, J. Ward, L. Hudson, et al. Susceptibility to SARS-CoV-2 Infection Among Children and Adolescents Compared With Adults: A Systematic Review and Meta-analysis. JAMA Pediatr., 175(2):143–156, 2021. doi: 10.1001/jamapediatrics.2020.4573.

[110] E. Goldstein, M. Lipsitch, and M. Cevik. On the effect of age on the transmission of SARS-CoV-2 in households, schools and the community. J. Infect. Dis., 223(3): 362–369, 2020. doi: 10.1093/infdis/jiaa691.

[111] J. F. Ludvigsson. Systematic review of COVID-19 in children shows milder cases and a better prognosis than adults. Acta Paediatr., 109(6):1088–1095, 2020. doi: 10.1111/apa.15270.

[112] Y. Dong, X. Mo, Y. Hu, X. Qi, F. Jiang, Z. Jiang, et al. Epidemiology of COVID-19 Among Children in China. Pediatrics, 145(6):e20200702, 2020. doi: 10.1542/peds.2020-0702.

[113] M. K. Siebach, G. Piedimonte, and S. H. Ley. COVID-19 in childhood: Transmission, clinical presentation, complications and risk factors. Pediatr. Pulmonol., 56(6):1342–1356, 2021. doi: 10.1002/ppul.25344.

[114] N. S. Mehta, O. T. Mytton, E. W. S. Mullins, T. A. Fowler, C. L. Falconer, O. B. Murphy, et al. SARS-CoV-2 (COVID-19): What Do We Know About Children? a Systematic Review. Clin. Infect. Dis., 71(9):2469–2479, 2020. doi: 10.1093/cid/ciaa556.

[115] M. S. Han, E. H. Choi, S. H. Chang, B.-L. Jin, E. J. Lee, B. N. Kim, et al. Clinical Characteristics and Viral RNA Detection in Children With Coronavirus Disease 2019 in the Republic of Korea. JAMA Pediatr., 175(1):73–80, 2021. doi: 10.1001/jamapediatrics.2020.3988.

[116] M. Lipsitch, T. Cohen, B. Cooper, J. M. Robins, S. Ma, L. James, et al. Transmission Dynamics and Control of Severe Acute Respiratory Syndrome. Science, 300(5627): 1966–1970, 2003. doi: 10.1126/science.1086616.

[117] R. M. Anderson, C. Fraser, A. C. Ghani, C. A. Donnelly, S. Riley, N. M. Ferguson, et al. Epidemiology, transmission dynamics and control of SARS: the 2002-2003 epidemic. Philos. Trans. R. Soc. Lond., B, Biol. Sci., 359(1447):1091–105, 2004. doi: 10.1098/rstb.2004.1490.

[118] World Health Organisation. Consensus document on the epidemiology of severe acute respiratory syndrome (SARS). Report, 2003. URL https://apps.who.int/iris/bitstream/handle/10665/70863/WHO_CDS_CSR_GAR_2003.11_eng.pdf?sequence=1&isAllowed=y.

[119] T. Day, A. Park, N. Madras, A. Gumel, and J. Wu. When is Quarantine a Useful Control Strategy for Emerging Infectious Diseases? Am. J. Epidemiol., 163(5):479–485, 2006. doi: 10.1093/aje/kwj056.

[120] W. J. Moss, D. E. Griffin, and W. H. Feinstone. Chapter 30 - Measles. In: Vaccines for Biodefense and Emerging and Neglected Diseases. Academic Press, London, 2009. doi: 10.1016/B978-0-12-369408-9.00030-5.

[121] World Health Organisation. Measles, Accessed 02 June 2023. URL https://www.who.int/news-room/fact-sheets/detail/measles.

[122] L. L. Lau, B. J. Cowling, V. J. Fang, K. H. Chan, E. H. Lau, M. Lipsitch, et al. Viral shedding and clinical illness in naturally acquired influenza virus infections. J. Infect. Dis., 201(10):1509–16, 2010. doi: 10.1086/652241.

[123] Y. Gu, N. Komiya, H. Kamiya, Y. Yasui, K. Taniguchi, and N. Okabe. Pandemic (H1N1) 2009 Transmission during Presymptomatic Phase, Japan. Emerging Infect. Dis., 17(9):1737–1739, 2011. doi: 10.3201/eid1709.101411.

[124] K. L. Roberts, H. Shelton, P. Stilwell, and W. S. Barclay. Transmission of a 2009 H1N1 Pandemic Influenza Virus Occurs before Fever is Detected, in the Ferret Model. PLoS ONE, 7(8):e43303, 2012. doi: 10.1371/journal.pone.0043303.

[125] D. K. M. Ip, L. L. H. Lau, N. H. L. Leung, V. J. Fang, K.-H. Chan, D. K. W. Chu, et al. Viral Shedding and Transmission Potential of Asymptomatic and Paucisymptomatic Influenza Virus Infections in the Community. Clin. Infect. Dis., 64(6):736–742, 2016. doi: 10.1093/cid/ciw841.

[126] C. Cohen, J. Kleynhans, J. Moyes, M. L. McMorrow, F. K. Treurnicht, O. Hellferscee, et al. Asymptomatic transmission and high community burden of seasonal influenza in an urban and a rural community in South Africa, 2017-18 (PHIRST): a population cohort study. Lancet Glob. Health, 9(6):e863–e874, 2021. doi: 10.1016/s2214-109x(21)00141-8.

[127] E. Patrozou and L. A. Mermel. Does influenza transmission occur from asymptomatic infection or prior to symptom onset? Public Health Rep., 124(2):193–6, 2009. doi: 10.1177/003335490912400205.

[128] A. Briggs, D. Jenkins, and C. Fraser. NHS Test and Trace: the journey so far. Report, The Health Foundation, 2020. URL https://www.health.org.uk/sites/default/files/2020-09/NHS%20Test%20and%20Trace.pdf.

[129] G. Iacobucci and R. Coombes. Covid-19: Government plans to spend £100bn on expanding testing to 10 million a day. BMJ, 370:3520, 2020. doi: 10.1136/bmj.m3520.

[130] G. Iacobucci. Covid-19: Government rolls out twice weekly rapid testing to all in England. BMJ, 373:902, 2021. doi: 10.1136/bmj.n902.

[131] UK Government. Twice weekly rapid testing to be available to everyone in England, 2021. URL https://www.gov.uk/government/news/twice-weekly-rapid-testing-to-be-available-to-everyone-in-england.

[132] Minutes of the fifty-sixth SAGE meeting on COVID-19, 10th September 2020. Report. URL https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/928699/S0740_Fifty-sixth_SAGE_meeting_on_Covid-19.pdf.

[133] L. Matthews, D. T. Haydon, D. J. Shaw, M. E. Chase-Topping, M. J. Keeling, and M. E. J. Woolhouse. Neighbourhood control policies and the spread of infectious diseases. Proc. Royal Soc. B, 270(1525):1659–1666, 2003. doi: 10.1098/rspb.2003.2429.

[134] C.-C. Lai, T.-P. Shih, W.-C. Ko, H.-J. Tang, and P.-R. Hsueh. Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and coronavirus disease-2019 (COVID-19): The epidemic and the challenges. Int. J. Antimicrob. Agents, 55(3):105924, 2020. doi: 10.1016/j.ijantimicag.2020.105924.

[135] Y. Liu, A. A. Gayle, A. Wilder-Smith, and J. Rocklov. The reproductive number of COVID-19 is higher compared to SARS coronavirus. J. Travel Med., 27(2):taaa021, 2020. doi: 10.1093/jtm/taaa021.

[136] P. Zhai, Y. Ding, X. Wu, J. Long, Y. Zhong, and Y. Li. The epidemiology, diagnosis and treatment of COVID-19. Int. J. Antimicrob. Agents, 55(5):105955, 2020. doi: 10.1016/j.ijantimicag.2020.105955.

[137] S. Zhao, Q. Lin, J. Ran, S. S. Musa, G. Yang, W. Wang, et al. Preliminary estimation of the basic reproduction number of novel coronavirus (2019-nCoV) in China, from 2019 to 2020: A data-driven analysis in the early phase of the outbreak. Int. J. Infect. Dis., 92:214–217, 2020. doi: 10.1016/j.ijid.2020.01.050.

[138] D. Buitrago-Garcia, D. Egli-Gany, M. J. Counotte, S. Hossman, H. Imeri, A. M. Ipekci, et al. Occurrence and transmission potential of asymptomatic and presymptomatic SARS-CoV-2 infections: A living systematic review and meta-analysis. PLoS Med., 17(9):e1003346, 2020. doi: 10.1371/journal.pmed.1003346.

[139] O. Byambasuren, M. Cardona, K. Bell, J. Clark, M.-L. McLaws, and P. Glasziou. Estimating the extent of asymptomatic COVID-19 and its potential for community transmission: Systematic review and meta-analysis. Can. J. Infect. Dis. Med. Microbiol., 5 (4):223–234, 2020. doi: 10.3138/jammi-2020-0030.

[140] J. He, Y. Guo, R. Mao, and J. Zhang. Proportion of asymptomatic coronavirus disease 2019: A systematic review and meta-analysis. J. Med. Virol., 93(2):820–830, 2020. doi: 10.1002/jmv.26326.

[141] L. Ferretti, C. Wymant, M. Kendall, L. Zhao, A. Nurtay, L. Abeler-Dorner, et al. Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. Science, 368(6491):eabb6936, 2020. doi: 10.1126/science.abb6936.

[142] Q. Li, X. Guan, P. Wu, X. Wang, L. Zhou, Y. Tong, et al. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus–Infected Pneumonia. N. Engl. J. Med., 382(13):1199–1207, 2020. doi: 10.1056/NEJMoa2001316.

[143] M. Casey-Bryars, J. Griffin, C. McAloon, A. Byrne, J. Madden, D. McEvoy, et al. Presymptomatic transmission of SARS-CoV-2 infection: a secondary analysis using published data. BMJ Open, 11(6):e041240, 2021. doi: 10.1136/bmjopen-2020-041240.

[144] M. M. Arons, K. M. Hatfield, S. C. Reddy, A. Kimball, A. James, J. R. Jacobs, et al. Presymptomatic SARS-CoV-2 Infections and Transmission in a Skilled Nursing Facility. N. Engl. J. Med., 382(22):2081–2090, 2020. doi: 10.1056/NEJMoa2008457.

[145] J. Bullard, K. Dust, D. Funk, J. E. Strong, D. Alexander, L. Garnett, et al. Predicting Infectious Severe Acute Respiratory Syndrome Coronavirus 2 From Diagnostic Samples. Clin. Infect. Dis., 71(10):2663–2666, 2020. doi: 10.1093/cid/ciaa638.

[146] R. Wolfel, V. M. Corman, W. Guggemos, M. Seilmaier, S. Zange, M. A. Muller, et al. Virological assessment of hospitalized patients with COVID-2019. Nature, 581(7809): 465–469, 2020. doi: 10.1038/s41586-020-2196-x.

[147] C. L. Althaus, N. Low, E. O. Musa, F. Shuaib, and S. Gsteiger. Ebola virus disease outbreak in Nigeria: Transmission dynamics and rapid control. Epidemics, 11:80–84, 2015. doi: 10.1016/j.epidem.2015.03.001.

[148] P. Boldog, T. Tekeli, Z. Vizi, A. Denes, F. A. Bartha, and G. Rost. Risk Assessment of Novel Coronavirus COVID-19 Outbreaks Outside China. J. Clin. Med., 9(2):571, 2020. doi: 10.3390/jcm9020571.

[149] Y. Daon, R. N. Thompson, and U. Obolski. Estimating COVID-19 outbreak risk through air travel. J. Travel Med., 27(5):taaa093, 2020. doi: 10.1093/jtm/taaa093.

[150] A. A. Lashari and P. Trapman. Branching process approach for epidemics in dynamic partnership network. J. Math. Biol., 76(1):265–294, 2018. doi: 10.1007/s00285-017-1147-0.

[151] R. N. Thompson, C. P. Thompson, O. Pelerman, S. Gupta, and U. Obolski. Increased frequency of travel in the presence of cross-immunity may act to decrease the chance of a global pandemic. Philos. Trans. R. Soc. Lond., B, Biol. Sci., 374(1775):20180274, 2019. doi: 10.1098/rstb.2018.0274.

[152] R. N. Thompson, C. A. Gilligan, and N. J. Cunniffe. Detecting Presymptomatic Infection Is Necessary to Forecast Major Epidemics in the Earliest Stages of Infectious Disease Outbreaks. PLoS Comput. Biol., 12(4):e1004836, 2016. doi: 10.1371/journal.pcbi.1004836.

[153] B. Charleston, B. M. Bankowski, S. Gubbins, M. E. Chase-Topping, D. Schley, R. Howey, et al. Relationship Between Clinical Signs and Transmission of an Infectious Disease and the Implications for Control. Science, 332(6030):726–729, 2011. doi: 10.1126/science.1199884.

[154] L. Martinez. Global infectious disease surveillance. Int. J. Infect. Dis., 4(4):222–228, 2000. doi: 10.1016/s1201-9712(00)90114-0.

[155] G. J. Milinovich, G. M. Williams, A. C. A. Clements, and W. Hu. Internet-based surveillance systems for monitoring emerging infectious diseases. Lancet Infect. Dis., 14(2):160–168, 2014. doi: 10.1016/S1473-3099(13)70244-5.

[156] Q. A. ten Bosch, H. E. Clapham, L. Lambrechts, V. Duong, P. Buchy, B. M. Althouse, et al. Contributions from the silent majority dominate dengue virus transmission. PLoS Pathog., 14(5):e1006965, 2018. doi: 10.1371/journal.ppat.1006965.

[157] S. M. Moghadas, M. C. Fitzpatrick, P. Sah, A. Pandey, A. Shoukat, B. H. Singer, et al. The implications of silent transmission for the control of COVID-19 outbreaks. Proc. Natl. Acad. Sci. U.S.A., 117(30):17513–17515, 2020. doi: 10.1073/pnas.2008373117.

[158] X. Ren, Y. Li, X. Yang, Z. Li, J. Cui, A. Zhu, et al. Evidence for pre-symptomatic transmission of coronavirus disease 2019 (COVID-19) in China. Influenza and Other Respiratory Viruses, 15(1):19–26, 2021. doi: 10.1111/irv.12787.

[159] M. Yanes-Lane, N. Winters, F. Fregonese, M. Bastos, S. Perlman-Arrow, J. R. Campbell, et al. Proportion of asymptomatic infection among COVID-19 positive persons and their transmission potential: A systematic review and meta-analysis. PLoS ONE, 15(11):e0241536, 2020. doi: 10.1371/journal.pone.0241536.

[160] E. A. Meyerowitz, A. Richterman, I. I. Bogoch, N. Low, and M. Cevik. Towards an accurate and systematic characterisation of persistently asymptomatic infection with SARS-CoV-2. Lancet Infect. Dis., 21(6):e163–e169, 2021. doi: 10.1016/S1473-3099(20)30837-9.

[161] D. P. Oran and E. J. Topol. Prevalence of Asymptomatic SARS-CoV-2 Infection: A Narrative Review. Ann. Intern. Med., 173(5):362–367, 2020. doi: 10.7326/M20-3012.

[162] D. Klinkenberg, C. Fraser, and H. Heesterbeek. The Effectiveness of Contact Tracing in Emerging Epidemics. PLoS ONE, 1(1):e12, 2006. doi: 10.1371/journal.pone.0000012.

[163] G. Webb, C. Browne, X. Huo, O. Seydi, M. Seydi, and P. Magal. A Model of the 2014 Ebola Epidemic in West Africa with Contact Tracing. PLoS Curr., 7, 2015. doi: 10.1371/currents.outbreaks.846b2a31ef37018b7d1126a9c8adf22a.

[164] R. N. Thompson. Epidemiological models are important tools for guiding COVID-19 interventions. BMC Med., 18(1):152, 2020. doi: 10.1186/s12916-020-01628-4.

[165] H. Nishiura and K. Kamiya. Fever screening during the influenza (H1N1-2009) pandemic at Narita International Airport, Japan. BMC Infect. Dis., 11(1):111, 2011. doi: 10.1186/1471-2334-11-111.

[166] M. Chinazzi, J. T. Davis, M. Ajelli, C. Gioannini, M. Litvinova, S. Merler, et al. The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. Science, 368(6489):395–400, 2020. doi: 10.1126/science.aba9757.

[167] B. L. Dickens, J. R. Koo, J. T. Lim, H. Sun, H. E. Clapham, A. Wilder-Smith, et al. Strategies at points of entry to reduce importation risk of COVID-19 cases and reopen travel. J. Travel Med., 27(8):taaa141, 2020. doi: 10.1093/jtm/taaa141.

[168] K. Prem, A. R. Cook, and M. Jit. Projecting social contact matrices in 152 countries using contact surveys and demographic data. PLoS Comput. Biol., 13(9):e1005697, 2017. doi: 10.1371/journal.pcbi.1005697.

[169] S.-J. Kang and S. I. Jung. Age-Related Morbidity and Mortality among Patients with COVID-19. J. Infect. Chemother., 52(2):154–164, 2020. doi: 10.3947/ic.2020.52.2.154.

[170] A. Endo, Centre for the Mathematical Modelling of Infectious Diseases COVID-19 Working Group, S. Abbott, A. Kucharski, and S. Funk. Estimating the overdispersion in COVID-19 transmission using outbreak sizes outside China [version 3; peer review: 2 approved]. Wellcome Open Res., 5(67), 2020. doi: 10.12688/wellcomeopenres.15842.3.

[171] M. S. Y. Lau, B. Grenfell, M. Thomas, M. Bryan, K. Nelson, and B. Lopman. Characterizing superspreading events and age-specific infectiousness of SARS-CoV-2 transmission in Georgia, USA. Proc. Natl. Acad. Sci. U.S.A., 117(36):22430–22435, 2020. doi: 10.1073/pnas.2011802117.

[172] M. E. J. Woolhouse, C. Dye, J.-F. Etard, T. Smith, J. D. Charlwood, G. P. Garnett, et al. Heterogeneities in the tranmission of infectious agents: Implications for the design of control programs. Proc. Natl. Acad. Sci. U.S.A., 94(1):338–342, 1997. doi: 10.1073/pnas.94.1.338.

[173] R. N. Thompson, T. D. Hollingsworth, V. Isham, D. Arribas-Bel, B. Ashby, T. Britton, et al. Key questions for modelling COVID-19 exit strategies. Proc. Royal Soc. B, 287 (1932):20201405, 2020. doi: 10.1098/rspb.2020.1405.

[174] J. O. Lloyd-Smith, S. J. Schreiber, P. E. Kopp, and W. M. Getz. Superspreading and the effect of individual variation on disease emergence. Nature, 438(17):355–359, 2005. doi: 10.1038/nature04153.

[175] N. Perra. Non-pharmaceutical interventions during the COVID-19 pandemic: A review. Phys. Rep., 913:1–52, 2021. doi: 10.1016/j.physrep.2021.02.001.

[176] K. Regmi and C. M. Lwin. Impact of non-pharmaceutical interventions for reducing transmission of COVID-19: a systematic review and meta-analysis protocol. BMJ Open, 10(10):e041383, 2020. doi: 10.1136/bmjopen-2020-041383.

[177] N. Imai, K. A. M. Gaythorpe, S. Abbott, S. Bhatia, S. van Elsland, K. Prem, et al. Adoption and impact of non-pharmaceutical interventions for COVID-19. Wellcome Open Res., 5:59, 2020. doi: 10.12688/wellcomeopenres.15808.1.

[178] R. J. Hatchett, C. E. Mecher, and M. Lipsitch. Public health interventions and epidemic intensity during the 1918 influenza pandemic. Proc. Natl. Acad. Sci. U.S.A., 104(18):7582, 2007. doi: 10.1073/pnas.0610941104.

[179] T. D. Kirsch, H. Moseson, M. Massaquoi, T. G. Nyenswah, R. Goodermote, I. Rodriguez-Barraquer, et al. Impact of interventions and the incidence of Ebola Virus Disease in Liberia—implications for future epidemics. Health Policy Plan., 32 (2):205–214, 2017. doi: 10.1093/heapol/czw113.

[180] B. E. Berkman. Mitigating Pandemic Influenza: The Ethics of Implementing a School Closure Policy. J. Public Health Manag. Pract., 14(4):372–378, 2008. doi: 10.1097/01.PHH.0000324566.72533.0b.

[181] S. Moore, E. M. Hill, M. J. Tildesley, L. Dyson, and M. J. Keeling. Vaccination and non-pharmaceutical interventions for COVID-19: a mathematical modelling study. Lancet Infect. Dis., 21(6):793–802, 2021. doi: 10.1016/S1473-3099(21)00143-2.

[182] M. A. Spinelli, D. V. Glidden, E. D. Gennatas, M. Bielecki, C. Beyrer, G. Rutherford, et al. Importance of non-pharmaceutical interventions in lowering the viral inoculum to reduce susceptibility to infection by SARS-CoV-2 and potentially disease severity. Lancet Infect. Dis., 21(9):e296–e301, 2021. doi: 10.1016/S1473-3099(20)30982-8.

[183] R. N. Thompson, E. M. Hill, and J. R. Gog. SARS-CoV-2 incidence and vaccine escape. Lancet Infect. Dis., 21(7):913–914, 2021. doi: 10.1016/S1473-3099(21)00202-4.

[184] J. M. Donohue and E. Miller. COVID-19 and School Closures. JAMA, 324(9):845–847, 2020. doi: 10.1001/jama.2020.13092.

[185] M. Silverman, R. Sibbald, and S. Stranges. Ethics of COVID-19-related school closures. Can. J. Public Health, 111(4):462–465, 2020. doi: 10.17269/s41997-020-00396-1.

[186] M. Z. Sadique, E. J. Adams, and W. J. Edmunds. Estimating the costs of school closure for mitigating an influenza pandemic. BMC Public Health, 8(1):135, 2008. doi: 10.1186/1471-2458-8-135.

[187] S. K. Brooks, R. K. Webster, L. E. Smith, L. Woodland, S. Wessely, N. Greenberg, et al. The psychological impact of quarantine and how to reduce it: rapid review of the evidence. Lancet, 395(10227):912–920, 2020. doi: 10.1016/S0140-6736(20)30460-8.

[188] M. E. Loades, E. Chatburn, N. Higson-Sweeney, S. Reynolds, R. Shafran, A. Brigden, et al. Rapid Systematic Review: The Impact of Social Isolation and Loneliness on the Mental Health of Children and Adolescents in the Context of COVID-19. J. Am. Acad. Child Adolesc. Psychiatry, 59(11):1218–1239.e3, 2020. doi: 10.1016/j.jaac.2020.05.009.

[189] J. Mossong, N. Hens, M. Jit, P. Beutels, K. Auranen, R. Mikolajczyk, et al. Social Contacts and Mixing Patterns Relevant to the Spread of Infectious Diseases. PLoS Med., 5(3):e74, 2008. doi: 10.1371/journal.pmed.0050074.

[190] L. Danon, J. M. Read, T. A. House, M. C. Vernon, and M. J. Keeling. Social encounter networks: characterizing Great Britain. Proc. Royal Soc. B, 280(1765):20131037, 2013. doi: 10.1098/rspb.2013.1037.

[191] K. Leung, M. Jit, E. H. Y. Lau, and J. T. Wu. Social contact patterns relevant to the spread of respiratory infectious diseases in Hong Kong. Sci. Rep., 7(7974), 2017. doi: 10.1038/s41598-017-08241-1.

[192] G. Béraud, S. Kazmercziak, P. Beutels, D. Levy-Bruhl, X. Lenne, N. Mielcarek, et al. The French Connection: The First Large Population-Based Contact Survey in France Relevant for the Spread of Infectious Diseases. PLoS ONE, 10(7):e0133203, 2015. doi: 10.1371/journal.pone.0133203.

[193] Y. Ibuka, Y. Ohkusa, T. Sugawara, G. B. Chapman, D. Yamin, K. E. Atkins, et al. Social contacts, vaccination decisions and influenza in Japan. J. Epidemiol. Comm. Health, 70(2):162–167, 2016. doi: 10.1136/jech-2015-205777.

[194] J. M. Read, J. Lessler, S. Riley, S. Wang, L. J. Tan, K. O. Kwok, et al. Social mixing patterns in rural and urban areas of southern China. Proc. Royal Soc. B, 281(1785): 20140268, 2014. doi: 10.1098/rspb.2014.0268.

[195] J. M. Read, W. J. Edmunds, S. Riley, J. Lessler, and D. A. T. Cummings. Close encounters of the infectious kind: methods to measure social mixing behaviour. Epidemiol. Infect., 140(12):2117–2130, 2012. doi: 10.1017/S0950268812000842.

[196] J. Wallinga, P. Teunis, and M. Kretzschmar. Using Data on Social Contacts to Estimate Age-specific Transmission Parameters for Respiratory-spread Infectious Agents. Am. J. Epidemiol., 164(10):936–944, 2006. doi: 10.1093/aje/kwj317.

[197] A. J. Kucharski, K. O. Kwok, V. W. I. Wei, B. J. Cowling, J. M. Read, J. Lessler, et al. The Contribution of Social Behaviour to the Transmission of Influenza A in a Human Population. PLoS Pathog., 10(6):e1004206, 2014. doi: 10.1371/journal.ppat.1004206.

[198] A. M. Pollock and J. Lancaster. Asymptomatic transmission of COVID-19. BMJ, 371: m4851, 2020. doi: 10.1136/bmj.m4851.

[199] W. C. Koh, L. Naing, L. Chaw, M. A. Rosledzana, M. F. Alikhan, S. A. Jamaludin, et al. What do we know about SARS-CoV-2 transmission? A systematic review and meta-analysis of the secondary attack rate and associated risk factors. PLoS ONE, 15 (10):e0240205, 2020. doi: 10.1371/journal.pone.0240205.

[200] Z. J. Madewell, Y. Yang, Jr. Longini, I. M., M. E. Halloran, and N. E. Dean. Household Transmission of SARS-CoV-2: A Systematic Review and Meta-analysis. JAMA Netw. Open, 3(12):e2031756, 2020. doi: 10.1001/jamanetworkopen.2020.31756.

[201] X. Qiu, A. I. Nergiz, A. E. Maraolo, I. I. Bogoch, N. Low, and M. Cevik. Defining the role of asymptomatic and pre-symptomatic SARS-CoV-2 transmission - a living systematic review. Clin. Microbiol. Infect., 27(4):511–519, 2021. doi: 10.1016/j.cmi.2021.01.011.

[202] B. G. Pijls, S. Jolani, A. Atherley, R. T. Derckx, J. I. R. Dijkstra, G. H. L. Franssen, et al. Demographic risk factors for COVID-19 infection, severity, ICU admission and death: a meta-analysis of 59 studies. BMJ Open, 11(1):e044640, 2021. doi: 10.1136/bmjopen-2020-044640.

[203] R. Verity, L. C. Okell, I. Dorigatti, P. Winskill, C. Whittaker, N. Imai, et al. Estimates of the severity of coronavirus disease 2019: a model-based analysis. Lancet Infect. Dis., 20(6):669–677, 2020. doi: 10.1016/S1473-3099(20)30243-7.

[204] T. Chen, Z. Dai, P. Mo, X. Li, Z. Ma, S. Song, et al. Clinical Characteristics and Outcomes of Older Patients with Coronavirus Disease 2019 (COVID-19) in Wuhan, China: A Single-Centered, Retrospective Study. J. Gerontol. A Biol. Sci. Med. Sci., 75(9):1788–1795, 2020. doi: 10.1093/gerona/glaa089.

[205] T. Liu, D. Gong, J. Xiao, J. Hu, G. He, Z. Rong, et al. Cluster infections play important roles in the rapid evolution of COVID-19 transmission: A systematic review. Int. J. Infect. Dis., 99:374–380, 2020. doi: 10.1016/j.ijid.2020.07.073.

[206] Q. Leclerc, N. Fuller, L. Knight, CMMID COVID-19 Working Group, S. Funk, and G. Knight. What settings have been linked to SARS-CoV-2 transmission clusters? [version 2; peer review: 2 approved]. Wellcome Open Res., 5(83), 2020. doi: 10.12688/wellcomeopenres.15889.2.

[207] H. Rossman, A. Keshet, S. Shilo, A. Gavrieli, T. Bauman, O. Cohen, et al. A framework for identifying regional outbreak and spread of COVID-19 from one-minute population-wide surveys. Nat. Med., 26(5):634–638, 2020. doi: 10.1038/s41591-020-0857-9.

[208] R. N. Thompson, K. Jalava, and U. Obolski. Sustained transmission of Ebola in new locations: more likely than previously thought. Lancet Infect. Dis., 19(10):1058–1059, 2019. doi: 10.1016/S1473-3099(19)30483-9.

[209] H. Nishiura, A. R. Cook, and B. J. Cowling. Assortativity and the Probability of Epidemic Extinction: A Case Study of Pandemic Influenza A (H1N1-2009). Interdiscip. Perspect. Infect. Dis., 2011:194507, 2011. doi: 10.1155/2011/194507.

[210] United Nations Department of Economic and Social Affairs (Population Division). World Population Prospects 2019, Online Edition, Rev. 1, 2019. URL `https://population.un.org/wpp/Download/Standard/Population/`.

[211] W. S. Hart, L. F. R. Hochfilzer, N. J. Cunniffe, H. Lee, H. Nishiura, and R. N. Thompson. Accurate forecasts of the effectiveness of interventions against Ebola may require models that account for variations in symptoms during infection. Epidemics, 29:100371, 2019. doi: 10.1016/j.epidem.2019.100371.

[212] Q. S. Huang, T. Wood, L. Jelley, T. Jennings, S. Jefferies, K. Daniells, et al. Impact of the COVID-19 nonpharmaceutical interventions on influenza and other respiratory viral infections in New Zealand. Nat. Commun., 12(1):1001, 2021. doi: 10.1038/s41467-021-21157-9.

[213] S. Flaxman, S. Mishra, A. Gandy, H. J. T. Unwin, T. A. Mellan, H. Coupland, et al. Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. Nature, 584(7820):257–261, 2020. doi: 10.1038/s41586-020-2405-7.

[214] R. H. Borse, C. B. Behravesh, T. Dumanovsky, J. R. Zucker, D. Swerdlow, P. Edelson, et al. Closing Schools in Response to the 2009 Pandemic Influenza A H1N1 Virus in New York City: Economic Impact on Households. Clin. Infect. Dis., 52(Suppl. 1): S168–S172, 2011. doi: 10.1093/cid/ciq033.

[215] A. L. Greer, A. Tuite, and D. N. Fisman. Age, influenza pandemics and disease dynamics. Epidemiol. Infect., 138(11):1542–1549, 2010. doi: 10.1017/S0950268810000579.

[216] S. Clohisey and J. K. Baillie. Host susceptibility to severe influenza A virus infection. Crit. Care., 23(1):303, 2019. doi: 10.1186/s13054-019-2566-7.

[217] G. N. Agrios. Plant Pathology. Academic Press, 5th edition, 2005.

[218] N. J. Cunniffe, R. C. Cobb, R. K. Meentemeyer, D. M. Rizzo, and C. A. Gilligan. Modeling when, where, and how to manage a forest epidemic, motivated by sudden oak death in California. Proc. Natl. Acad. Sci. U.S.A., 113(20):5640–5645, 2016. doi: 10.1073/pnas.1602153113.

[219] B. M. Dala-Paula, A. Plotto, J. Bai, J. A. Manthey, E. A. Baldwin, R. S. Ferrarezi, et al. Effect of Huanglongbing or Greening Disease on Orange Juice Quality, a Review. Front. Plant Sci., 9(1976), 2019. doi: 10.3389/fpls.2018.01976.

[220] A. K. Das. Citrus canker - A review. J. Appl. Hortic., 5(1):52–60, 2003. doi: 10.37855/jah.2003.v05i01.15.

[221] D.-c. He, J.-s. Zhan, and L.-h. Xie. Problems, challenges and future of plant disease management: from an ecological point of view. J. Integr. Agric., 15(4):705–715, 2016. doi: 10.1016/S2095-3119(15)61300-4.

[222] R. C. Ploetz. Panama disease: Return of the first banana menace. Int. J. Pest Manag., 40(4):326–336, 1994. doi: 10.1080/09670879409371908.

[223] K. Schneider, W. van der Werf, M. Cendoya, M. Mourits, J. A. Navas-Cortes, A. Vicent, et al. Impact of *Xylella fastidiosa* subspecies *pauca* in European olives. Proc. Natl. Acad. Sci. U.S.A., 117(17):9250–9259, 2020. doi: 10.1073/pnas.1912206117.

[224] J. Ristaino, K. Anderson, D. P. Bebber, K. A. Brauman, N. J. Cunniffe, N. V. Fedoroff, et al. The Persistent Threat of Emerging Plant Disease Pandemics to Global Food Security. Proc. Natl. Acad. Sci. U.S.A., 118(23):e2022239118, 2021. doi: 10.1073/pnas.2022239118.

[225] I. L. Boyd, P. H. Freer-Smith, C. A. Gilligan, and H. C. J. Godfray. The Consequence of Tree Pests and Diseases for Ecosystem Services. Science, 342(6160):1235773, 2013. doi: 10.1126/science.1235773.

[226] H. N. Fones, D. P. Bebber, T. M. Chaloner, W. T. Kay, G. Steinberg, and S. J. Gurr. Threats to global food security from emerging fungal and oomycete crop pathogens. Nat. Food, 1(6):332–342, 2020. doi: 10.1038/s43016-020-0075-0.

[227] H. C. Godfray, D. Mason-D'Croz, and S. Robinson. Food system consequences of a fungal disease epidemic in a major crop. Philos. Trans. R. Soc. Lond., B, Biol. Sci., 371(1709):20150467, 2016. doi: 10.1098/rstb.2015.0467.

[228] S. Savary, L. Willocquet, S. J. Pethybridge, P. Esker, N. McRoberts, and A. Nelson. The global burden of pathogens and pests on major food crops. Nat. Ecol. Evol., 3(3): 430–439, 2019. doi: 10.1038/s41559-018-0793-y.

[229] M. Carvajal-Yepes, K. Cardwell, A. Nelson, K. A. Garrett, B. Giovani, D. G. O. Saunders, et al. A global surveillance system for crop diseases. Science, 364(6447): 1237–1239, 2019. doi: 10.1126/science.aaw1572.

[230] K. Yoshida, V. J. Schuenemann, L. M. Cano, M. Pais, B. Mishra, R. Sharma, et al. The rise and fall of the *Phytophthora infestans* lineage that triggered the Irish potato famine. eLife, 2:e00731, 2013. doi: 10.7554/eLife.00731.

[231] O. J. Alabi, P. L. Kumar, and R. A. Naidu. Cassava Mosaic Disease: A Curse to Food Security in Sub-Saharan Africa. APSnet Features, 2011. doi: 10.1094/APSnetFeature-2011-0701.

[232] A. Singerman and M. E. Rogers. The Economic Challenges of Dealing with Citrus Greening: The Case of Florida. J. Integr. Pest Manag., 11(1):3, 2020. doi: 10.1093/jipm/pmz037.

[233] D. Farnsworth, K. A. Grogan, A. H. C. van Bruggen, and C. B. Moss. The Potential Economic Cost and Response to Greening in Florida Citrus. Choices, 29(3), 2014. URL https://www.choicesmagazine.org/UserFiles/file/cmsarticle_393.pdf.

[234] S. Alvarez, E. Rohrig, D. Solís, and M. H. Thomas. Citrus Greening Disease (Huanglongbing) in Florida: Economic Impact, Management and the Potential for Biological Control. Agric. Res., 5(2):109–118, 2016. doi: 10.1007/s40003-016-0204-z.

[235] G. S. Laha, R. Singh, D. Ladhalakshmi, S. Sunder, M. S. Prasad, C. S. Dagar, et al. Importance and Management of Rice Diseases: A Global Perspective. In: Rice Production Worldwide. Springer International Publishing, 2017. doi: 10.1007/978-3-319-47516-5_13.

[236] M. Figueroa, K. E. Hammond-Kosack, and P. S. Solomon. A review of wheat diseases - a field perspective. Mol. Plant Pathol., 19(6):1523–1536, 2018. doi: 10.1111/mpp.12618.

[237] A. Y. Bandara, D. K. Weerasooriya, C. A. Bradley, T. W. Allen, and P. D. Esker. Dissecting the economic impact of soybean diseases in the United States over two decades. PLoS ONE, 15(4):e0231141, 2020. doi: 10.1371/journal.pone.0231141.

[238] R. S. García-Estrada, A. Diaz-Lara, V. H. Aguilar-Molina, and J. M. Tovar-Pedraza. Viruses of Economic Impact on Tomato Crops in Mexico: From Diagnosis to Management - A Review. Viruses, 14(6):1251, 2022. doi: 10.3390/v14061251.

[239] G. Lu, Z. Wang, F. Xu, Y. B. Pan, M. P. Grisham, and L. Xu. Sugarcane Mosaic Disease: Characteristics, Identification and Control. Microorganisms, 9(9):1984, 2021. doi: 10.3390/microorganisms9091984.

[240] K. Tumber, J. Alston, and K. Fuller. Pierce's disease costs California $104 million per year. Calif. Agric., 68(1):20–29, 2014. doi: 10.3733/ca.v068n01p20.

[241] D. C. Cook, A. S. Taylor, R. A. Meldrum, and A. Drenth. Potential economic impact of Panama disease (tropical race 4) on the Australian banana industry. J. Plant. Dis. Prot., 122(5/6):229–237, 2015. doi: 10.1007/BF03356557.

[242] A. Ficke, C. Cowger, G. Bergstrom, and G. Brodal. Understanding Yield Loss and Pathogen Biology to Improve Disease Management: Septoria Nodorum Blotch - A Case Study in Wheat. Plant Dis., 102(4):696–707, 2018. doi: 10.1094/pdis-09-17-1375-fe.

[243] Editorial. Pathogens, precipitation and produce prices. Nat. Clim. Chang., 11(8):635–635, 2021. doi: 10.1038/s41558-021-01124-4.

[244] S. Parnell, F. van den Bosch, T. Gottwald, and C. A. Gilligan. Surveillance to Inform Control of Emerging Plant Diseases: An Epidemiological Perspective. Annu. Rev. Phytopathol., 55:591–610, 2017. doi: 10.1146/annurev-phyto-080516-035334.

[245] R. N. Thompson, R. C. Cobb, C. A. Gilligan, and N. J. Cunniffe. Management of invading pathogens should be informed by epidemiology rather than administrative boundaries. Ecol. Modell., 324:28–32, 2016. doi: 10.1016/j.ecolmodel.2015.12.014.

[246] A. P. Craig, N. J. Cunniffe, M. Parry, F. F. Laranjeira, and C. A. Gilligan. Grower and regulator conflict in management of the citrus disease Huanglongbing in Brazil: a modelling study. J. Appl. Ecol., 55(4):1956–1965, 2018. doi: 10.1111/1365-2664.13122.

[247] F. Hilker, A. Allen, V. Bokil, C. Briggs, Z. Feng, K. A. Garrett, et al. Modelling virus coinfection to inform management of maize lethal necrosis in Kenya. Phytopathology, 107(10):1095–1108, 2017. doi: 10.1094/PHYTO-03-17-0080-FI.

[248] S. R. Hyatt-Twynam, S. Parnell, R. O. J. H. Stutt, T. R. Gottwald, C. A. Gilligan, and N. J. Cunniffe. Risk-based management of invading plant disease. New Phytol., 214(3):1317–1329, 2017. doi: 10.1111/nph.14488.

[249] H. K. Adrakey, G. Streftaris, N. J. Cunniffe, T. R. Gottwald, C. A. Gilligan, and G. J. Gibson. Evidence-based controls for epidemics using spatio-temporal stochastic models in a Bayesian framework. J. R. Soc. Interface, 14(136):20170386, 2017. doi: 10.1098/rsif.2017.0386.

[250] D. S. Chapman, S. M. White, D. A. P. Hooftman, and J. M. Bullock. Inventory and review of quantitative models for spread of plant pests for use in pest risk assessment for the EU territory. EFSA Supporting Publication, 12(4), 2015. doi: 10.2903/sp.efsa.2015.EN-795.

[251] N. J. Cunniffe and C. A. Gilligan. Using mathematical models to predict epidemics and to optimise disease detection and management. In: Emerging Plant Diseases and Global Food Security. APS Press, 2020.

[252] F. Fabre, J. Coville, and N. J. Cunniffe. Optimising reactive disease management using spatially explicit models at the landscape scale. In: Plant Diseases and Food Security in the 21st Century. Springer, 2021.

[253] R. S. Epanchin-Niell, R. G. Haight, L. Berec, J. M. Kean, and A. M. Liebhold. Optimal surveillance and eradication of invasive species in heterogeneous landscapes. Ecol. Lett., 15(8):803–812, 2012. doi: 10.1111/j.1461-0248.2012.01800.x.

[254] E. H. Bussell and N. J. Cunniffe. Applying optimal control theory to a spatial simulation model of sudden oak death: ongoing surveillance protects tanoak whilst conserving biodiversity. J. R. Soc. Interface, 17(165):20190671, 2020. doi: 10.1098/rsif.2019.0671.

[255] E. H. Bussell and N. J. Cunniffe. Optimal strategies to protect a sub-population at risk due to an established epidemic. J. R. Soc. Interface, 19(186):20210718, 2022. doi: 10.1098/rsif.2021.0718.

[256] E. H. Bussell, C. Dangerfield, C. A. Gilligan, and N. J. Cunniffe. Applying optimal control theory to complex epidemiological models to inform real-world disease management. Philos. Trans. R. Soc. Lond., B, Biol. Sci., 374(1776):20180284, 2019. doi: 10.1098/rstb.2018.0284.

[257] S. Parnell, T. R. Gottwald, F. van den Bosch, and C. A. Gilligan. Optimal Strategies for the Eradication of Asiatic Citrus Canker in Heterogeneous Host Landscapes. Phytopathology, 99(12):1370–1376, 2009. doi: 10.1094/phyto-99-12-1370.

[258] S. Parnell, T. R. Gottwald, C. A. Gilligan, N. J. Cunniffe, and F. van den Bosch. The Effect of Landscape Pattern on the Optimal Eradication Zone of an Invading Epidemic. Phytopathology, 100(7):638–644, 2010. doi: 10.1094/phyto-100-7-0638.

[259] EFSA Panel on Plant Health. Scientific opinion on the risk to plant health posed by *Xylella fastidiosa* in the EU territory, with the identification and evaluation of risk reduction options. EFSA Journal, 13(1):3989, 2015. doi: 10.2903/j.efsa.2015.3989.

[260] S. M. White, J. M. Bullock, D. A. P. Hooftman, and D. S. Chapman. Modelling the spread and control of *Xylella fastidiosa* in the early stages of invasion in Apulia, Italy. Biol. Invasions, 19(6):1825–1837, 2017. doi: 10.1007/s10530-017-1393-5.

[261] N. J. Cunniffe, B. Koskella, C. J. E. Metcalf, S. Parnell, T. R. Gottwald, and C. A. Gilligan. Thirteen challenges in modelling plant diseases. Epidemics, 10:6–10, 2015. doi: 10.1016/j.epidem.2014.06.002.

[262] IPPC Secretariat. Surveillance guide – A guide to understand the principal requirements of surveillance programmes for national plant protection organizations. Report, FAO on behalf of the Secretariat of the International Plant Protection Convention, 2021. URL `https://www.fao.org/documents/card/en/c/cb7139en`.

[263] A. J. Mastin, F. van den Bosch, Y. Bourhis, and S. Parnell. Epidemiologically-based strategies for the detection of emerging plant pathogens. Sci. Rep., 12(1):10972, 2022. doi: 10.1038/s41598-022-13553-y.

[264] S. Parnell, T. R. Gottwald, W. R. Gilks, and F. van den Bosch. Estimating the incidence of an epidemic when it is first discovered and the design of early detection monitoring. J. Theor. Biol., 305:30–36, 2012. doi: 10.1016/j.jtbi.2012.03.009.

[265] S. Parnell, T. R. Gottwald, N. J. Cunniffe, V. Alonso Chavez, and F. van den Bosch. Early detection surveillance for an emerging plant pathogen: a rule of thumb to predict prevalence at first discovery. Proc. Royal Soc. B, 282(1814), 2015. doi: 10.1098/rspb.2015.1478.

[266] EFSA Panel on Plant Health. Update of the Scientific Opinion on the risks to plant health posed by *Xylella fastidiosa* in the EU territory. EFSA Journal, 17(5):5665, 2019. doi: 10.2903/j.efsa.2019.5665.

[267] M.-A. Jacques, H. D. Coletta-Filho, L. Burbank, G. Clover, S. Elansky, F. Takashi, et al. G20 MACS Transboundary and emerging pests: *Xylella fastidiosa*. Report, G20 MACS, 2019.

[268] T. R. Gottwald. Current Epidemiological Understanding of Citrus Huanglongbing. Annu. Rev. Phytopathol., 48:119–39, 2010. doi: 10.1146/annurev-phyto-073009-114418.

[269] J. A. Lee, S. E. Halbert, W. O. Dawson, C. J. Robertson, J. E. Keesling, and B. H. Singer. Asymptomatic spread of huanglongbing and implications for disease control. Proc. Natl. Acad. Sci. U.S.A., 112(24):7605–7610, 2015. doi: 10.1073/pnas.1508253112.

[270] M. Leclerc, T. Doré, C. A. Gilligan, P. Lucas, and J. A. N. Filipe. Estimating the Delay between Host Infection and Disease (Incubation Period) and Assessing Its

Significance to the Epidemiology of Plant Diseases. PLoS ONE, 9(1):e86568, 2014. doi: 10.1371/journal.pone.0086568.

[271] K. G. Pegg, L. M. Coates, W. T. O'Neill, and D. W. Turner. The Epidemiology of Fusarium Wilt of Banana. Front. Plant Sci., 10:1395, 2019. doi: 10.3389/fpls.2019.01395.

[272] R. B. Bassanezi and R. C. Bassanezi. An approach to model the impact of Huanglongbing on citrus yield. From the proceedings of the International Research Conference on Huanglongbing (Orlando, Florida). Plant Management Network, 2008.

[273] S. Mansfield, M. R. McNeill, L. T. Aalders, N. L. Bell, J. M. Kean, B. I. P. Barratt, et al. The value of sentinel plants for risk assessment and surveillance to support biosecurity. NeoBiota, 48:1–24, 2019. doi: 10.3897/neobiota.48.34205.

[274] C. Pollock. The Canary in the Coal Mine. J. Avian Med. Surg., 30(4):386–391, 2016. doi: 10.1647/1082-6742-30.4.386.

[275] S. K. Estreicher. Wine and France: A Brief History. Eur. Rev., 31(2):91–179, 2023. doi: 10.1017/S1062798722000370.

[276] Jean Leon Blog. The rose and the vineyard, a lasting friendship, 2018. URL `https://www.jeanleon.com/en/the-rose-and-the-vineyard-lasting-friendship/`.

[277] M. L. Daughtrey, K. T. Hodge, and N. Shishkoff. The Powdery Mildews. In: Plant Pathology: Concepts and Laboratory Exercises. CRC Press, 2008.

[278] R. Eschen, R. O'Hanlon, A. Santini, A. Vannini, A. Roques, N. Kirichenko, et al. Safeguarding global plant health: the rise of sentinels. J. Pest. Sci., 92:29–36, 2019. doi: 10.1007/s10340-018-1041-6.

[279] J. M. Wells, B. C. Raju, H.-Y. Hung, W. G. Weisburg, L. Mandelco-Paul, and D. J. Brenner. *Xylella fastidiosa* gen. nov., sp. nov: Gram-Negative, Xylem-Limited, Fastidious Plant Bacteria Related to Xanthomonas spp. Int. J. Syst. Evol. Microbiol., 37(2):136–143, 1987. doi: 10.1099/00207713-37-2-136.

[280] L. Catalano, S. Al-Dobai, D. Boscia, and G. P. Martelli. Guidelines for the prevention, eradication and containment of *Xylella fastidiosa* in olive-growing areas. Report, FAO, 2019. URL `https://www.fao.org/documents/card/fr/c/i5994en/`.

[281] M. J. Davis, A. H. Purcell, and S. V. Thomson. Pierce's disease of grapevines: isolation of the causal bacterium. Science, 199(4324):75–77, 1978. doi: 10.1126/science.199.4324.75.

[282] A. J. G. Simpson, F. C. Reinach, P. Arruda, F. A. Abreu, M. Acencio, R. Alvarenga, et al. The genome sequence of the plant pathogen *Xylella fastidiosa*. Nature, 406 (6792):151–157, 2000. doi: 10.1038/35018003.

[283] J. Rapicavoli, B. Ingel, B. Blanco-Ulate, D. Cantu, and C. Roper. *Xylella fastidiosa*: an examination of a re-emerging plant pathogen. Mol. Plant Pathol., 19(4):786–800, 2018. doi: 10.1111/mpp.12585.

[284] European Food Safety Authority (EFSA), A. Delbianco, D. Gibin, L. Pasinato, D. Boscia, and M. Morelli. Report on the update of the *Xylella* spp. host plant database - systematic literature search up to 31 December 2021. EFSA Journal, 20(6):7356, 2022. doi: 10.2903/j.efsa.2023.8061.

[285] H. D. Coletta-Filho, A. I. Castillo, F. F. Laranjeira, E. C. de Andrade, N. T. Silva, A. A. de Souza, et al. Citrus Variegated Chlorosis: an Overview of 30 years of Research and Disease Management. Trop. Plant Pathol., 45:175–191, 2020. doi: 10.1007/s40858-020-00358-5.

[286] I. Kyrkou, T. Pusa, L. Ellegaard-Jensen, M. F. Sagot, and L. H. Hansen. Pierce's Disease of Grapevines: A Review of Control Strategies and an Outline of an Epidemiological Model. Front. Microbiol., 9:2141, 2018. doi: 10.3389/fmicb.2018.02141.

[287] D. L. Hopkins and A. H. Purcell. *Xylella fastidiosa*: Cause of Pierce's Disease of Grapevine and Other Emergent Diseases. Plant Dis., 86(10):1056–1066, 2002. doi: 10.1094/PDIS.2002.86.10.1056.

[288] F. P. Gonçalves, E. S. Stuchi, S. A. Lourenço, B. Hau, and L. Amorim. Relationship between sweet orange yield and intensity of Citrus Variegated Chlorosis. Plant Pathol., 61(4):641–647, 2012. doi: 10.1111/j.1365-3059.2011.02557.x.

[289] F. P. Gonçalves, E. S. Stuchi, S. A. Lourenço, A. B. Kriss, T. R. Gottwald, and L. Amorim. The effect of irrigation on development of citrus variegated chlorosis symptoms. Crop Prot., 57:8–14, 2014. doi: 10.1016/j.cropro.2013.11.016.

[290] C. M. Rodrigues, A. A. de Souza, M. A. Takita, L. T. Kishi, and M. A. Machado. RNA-Seq analysis of Citrus reticulata in the early stages of *Xylella fastidiosa* infection reveals auxin-related genes as a defense response. BMC Genom., 14(1):676, 2013. doi: 10.1186/1471-2164-14-676.

[291] L. Nunney, B. Ortiz, S. A. Russell, R. Ruiz Sánchez, and R. Stouthamer. The Complex Biogeography of the Plant Pathogen *Xylella fastidiosa*: Genetic Evidence of Introductions and Subspecific Introgression in Central America. PLoS ONE, 9(11): e112463, 2014. doi: 10.1371/journal.pone.0112463.

[292] M. Godefroid, A. Cruaud, J.-C. Streito, J.-Y. Rasplus, and J.-P. Rossi. *Xylella fastidiosa*: climate suitability of European continent. Sci. Rep., 9(1):8844, 2019. doi: 10.1038/s41598-019-45365-y.

[293] A. Giménez-Romero, J. Galván, M. Montesinos, J. Bauzà, M. Godefroid, A. Fereres, et al. Global predictions for the risk of establishment of Pierce's disease of grapevines. Commun. Biol., 5(1):1389, 2022. doi: 10.1038/s42003-022-04358-w.

[294] A. Luvisi, F. Nicoli, and L. De Bellis. Sustainable Management of Plant Quarantine Pests: The Case of Olive Quick Decline Syndrome. Sustainability, 9(4):659, 2017. doi: 10.3390/su9040659.

[295] D. Cornara, M. Morente, A. Markheiser, N. Bodino, C.-W. Tsai, A. Fereres, et al. An overview on the worldwide vectors of *Xylella fastidiosa*. Entomol. Gen., 39(3-4): 157–181, 2019. doi: 10.1127/entomologia/2019/0811.

[296] M. Saponari, D. Boscia, F. Nigro, and G. P. Martelli. Identification of DNA Sequences Related to *Xylella fastidiosa* in Oleander, Almond and Olive Trees Exhibiting Leaf Scorch Symptoms in Apulia (southern Italy). J. Plant Pathol., 95(3):659–668, 2013. doi: 10.4454/JPP.V95I3.035.

[297] G. P. Martelli, D. Boscia, F. Porcelli, and M. Saponari. The olive quick decline syndrome in south-east Italy: a threatening phytosanitary emergency. Eur. J. Plant. Pathol., 144(2):235–243, 2016. doi: 10.1007/s10658-015-0784-7.

[298] A. Sicard, M. Saponari, M. Vanhove, A. I. Castillo, A. Giampetruzzi, G. Loconsole, et al. Introduction and adaptation of an emerging pathogen to olive trees in Italy. Microb. Genom., 7(12):000735, 2021. doi: 10.1099/mgen.0.000735.

[299] EPPO Global Database. *Xylella fastidiosa* distribution, 2023. URL `https://gd.eppo.int/taxon/XYLEFA/distribution`.

[300] G. Strona, C. J. Carstens, and P. S. A. Beck. Network analysis reveals why *Xylella fastidiosa* will persist in Europe. Sci. Rep., 7(71):1–8, 2017. doi: 10.1038/s41598-017-00077-z.

[301] R. P. P. Almeida. Can Apulia's olive trees be saved? Science, 353(6297):346–348, 2016. doi: doi:10.1126/science.aaf9710.

[302] M. Saponari, A. Giampetruzzi, G. Loconsole, D. Boscia, and P. Saldarelli. *Xylella fastidiosa* in Olive in Apulia: Where We Stand. Phytopathology, 109(2):175–186, 2019. doi: 10.1094/phyto-08-18-0319-fi.

[303] European Commission. *Xylella fastidiosa*, Accessed 20 June 2023. URL `https://food.ec.europa.eu/plants/plant-health-and-biosecurity/legislation/control-measures/xylella-fastidiosa_en`.

[304] G. Cardone, M. Digiaro, K. Djelouah, M. Frem, C. Rota, A. Lenders, et al. Socio-Economic Risks Posed by a New Plant Disease in the Mediterranean Basin. Diversity, 14(11):975, 2022. doi: 10.3390/d14110975.

[305] P. S. Pereira. *Xylella fastidiosa* - a new menace for Portuguese agriculture and forestry. Rev. Fac. Cienc. Agrar., 38(2):149–154, 2015. doi: 10.19084/RCA.16907.

[306] S. J. Harper, L. I. Ward, and G. R. G. Clover. Development of LAMP and Real-Time PCR Methods for the Rapid Detection of *Xylella fastidiosa* for Quarantine and Field Applications. Phytopathology, 100(12):1282–1288, 2010. doi: 10.1094/PHYTO-06-10-0168.

[307] G. V. Minsavage, C. M. Thompson, D. L. Hopkins, R. M. Leite, and S. R. E. Development of a Polymerase Chain Reaction Protocol for Detection of *Xylella fastidiosa* in Plant Tissue. Phytopathology, 84:456–461, 1994. doi: 10.1094/Phyto-84-456.

[308] GBIF: The Global Biodiversity Information Facility, Accessed 03 August 2022. URL `https://www.gbif.org/occurrence/search?q=Catharanthus%20roseus`.

[309] S. M. White, J. A. Navas-Cortés, J. M. Bullock, D. Boscia, and D. S. Chapman. Estimating the epidemiology of emerging *Xylella fastidiosa* outbreaks in olives. Plant Pathol., 69(8):1403–1413, 2020. doi: 10.1111/ppa.13238.

[310] D. Cornara, V. Cavalieri, C. Dongiovanni, G. Altamura, F. Palmisano, D. Bosco, et al. Transmission of *Xylella fastidiosa* by naturally infected *Philaenus spumarius* (Hemiptera, Aphrophoridae) to different host plants. J. Appl. Entomol., 141(1-2): 80–87, 2017. doi: 10.1111/jen.12365.

[311] N. Bodino, V. Cavalieri, M. Saponari, C. Dongiovanni, G. Altamura, and D. Bosco. Transmission of *Xylella fastidiosa* subsp. *pauca* ST53 by the Sharpshooter *Cicadella viridis* From Different Source Plants and Artificial Diets. J. Econ. Entomol., 115(6): 1852–1858, 2022. doi: 10.1093/jee/toac172.

[312] E. Brochu, V. M. Cora, and N. de Freitas. A Tutorial on Bayesian Optimization of Expensive Cost Functions, with Application to Active User Modeling and Hierarchical Reinforcement Learning. arXiv:1012.2599, 2010.

[313] P. I. Frazier. A Tutorial on Bayesian Optimization. arXiv:1807.02811, 2018.

[314] MathWorks. Bayesian Optimization Algorithm, 2021. URL `https://uk.mathworks.com/help/stats/bayesian-optimization-algorithm.html`.

[315] MathWorks. bayesopt, 2021. URL `https://uk.mathworks.com/help/stats/bayesopt.html`.

[316] MathWorks. Constraints in Bayesian Optimization, 2021. URL `https://uk.mathworks.com/help/stats/constraints-in-bayesian-optimization.html`.

[317] EPPO. PM 7/24 (4) *Xylella fastidiosa*. Bull. OEPP, 49(2):175–227, 2019. doi: 10.1111/epp.12575.

[318] S. Gaire, U. Albrecht, O. Batuman, J. Qureshi, M. Zekri, and F. Alferez. Individual protective covers (IPCs) to prevent Asian citrus psyllid and *Candidatus* Liberibacter asiaticus from establishing in newly planted citrus trees. Crop Prot., 152:105862, 2022. doi: 10.1016/j.cropro.2021.105862.

[319] D. G. Hall, U. Albrecht, and K. D. Bowman. Transmission rates of 'ca. liberibacter asiaticus' by asian citrus psyllid are enhanced by the presence and developmental stage of citrus flush. J. Econ. Entomol., 109(2):558–63, 2016. doi: 10.1093/jee/tow009.

[320] M. P. Daugherty, A. Rashed, R. P. P. Almeida, and T. M. Perring. Vector preference for hosts differing in infection status: sharpshooter movement and *Xylella fastidiosa* transmission. Ecol. Entomol., 36(5):654–662, 2011. doi: 10.1111/j.1365-2311.2011.01309.x.

[321] A. J. Mastin, T. R. Gottwald, F. van den Bosch, N. J. Cunniffe, and S. R. Parnell. Optimising risk-based surveillance for early detection of invasive plant pathogens. PLoS Biol., 18(10):e3000863, 2020. doi: 10.1371/journal.pbio.3000863.

[322] EPPO. PM 7/24 (2) *Xylella fastidiosa*. Bull. OEPP, 46(3):463–500, 2016. doi: 10.1111/epp.12327.

[323] L. L. Raiol-Junior, J. C. Cifuentes-Arenas, N. J. Cunniffe, R. Turgeon, and S. A. Lopes. Modeling '*Candidatus* Liberibacter asiaticus' Movement Within Citrus Plants. Phytopathology, 111(10):1711–1719, 2021. doi: 10.1094/PHYTO-12-20-0559-R.

[324] N. J. Cunniffe, R. O. J. H. Stutt, F. van den Bosch, and C. A. Gilligan. Time-Dependent Infectivity and Flexible Latent and Infectious Periods in Compartmental Models of Plant Disease. Phytopathology, 102(4):365–380, 2012. doi: 10.1094/phyto-12-10-0338.

[325] R. N. Thompson and W. S. Hart. Effect of Confusing Symptoms and Infectiousness on Forecasting and Control of Ebola Outbreaks. Clin. Infect. Dis., 67(9):1472–1474, 2018. doi: 10.1093/cid/ciy248.

[326] M. Saponari, G. Loconsole, D. Cornara, R. K. Yokomi, A. De Stradis, D. Boscia, et al. Infectivity and Transmission of *Xylella fastidiosa* by *Philaenus spumarius* (Hemiptera: Aphrophoridae) in Apulia, Italy. J. Econ. Entomol., 107(4):1316–1319, 2014. doi: 10.1603/ec14142.

[327] D. Cornara, M. Saponari, A. R. Zeilinger, A. de Stradis, D. Boscia, G. Loconsole, et al. Spittlebugs as vectors of *Xylella fastidiosa* in olive orchards in Italy. J. Pest. Sci., 90 (2):521–530, 2017. doi: 10.1007/s10340-016-0793-0.

[328] J. Albre, J. M. García Carrasco, and M. Gibernau. Ecology of the meadow spittlebug *Philaenus spumarius* in the Ajaccio region (Corsica) – i: spring. Bull. Entomol. Res., 111(2):246–256, 2021. doi: 10.1017/S0007485320000711.

[329] N. Bodino, V. Cavalieri, C. Dongiovanni, E. Plazio, M. A. Saladini, S. Volani, et al. Phenology, seasonal abundance and stage-structure of spittlebug (Hemiptera: Aphrophoridae) populations in olive groves in italy. Sci. Rep., 9(1):17725, 2019. doi: 10.1038/s41598-019-54279-8.

[330] D. Cornara, D. Bosco, and A. Fereres. *Philaenus spumarius*: when an old acquaintance becomes a new threat to European agriculture. J. Pest. Sci., 91(3):957–972, 2018. doi: 10.1007/s10340-018-0966-0.

[331] G. Santoiemma, G. Tamburini, F. Sanna, N. Mori, and L. Marini. Landscape composition predicts the distribution of *Philaenus spumarius*, vector of *Xylella fastidiosa*, in olive groves. J. Pest. Sci., 92(3):1101–1109, 2019. doi: 10.1007/s10340-019-01095-8.

[332] I. E. Ben Moussa, V. Mazzoni, F. Valentini, T. Yaseen, D. Lorusso, S. Speranza, et al. Seasonal Fluctuations of Sap-Feeding Insect Species infected by *Xylella fastidiosa* in Apulian Olive Groves of Southern Italy. J. Econ. Entomol., 109(4):1512–8, 2016. doi: 10.1093/jee/tow123.

[333] A. J. McElrone, J. L. Sherald, and I. N. Forseth. Interactive effects of water stress and xylem-limited bacterial infection on the water relations of a host vine. J. Exp. Bot., 54(381):419–430, 2003. doi: 10.1093/jxb/erg046.

[334] M. De Pascali, M. Vergine, C. Negro, D. Greco, F. Vita, E. Sabella, et al. *Xylella fastidiosa* and Drought Stress in Olive Trees: A Complex Relationship Mediated by Soluble Sugars. Biology, 11(1):112, 2022. doi: 10.3390/biology11010112.

[335] H. Feil and A. H. Purcell. Temperature-Dependent Growth and Survival of *Xylella fastidiosa* in Vitro and in Potted Grapevines. Plant Dis., 85(12):1230–1234, 2001. doi: 10.1094/PDIS.2001.85.12.1230.

[336] Weather Spark. Climate and Average Weather Year Round in Lecce. URL https://weatherspark.com/y/83079/Average-Weather-in-Lecce-Italy-Year-Round#Figures-Daylight.

[337] ISPM 27 Diagnostic protocols for regulated pests. DP 25: *Xylella fastidiosa*. Report, International Plant Protection Convention, 2018. URL https://www.ippc.int/static/media/files/publication/en/2018/09/DP_25_2018_Xylellafastidiosa_2018-09-21.pdf.

[338] G. C. Adams, M. Catall, J. Walla, and A. B. Gould. Bacterial leaf scorch distribution and isothermal lines (PROJECT NC-EM-08-02). In: Forest health monitoring: national status, trends, and analysis 2011. Gen. Tech. Rep. SRS-GTR-185. USDA Forest Service, Southern Research Station, 2013. doi: 10.2737/SRS-GTR-185.

[339] M. De Pascali, M. Vergine, E. Sabella, A. Aprile, E. Nutricati, F. Nicolì, et al. Molecular Effects of *Xylella fastidiosa* and Drought Combined Stress in Olive Trees. Plants, 8 (11):437, 2019. doi: 10.3390/plants8110437.

[340] A. J. McElrone, J. L. Sherald, and I. N. Forseth. Effects of Water Stress on Symptomatology and Growth of *Parthenocissus quinquefolia* Infected by *Xylella fastidiosa*. Plant Dis., 85(11):1160–1164, 2001. doi: 10.1094/PDIS.2001.85.11.1160.

[341] European Commission, Directorate-General for Health, and Food Safety. Guidelines for the survey of *Xylella fastidiosa* (Wells et al.) in the Union territory. Report, 2015. URL https://food.ec.europa.eu/system/files/2016-10/ph_biosec_legis_guidelines_xylella-survey.pdf.

[342] D. Cornara, A. Sicard, A. R. Zeilinger, F. Porcelli, A. H. Purcell, and R. P. P. Almeida. Transmission of *Xylella fastidiosa* to Grapevine by the Meadow Spittlebug. Phytopathology, 106(11):1285–1290, 2016. doi: 10.1094/PHYTO-05-16-0202-R.

[343] D. F. Anderson. A modified next reaction method for simulating chemical systems with time dependent propensities and delays. J. Chem. Phys., 127(21):214107, 2007. doi: 10.1063/1.2799998.

[344] V. H. Thanh and C. Priami. Simulation of biochemical reactions with time-dependent rates by the rejection-based algorithm. J. Chem. Phys., 143(5):054104, 2015. doi: 10.1063/1.4927916.

[345] S. Waliullah, O. Hudson, J. E. Oliver, P. M. Brannen, P. Ji, and M. E. Ali. Comparative analysis of different molecular and serological methods for detection of *Xylella fastidiosa* in blueberry. PLoS ONE, 14(9):e0221903, 2019. doi: 10.1371/journal.pone.0221903.

[346] R. A. Weiss and N. Sankaran. Emergence of epidemic diseases: zoonoses and other origins. Fac. Rev., 11:2, 2022. doi: 10.12703/r/11-2.

[347] R. E. Baker, A. S. Mahmud, I. F. Miller, M. Rajeev, F. Rasambainarivo, B. L. Rice, et al. Infectious disease in an era of global change. Nat. Rev. Microbiol., 20(4):193–205, 2022. doi: 10.1038/s41579-021-00639-z.

[348] S. Milne-Price, K. L. Miazgowicz, and V. J. Munster. The emergence of the Middle East respiratory syndrome coronavirus. Pathog. Dis., 71(2):121–36, 2014. doi: 10.1111/2049-632x.12166.

[349] S. Guth, E. Visher, M. Boots, and C. E. Brook. Host phylogenetic distance drives trends in virus virulence and transmissibility across the animal–human interface. Philos. Trans. R. Soc. Lond., B, Biol. Sci., 374(1782):20190296, 2019. doi: 10.1098/rstb.2019.0296.

[350] E. Visher, C. Evensen, S. Guth, E. Lai, M. Norfolk, C. Rozins, et al. The three Ts of virulence evolution during zoonotic emergence. Proc. Royal Soc. B, 288(1956): 20210900, 2021. doi: 10.1098/rspb.2021.0900.

[351] P. Renault, J.-L. Solet, D. Sissoko, E. Balleydier, S. Larrieu, L. Filleul, et al. A Major Epidemic of Chikungunya Virus Infection on Reunion Island, France, 2005–2006. Am. J. Trop. Med. Hyg., 77(4):727–731, 2007. doi: 10.4269/ajtmh.2007.77.727.

[352] M. Vazeille, S. Moutailler, D. Coudrier, C. Rousseaux, H. Khun, M. Huerre, et al. Two Chikungunya Isolates from the Outbreak of La Reunion (Indian Ocean) Exhibit Different Patterns of Infection in the Mosquito, *Aedes albopictus*. PLoS ONE, 2(11): e1168, 2007. doi: 10.1371/journal.pone.0001168.

[353] V. N. Petrova and C. A. Russell. The evolution of seasonal influenza viruses. Nat. Rev. Microbiol., 16(1):47–60, 2018. doi: 10.1038/nrmicro.2017.118.

[354] A. Blagodatski, K. Trutneva, O. Glazova, O. Mityaeva, L. Shevkova, E. Kegeles, et al. Avian Influenza in Wild Birds and Poultry: Dissemination Pathways, Monitoring Methods, and Virus Ecology. Pathogens, 10(5):630, 2021. doi: 10.3390/pathogens10050630.

[355] V. Caliendo, N. S. Lewis, A. Pohlmann, S. R. Baillie, A. C. Banyard, M. Beer, et al. Transatlantic spread of highly pathogenic avian influenza H5N1 by wild birds from Europe to North America in 2021. Sci. Rep., 12(1):11729, 2022. doi: 10.1038/s41598-022-13447-z.

[356] R. N. Binny, M. G. Baker, S. C. Hendy, A. James, A. Lustig, M. J. Plank, et al. Early intervention is the key to success in COVID-19 control. R. Soc. Open Sci., 8(11): 210488, 2021. doi: 10.1098/rsos.210488.

[357] B. Y. Lee and L. A. Haidari. The importance of vaccine supply chains to everyone in the vaccine world. Vaccine, 35(Pt A):4475–4479, 2017. doi: 10.1016/j.vaccine.2017.05.096.

[358] H. Onyeaka, C. K. Anumudu, Z. T. Al-Sharify, E. Egele-Godswill, and P. Mbaegbu. COVID-19 pandemic: A review of the global lockdown and its far-reaching effects. Sci. Prog., 104(2):00368504211019854, 2021. doi: 10.1177/00368504211019854.

[359] A. M. Clarfield. Demography Will Continue to Shape the Course of History. Am. J. Public Health, 110(7):919–920, 2020. doi: 10.2105/AJPH.2020.305652.

[360] N. Sudharsanan and D. Bloom. The Demography of Aging in Low- and Middle-Income Countries: Chronological versus Functional Perspectives. National Academies Press (US), 2018.

[361] K. Prem, K. van Zandvoort, P. Klepac, R. M. Eggo, N. G. Davies, Centre for the Mathematical Modelling of Infectious Diseases Covid Working Group, et al. Projecting contact matrices in 177 geographical regions: An update and comparison with empirical data for the COVID-19 era. PLoS Comput. Biol., 17(7):e1009098, 2021. doi: 10.1371/journal.pcbi.1009098.

[362] X. Feng, D. D. Kim, J. T. Cohen, P. J. Neumann, and D. A. Ollendorf. Using QALYs versus DALYs to measure cost-effectiveness: How much does it matter? Int. J. Technol. Assess. Health Care, 36(2):96–103, 2020. doi: 10.1017/S0266462320000124.

[363] P. J. Neumann, J. E. Anderson, A. D. Panzer, E. F. Pope, B. N. D'Cruz, D. D. Kim, et al. Comparing the cost-per-QALYs gained and cost-per-DALYs averted literatures. Gates Open Res., 2:5, 2018. doi: 10.12688/gatesopenres.12786.2.

[364] L. Prieto and J. A. Sacristán. Problems and solutions in calculating quality-adjusted life years (QALYs). Health Qual. Life Outcomes, 1:80, 2003. doi: 10.1186/1477-7525-1-80.

[365] A. Tsuchiya. QALYs and ageism: philosophical theories and age weighting. Health Econ., 9(1):57–68, 2000. doi: 10.1002/(sici)1099-1050(200001)9:1⟨57::aid-hec484⟩3.0.co;2-n.

[366] R. Nathan, E. Klein, J. J. Robledo-Anuncio, and E. Revilla. Dispersal kernels: review. In: Dispersal Ecology and Evolution. Oxford University Press, first edition, 2012.

[367] N. J. Cunniffe, N. P. Taylor, F. M. Hamelin, and M. J. Jeger. Epidemiological and ecological consequences of virus manipulation of host and vector in plant virus transmission. PLoS Comput. Biol., 17(12):e1009759, 2022. doi: 10.1371/journal.pcbi.1009759.

[368] R. Prasad, S. K. Sagar, S. Parveen, and R. Dohare. Mathematical modeling in perspective of vector-borne viral infections: a review. Beni-Suef Univ. J. Basic Appl. Sci., 11(1):102, 2022. doi: 10.1186/s43088-022-00282-4.

[369] European Food Safety Authority, E. Lázaro, S. Parnell, A. Vicent Civera, J. Schans, M. Schenk, et al. Guidelines for statistically sound and risk-based surveys of *Xylella fastidiosa*. EFSA Supporting Publications, 17(6):1873E, 2020. doi: 10.2903/sp.efsa.2020.EN-1873.

[370] World Health Organisation. 2022-23 Mpox (Monkeypox) Outbreak: Global Trends, 2023. URL `https://worldhealthorg.shinyapps.io/mpx_global/`.

[371] C. Adlhoch and F. Baldinelli. Avian influenza, new aspects of an old threat. Euro Surveill., 28(19):2300227, 2023. doi: 10.2807/1560-7917.Es.2023.28.19.2300227.

[372] T. R. Frieden, C. T. Lee, A. F. Bochner, M. Buissonnière, and A. McClelland. 7-1-7: an organising principle, target, and accountability metric to make the world safer from pandemics. Lancet, 398(10300):638–640, 2021. doi: 10.1016/s0140-6736(21)01250-2.

[373] Regional Committee for Africa. Regional strategy for health security and emergencies 2022–2030: report of the Secretariat. Report, World Health Organization Regional Office for Africa, 2022. URL `https://apps.who.int/iris/handle/10665/361858`.

[374] Prevent Epidemics. 7-1-7: Rapid Improvement for Early Disease Detection & Response, 2023. URL `https://preventepidemics.org/preparedness/7-1-7/`.

[375] A. F. Bochner, I. Makumbi, O. Aderinola, A. Abayneh, R. Jetoh, R. L. Yemanaberhan, et al. Implementation of the 7-1-7 target for detection, notification, and response to public health threats in five countries: a retrospective, observational study. Lancet Glob. Health, 11(6):e871–e879, 2023. doi: 10.1016/s2214-109x(23)00133-x.