

Population and Within-Host Modelling of Virus-Disease Dynamics



Rahil Sachak-Patwa
St Anne's College
University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy

Michaelmas 2021

Abstract

Mathematical modelling is an important tool in helping make projections of infectious disease dynamics. In this thesis, we address epidemiological modelling questions at both the population and within-host scales. By focussing on problems relating to immunity and within-host immunological dynamics, we aim to establish how the accuracy of model projections can be improved.

First, we consider the spread of SARS-CoV-2, using a stochastic model to determine the probability that infected individuals introduced into a population with low case prevalence initiate an outbreak. We consider Israel and the Isle of Man as case studies, two regions that have had few cases and that are quickly vaccinating their populations. Our results suggest that when travel restrictions are relaxed, in order to prevent local transmission, it will still be necessary to implement surveillance of incoming passengers to identify infected individuals.

We then consider forecasting influenza epidemics. Previous exposure to influenza viruses confers cross-immunity against future infections with related strains, however this effect is not always included in influenza forecasting models. We use case notification data from Japan from the 2009 H1N1 influenza pandemic to estimate parameter values and then consider a hypothetical future epidemic. Our results show that a more epidemiologically realistic model, in which cross-immunity is represented explicitly, can generate more accurate projections during outbreaks than if cross-immunity is not included.

We next model influenza virus dynamics at the within-host scale. We consider a hierarchy of three ordinary differential equation models which describe the kinetics of influenza infection; a complex model which explicitly accounts for an immune response to the virus, and two simpler models in which the infection is limited by the availability of target-cells. We show that the simplest target-cell limited model can reproduce synthetic data generated from the complex immune response model, and furthermore that the parameters of the target-cell limited model vary in a systematic manner as the parameters in the complex model used to generate the data are changed. Finally, we extend these within-host models to describe the severity of symptoms experienced by a host. We show that a simple model can be used to make real-time forecasts of the severity of influenza symptoms and be used to classify patients based on estimated model parameter values.

Acknowledgements

First and foremost, I would like to thank my supervisors Helen Byrne and Robin Thompson for their continued support over the past few years. I count myself very lucky to have had the pleasure of working with them and am grateful for all their help. A big thank you also to Erin Lafferty and Claude Schmit from Biosensors Beyond Borders for their guidance and collaboration.

This journey would have been nowhere near as enjoyable without the camaraderie of InFoMM and Cohort 4. Thanks to all my coursemates and friends, including of course members of The Machine.

My time in Oxford has been largely spent running with OUCCC. Thanks for all the memories, in particular to Edward Brooks, Alexander Betts, and Richard Burman for dragging me out of bed in the early hours of the morning.

This DPhil was everything I wanted it to be, thanks everyone for making that happen.

Contents

1	Introduction	1
1.1	Public health impacts of viral pathogens	1
1.2	Role of mathematical modelling	2
1.3	Viral emergence and immunity	3
1.4	Epidemiology of viral infection and outbreaks	5
1.5	Infectious disease models	8
1.5.1	Population level epidemic models	8
1.5.2	Within-host models	10
1.6	Data and parameter estimation	12
1.6.1	Data	12
1.6.2	Parameter estimation	13
1.7	Thesis outline	15
2	The risk of SARS-CoV-2 outbreaks in low prevalence settings following the removal of travel restrictions	18
2.1	Introduction	18
2.2	Methods	20
2.2.1	Epidemiological model	20
2.2.2	Vaccination data	21
2.2.3	Outbreak risk metrics	22
2.2.3.1	Instantaneous Outbreak Risk (IOR)	22
2.2.3.2	Case Outbreak Risk (COR)	24
2.2.3.3	Simulated Outbreak Risk (SOR)	25
2.2.3.4	Numerical Outbreak Risk (NOR)	26
2.3	Results	27
2.3.1	Modelling the vaccination process and population susceptibility	27
2.3.2	Modelling the risk of an outbreak	27
2.4	Discussion	30

3	Accounting for cross-immunity can improve forecast accuracy during influenza epidemics	34
3.1	Introduction	34
3.2	Methods	35
3.2.1	Data	35
3.2.2	Models	36
3.2.2.1	1-group model	36
3.2.2.2	2-group model	37
3.2.3	Parameter estimation and forecasting	38
3.2.4	Modelling the size of the cross-immune population between epidemics .	39
3.2.5	Approaches to forecasting a future epidemic	40
3.3	Results	40
3.3.1	Fitting the models to the 2009 H1N1 influenza epidemic in Japan	40
3.3.2	Forecasting an epidemic in real-time without prior information	41
3.3.3	Forecasting epidemics in advance	43
3.3.4	Forecasting an epidemic in real-time with prior information	44
3.4	Discussion	46
4	A target-cell limited model can reproduce influenza infection dynamics in hosts with differing immune responses	51
4.1	Introduction	51
4.2	Methods	53
4.2.1	Models	53
4.2.1.1	IR model	53
4.2.1.2	TEIV model	55
4.2.1.3	TV model	56
4.2.2	Experimental data	56
4.2.3	Structural identifiability analysis	57
4.2.4	Practical identifiability analysis	58
4.2.5	Parameter estimation	59
4.2.6	Global sensitivity analysis	59
4.3	Results	60
4.3.1	Structural identifiability analysis	60
4.3.2	Practical identifiability analysis	61
4.3.3	Fitting the models to experimental viral load data	61
4.3.4	Fitting the IR model to pooled experimental data	62
4.3.5	Modulating the immune response in the IR model	65
4.3.6	Using the TEIV and TV models to reproduce viral load dynamics of the IR Model	67

4.3.7	Using the TV model to reproduce viral load dynamics in hosts with different immunological characteristics	71
4.3.8	Forecasting influenza viral load dynamics	74
4.4	Discussion	75
5	Modelling and forecasting the symptom dynamics of influenza	79
5.1	Introduction	79
5.2	Methods	80
5.2.1	Models	80
5.2.1.1	SD model	81
5.2.1.2	TS model	82
5.2.2	Experimental data	83
5.2.3	Parameter estimation and forecasting	84
5.2.4	Generating synthetic data	85
5.2.5	Classification	85
5.3	Results	86
5.3.1	Fitting the SD model to pooled experimental data	86
5.3.2	Fitting the TS model to synthetic symptom score data	86
5.3.3	Forecasting symptom scores in real-time using the TS model	88
5.3.4	Fitting the TS model to symptom score data to classify patients	95
5.4	Discussion	96
6	Discussion	100
6.1	Introduction	100
6.2	Summary of results	100
6.3	Model complexity and incorporating immunity	102
6.4	Areas for future research	104
6.5	Concluding remarks	106
A	Appendix to Chapter 2	108
A.1	The effect of the assumed vaccine uptake on the outbreak risk	108
A.2	The effect of immunity due to prior infections on the outbreak risk	109
A.3	The effect of the assumed effectiveness of two vaccine doses on the outbreak risk	110
A.4	Sensitivity analysis of the outbreak risk to model parameters	111
A.5	Individual level variation in SARS-CoV-2 transmission	111

B	Appendix to Chapter 3	114
B.1	Sensitivity analysis: latent and infectious periods	114
B.2	Sensitivity analysis: the cross-immune fraction	115
B.3	Calculation of R_0 for the 2-group model	118
B.4	Residual analysis of models fitted to the 2009 epidemic	119
B.5	Parameter estimation	120
B.6	Estimation of initial conditions	121
B.7	Deviance information criterion (DIC)	122
B.8	Modelling the cross-immune fraction of the population	123
B.9	The 2-group model fitted to the 2009 epidemic: total recorded and unrecorded cases	125
B.10	Prior fitting parameter distributions	126
B.11	Forecasting epidemics in advance using different prior parameter distributions .	127
B.12	Outbreak statistics of future epidemics	129
B.13	Forecasting an epidemic in real-time with prior information using different prior parameter distributions	131
C	Appendix to Chapter 4	133
C.1	Parameter estimation	133
C.2	Deviance information criterion (DIC) and Akaike information criterion (AIC) .	134
C.3	Global sensitivity analysis	135
C.4	Structural identifiability analysis of the IR model	137
C.4.1	Nondimensionalisation	137
C.4.2	Taylor series calculations: $\tau = 0$ (Case 1b extension)	137
C.4.3	Taylor series calculations: $\tau > 0$ (Case 1a extension)	140
C.5	Practical identifiability analysis: profile likelihood curves	144
D	Appendix to Chapter 5	147
D.1	Parameter estimation	147
D.2	Scatter plots of ω and η	148
D.3	Modulation of ω and η	148
D.4	Synthetic prior patient data	149
D.5	Deviance information criterion (DIC)	149
E	MATLAB code	150
F	The risk of SARS-CoV-2 outbreaks in low prevalence settings following the removal of travel restrictions	151

G Accounting for cross-immunity can improve forecast accuracy during influenza epidemics	161
Bibliography	175

List of Figures

1.1	A diagram of influenza viral infection at the cellular level	6
1.2	Influenza immune response timeline	7
1.3	Schematic of the SEIR model and example of population level epidemic data .	8
1.4	Comparison of stochastic and deterministic SEIR epidemic models in small and large populations	10
1.5	Schematic of the TEIV model and example of within-host infection data	11
1.6	Schematic of within-host infection and simulation of the TEIV target-cell model	12
1.7	SEIR model fitted to synthetic epidemic case data using MLE and MCMC . .	15
2.1	The epidemiological model used to describe transmission and the vaccination process governing population susceptibility	21
2.2	The proportion of the population vaccinated and the vaccine-associated reproduction number	28
2.3	The risk that an infectious case introduced at each stage of the vaccination campaign initiates an outbreak if travel restrictions and other NPIs are removed	29
3.1	Schematic of the 1-group and 2-group transmission models	36
3.2	Schematic illustrating the modelling approaches considered for forecasting the dynamics of a future epidemic	41
3.3	The 1-group and 2-group models fitted to data describing the numbers of new recorded cases each week and scatter plots of the posterior distributions of R_0 and N	42
3.4	Forecast Strategy 1: Real-time forecasts obtained from the 1-group and 2-group models for a future epidemic occurring 25 years after the 2009 epidemic, calibrated by fitting the model parameters β and N to data describing the weekly numbers of new cases, using uninformative uniform priors	44
3.5	Forecasts Strategy 2: Forecasts obtained at the start of future epidemics using the 1-group and 2-group models, for epidemics occurring 25, 50, and 75 years after the 2009 epidemic	45

3.6	Forecast Strategy 3: Forecasts obtained using the 1-group and 2-group models during a future epidemic occurring 25 years after the 2009 epidemic, at the start of the epidemic, and using real-time case notification data obtained during the epidemic	47
4.1	Schematic of the IR, TEIV, and TV models	54
4.2	Profile likelihood (χ^2_{PL}) curves generated by fitting the IR, TEIV, and TV models to synthetic data for the specific cases: IR model Case 1, TEIV model Case 8, and TV model Case 9	63
4.3	The IR, TEIV, and TV models fitted to experimental data on the viral load of infected immunocompetent mice and the posterior distributions of fitting parameters	64
4.4	The IR model fitted to pooled experimental data on the viral load, fraction of dead cells, and adaptive immune response in immunosuppressed and immunocompetent mice and the posterior distributions of fitting parameters	66
4.5	The change in the three infection-related quantities of interest as the IR model innate and adaptive immune response parameters are varied	68
4.6	The TEIV model fitted to synthetic data as the immune response parameters in the IR model used to generate the data are individually varied	69
4.7	The TV model fitted to synthetic data as the immune response parameters in the IR model used to generate the data are individually varied	70
4.8	The TV model fitted to synthetic data relating to different patient groups . . .	73
4.9	Comparison of viral load forecasts obtained using the TV and IR models	75
5.1	Schematic of the SD, TVS, and TS models	81
5.2	The SD model fitted to experimental data on the viral load, innate immune response (cytokines), and symptom score and the posterior distributions of fitting parameters	87
5.3	The TS model fitted to synthetic symptom score data generated from the SD model for a healthy and comorbid patient and the posterior distributions of fitting parameters	88
5.4	Real-time forecasts obtained from the TS model of the symptom score dynamics of a healthy and comorbid patient using no prior information from previous patients	90
5.5	Real-time forecasts obtained from the TS model of the symptom score dynamics of a healthy and comorbid patient using prior information of symptom scores from 24 previous patients	91

5.6	The TS model fitted to the symptom score data of healthy and comorbid patients, when the first symptom score datapoint is observed at some unknown time t_0 after infection	93
5.7	Real-time forecasts obtained from the TS model of the symptom score dynamics of a healthy and comorbid patient where the time of infection is unknown . . .	94
5.8	The TS model used to classify patients as healthy or comorbid	96
A.1	The effect of the assumed vaccine uptake on the outbreak risk	108
A.2	The effect of immunity due to prior infections on the outbreak risk	109
A.3	The effect of the assumed effectiveness of two vaccine doses on the outbreak risk	110
A.4	Sensitivity analysis of the outbreak risk to model parameters	111
A.5	The effect of individual level variation in transmission between different infected individuals on the IOR following the completion of the vaccination programmes	113
B.1	Sensitivity analysis of the latent and infectious period parameters in the 1-group and 2-group models	114
B.2	The primary results of the study, repeated with half the value of the cross-immune fraction ν at the beginning of the 2009 epidemic compared to the value assumed in the main text	116
B.3	The primary results of the study, repeated with double the value of the cross-immune fraction ν at the beginning of the 2009 epidemic compared to the value assumed in the main text	117
B.4	Residuals of the 1-group and 2-group models fitted to data from the 2009 H1N1 influenza epidemic in Japan	119
B.5	The age distribution of the population of Japan in 2009 and the age distribution of the probability that an individual in Japan dies within one year, based on mortality and population data from 2009	124
B.6	The weekly number of recorded, unrecorded, and total (recorded and unrecorded) cases as estimated by the 2-group model with mean estimated parameters, fitted to data from the 2009 H1N1 influenza epidemic in Japan . . .	125
B.7	Forecasts obtained from the 1-group and 2-group models of epidemics occurring 25, 50, and 75 years after the 2009 epidemic as in Figure 3.5, but with variances of the gamma distributions of β and N five times larger	127
B.8	Forecasts obtained from the 1-group and 2-group models of epidemics occurring 25, 50, and 75 years after the 2009 epidemic as in Figure 3.5, but with variances of the gamma distributions of β and N five times smaller	128

B.9	The cross-immune fraction of the 2-group model, the basic reproduction numbers of the 1-group and 2-group models, the total number of recorded cases, the duration of the epidemic, the maximum number of weekly recorded cases, and the time at which this maximum occurs, if a second epidemic were to occur at a given time after the 2009 epidemic	130
B.10	Forecasts obtained from the 1-group and 2-group models of a future epidemic occurring 25 years after the 2009 epidemic in real-time using information from the 2009 epidemic as in Figure 3.6, but with variances of the gamma distributions of β and N five times larger	131
B.11	Forecasts obtained from the 1-group and 2-group models of a future epidemic occurring 25 years after the 2009 epidemic in real-time using information from the 2009 epidemic as in Figure 3.6, but with variances of the gamma distributions of β and N five times smaller	132
C.1	eFAST and Sobol sensitivity analyses of the IR model	136
C.2	Profile likelihood (χ^2_{PL}) curves for each fitting parameter of the IR model, when data for V and C are observed	144
C.3	Profile likelihood (χ^2_{PL}) curves for each fitting parameter of the IR model, when data for V and X are observed	144
C.4	Profile likelihood (χ^2_{PL}) curves for each fitting parameter of the IR model, when data for C and X are observed	145
C.5	Profile likelihood (χ^2_{PL}) curves for each fitting parameter of the IR model, when data for V are observed	145
C.6	Profile likelihood (χ^2_{PL}) curves for each fitting parameter of the IR model, when data for C are observed	146
C.7	Profile likelihood (χ^2_{PL}) curves for each fitting parameter of the IR model, when data for X are observed	146
D.1	Scatter plots of posterior distributions of ω and η and ω and $\tilde{\eta} = \eta/\omega$ when the SD model is fitted to data on the viral load, innate immune response (cytokines), and symptom score	148
D.2	Infection-related quantities of the SD model as parameters ω and η are varied, and ω and $\tilde{\eta} = \eta/\omega$ are varied	148
D.3	The synthetic data relating to 12 previously observed healthy and 12 previously observed comorbid patients used in Figures 5.5–5.8	149

Chapter 1

Introduction

1.1 Public health impacts of viral pathogens

Viral infectious diseases have caused a significant global burden throughout history. Numerous epidemics have occurred around the world caused by a multitude of viral pathogens. In recent years these include HIV/AIDS since 1981 [50], the 2002-2004 SARS outbreak [13], the 2012 MERS outbreak [114], the 2013-2016 Western African Ebola virus epidemic [331], and the 2015-2016 Zika virus epidemic [232].

Most notably since the turn of the 20th century, the 1918 Spanish Flu influenza pandemic has been estimated to have caused 50 million deaths [193]. As viral agents adapt and evolve, new variants emerge [222], and increasing international travel provides the opportunity for pathogens to spread [33, 178]. This has been seen starkly over the past year by the worldwide impact of the COVID-19 pandemic caused by the SARS-CoV-2 virus, which as of June 2021 has caused over three million deaths [390]. The burden is not limited to mortality, with epidemics also putting substantial pressures on healthcare systems, and non-pharmaceutical interventions such as national lockdowns causing disruption to the daily lives of individuals [14]. Secondary effects of the COVID-19 pandemic, and the interventions introduced to reduce its transmission, have been seen in trade, unemployment, education, and mental health [2].

In this thesis we shall focus on influenza viruses and SARS-CoV-2 as they have caused the largest pandemics in recent history. Influenza is a highly contagious disease caused by influenza viruses. Seasonal strains are estimated to cause approximately three to five million severe cases and result in 290,000 to 650,000 respiratory deaths annually worldwide. The majority of these deaths have been in high-risk groups, such as young children, the elderly, and individuals with chronic health conditions [187, 389]. These strains cause epidemics during the winter months in temperate regions in the northern and southern parts of the world, while epidemics may occur at any time of the year near the equator [389]. Considerably more lethal strains emerge periodically, giving rise to global pandemics in 1918, 1957, and 1968 [202].

In 2009, reassortment of North American and Eurasian swine viruses generated a new strain of H1N1, triggering the first influenza pandemic of the 21st century [75, 372]. The virus is believed to have originated in Mexico in April 2009, and then spread quickly across

the globe, reaching 43 countries by May that year [119, 372]. The case fatality rate was lower than that of previous pandemics in the 20th century [195]. However, the scale of the pandemic, with estimates that 11%–21% of the global population contracted the virus, led to pressures on healthcare systems [200].

In December 2019, a new coronavirus (later named SARS-CoV-2) emerged in the city of Wuhan in China [280]. It spread rapidly throughout Wuhan, and then across the world, with over 170 million cases reported as of June 2021 [390]. Many countries were unprepared for a global pandemic, with poor initial surveillance and control strategies leading to large case numbers [224], and healthcare systems being overwhelmed requiring restrictions on travel and ‘stay-at-home’ orders to be put in place [115, 215, 219]. While the pandemic remains ongoing, governments around the world face numerous challenges such as deciding when to lift restrictions and balancing the risks of cases against economic damage and non-COVID-19 health harms. With the emergence of effective vaccines, there are promising signs that the pandemic is starting to be controlled in many countries [81, 380]. However, as vaccines do not completely prevent transmission [157, 199], the virus may continue to spread even if severe disease and mortality is reduced.

This thesis considers epidemiological modelling questions at the population and host level scales related to how model projections of influenza and SARS-CoV-2 disease dynamics can be improved. In this chapter we provide a background to the role of mathematical modelling in infectious disease forecasting and an overview of viral emergence, immunity, and viral infection. We then review the literature on population level epidemic models and within-host infection models. We describe the data we use in our research, parameter estimation techniques which are a key component of the methodology, and then outline the research content of the thesis.

1.2 Role of mathematical modelling

Mathematical modelling is an important tool for making projections about the dynamics of an epidemic, guiding intervention strategies, and developing an understanding of the epidemiological and immunological processes which drive transmission and infection [176]. Recently, mathematical models have helped guide interventions for COVID-19 [134, 364]. They have been used to determine when interventions should be put in place and when they should be relaxed [110]. They are being used to inform decisions on how to exit lockdowns [367], and make long-term forecasts of the transmission dynamics through the postpandemic period [206]. Similarly, mathematical models have been used to guide decision making and policy to control influenza pandemics [24, 177, 179, 229, 376], and other viral diseases such as Ebola [223] and SARS [146]. Models also play an important role in determining optimal vaccine strategies; such as whether to prioritise vaccinating high-risk individuals to directly protect

them from severe disease or death, or vaccinating low-risk individuals who may be more likely to contribute to transmission to stop the spread of the virus [30, 51, 98, 246].

As we shall explore in Chapter 2, in the context of the ongoing COVID-19 pandemic, mathematical models can be used to determine the risk that an outbreak will occur following the introduction of a small number of infected individuals into a susceptible population [49, 83, 214]. In Chapter 3, we investigate making real-time forecasts of the number of cases during an epidemic. Much attention has been paid to these (often short-term) forecasts which are updated with new data, such as during the 2009 H1N1 influenza pandemic [278, 369], for seasonal influenza epidemics [41, 337], and currently for the COVID-19 pandemic [325]. Using historical data, models can also be used to provide insight into past epidemics, to better understand pathogen transmission and the impact of intervention strategies [99, 161].

However, mathematical models are not only used to explore transmission dynamics at the population scale. Within-host models, describing the viral dynamics in an infected individual, can aid understanding of epidemiological processes and analyses of experimental results [298]. They can elucidate the role of the immune response in clearing a viral infection [77, 93, 150, 153, 344, 397], and be used to make predictions about the differences in viral kinetics between strains [293, 346] as well as the emergence of pandemic strains [322]. Furthermore, models can be used to help assess the efficacy of antiviral treatments [32, 56, 92, 152, 194, 220, 259]. In Chapters 4 and 5 we will show how models can be used for forecasting purposes, to predict the dynamics of the viral load and the severity of symptoms.

Making accurate projections using mathematical models is challenging [324]. It is impossible to capture every detail in an infection process, and thus simplifying assumptions must be made [12]. Compromises have to be made between the level of detail included, and hence a model's ability to reproduce and forecast infectious disease dynamics accurately - at either a population or within-host level - and its transparency and utility, that is, how well we can understand the interplay between various model components. We may be able to improve the accuracy of a model by including increasing levels of epidemiological detail, but we may sacrifice the ability to qualitatively understand the system's behaviour and estimate model parameters accurately when fitting models to data.

1.3 Viral emergence and immunity

There are four types of influenza viruses, A, B, C, and D, with influenza A viruses being responsible for most human infections and severe illness. Influenza viruses can also infect a range of animal species, including birds and pigs, and new zoonotic strains may spread from animals to humans [386]. Zoonotic influenza appears to be transmitted from animals to humans via direct contact with pigs for swine flu, and exposure to contaminated bird droppings for avian influenza [388]. Influenza A is classified into subtypes, for example H1N1 or H3N2, based on the surface proteins haemagglutinin (HA) and neuraminidase (NA). Immunologically,

HA and NA are critical for virulence as they are targets for neutralizing antibodies of acquired immunity. HA binds to susceptible cells causing cellular infection, and NA frees the virus to infect other cells, resulting in the spread of the virus within a host [145].

It is thought that SARS-CoV-2, the virus responsible for the COVID-19 pandemic, has zoonotic origins, having emerged from a bat-borne coronavirus [300, 403]. Ongoing research seeks to determine whether the virus came directly from bats or through an intermediate host [357]. The virus belongs to the same group of coronaviruses that caused SARS in 2003 [295] and MERS in 2012 [407]. New variants of the SARS-CoV-2 virus have emerged since its initial discovery, and there is evidence that some of these new variants spread more easily and are associated with an increased risk of mortality [63, 116]. So far though, studies suggest that current vaccines confer protection against these new variants [104, 238], although they do not completely prevent transmission [157, 199]. Booster vaccines to increase protection against newly emerging variants may be required [237, 382]. Furthermore, assessing the effects of waning immunity is an important area for future research [169].

Influenza A viruses mutate over time, either due to antigenic drift, where a small mutation in the genes of the virus leads to changes in the surface proteins, producing a closely related strain, or, antigenic shift, an abrupt, major change in the virus due to recombination and reassortment, resulting in new surface proteins [46, 204]. A new form of the virus that has emerged due to drift is likely to have similar antigenic properties to the original virus and, thus, individuals who were infected with the original virus may have acquired some immunity to the altered virus. However, when antigenic shift occurs, many individuals are likely to have little or no immunity to the new virus (depending on their lifetime history of infection). Due to the random nature in which new influenza strains seem to emerge, it is not currently possible to predict when future pandemics will occur, and which strains will cause these pandemics [272].

Previous exposure to an influenza virus confers homosubtypic cross-immunity, that is cross-immunity to strains of the same subtype, due to cross-reactive neutralizing antibodies [142, 210, 211, 288]. These neutralizing antibodies, predominantly against HA proteins on the surface of the virus, block viral entry to the host cell [182, 355]. It has also been suggested that cross-protective non-HA antibody immunity may lead to reduced disease severity [109]. NA specific antibodies have also been shown to reduce influenza viral titres, morbidity, and virus shedding [79, 106, 264, 267, 335, 383]. As well as homosubtypic immunity, there is evidence of heterosubtypic immunity, namely cross-protection generated by prior exposure to a different but related influenza subtype [117, 142, 273]. Cross-immunity - either homosubtypic or heterosubtypic - may explain why there has not been a global influenza pandemic as severe as the 1918 pandemic in the last century [361].

SARS-CoV-2 has been able to spread so rapidly because it is a novel virus, hence individuals have little pre-existing immunity to the disease. As of June 2021, 17 vaccines have been

administered across the globe [245]. In the UK, the elderly, health and social care workers, and individuals with pre-existing health conditions have been prioritised for vaccination [88]. In general, many vaccines work by leaving the host with a supply of ‘memory’ T-cells which fight pathogens as well as B-cells which produce antibodies to neutralize the virus. If the host then encounters the virus again, these T- and B-cells quickly act to attack the virus and infected cells. Typically, it takes a few weeks after vaccination for the host to produce these T- and B-cells. For the majority of COVID-19 vaccines, two doses are required, administered at least three weeks apart, as it has been shown that having a second dose results in a stronger immune response than having a single dose [379]. However, a single dose has been shown to confer some protection compared to being unvaccinated [380].

1.4 Epidemiology of viral infection and outbreaks

Influenza is a short-lived infection; its viral kinetics in humans are characterised by rapid exponential growth over one to three days, to a peak viral load, followed by a decline over the subsequent week [60]. In uncomplicated cases, the virus infects epithelial cells in the respiratory tract of the host [276]. Symptoms include high fevers, coughs, headaches, muscle and joint pains, a sore throat, and a runny nose. The majority of individuals recover within a week [389], however in high-risk groups the duration of infection and symptoms may be prolonged by several days or even weeks [321, 375]

The infection lifecycle begins when the pathogen enters the body. The primary mechanism of transmission is via respiratory droplets released when a nearby infected individual coughs or sneezes [203]. Less commonly, transmission may occur when an individual touches a surface which has the pathogen on it and then touches their mouth, nose, or eyes.

Viral replication begins when a virion (virus particle) binds to a susceptible target-cell (Figure 1). The virus then replicates within the nucleus of the cell [95]. After a period of between five and twelve hours new virions are released [31, 339], and infect other cells, leading to the initial exponential growth of infected cells and virions. Infection results in cellular dysfunction and apoptosis (programmed cell death) [111]. Free virions outside cells also lose infectivity and are eliminated by non-specific mechanisms such as mucociliary clearance [345].

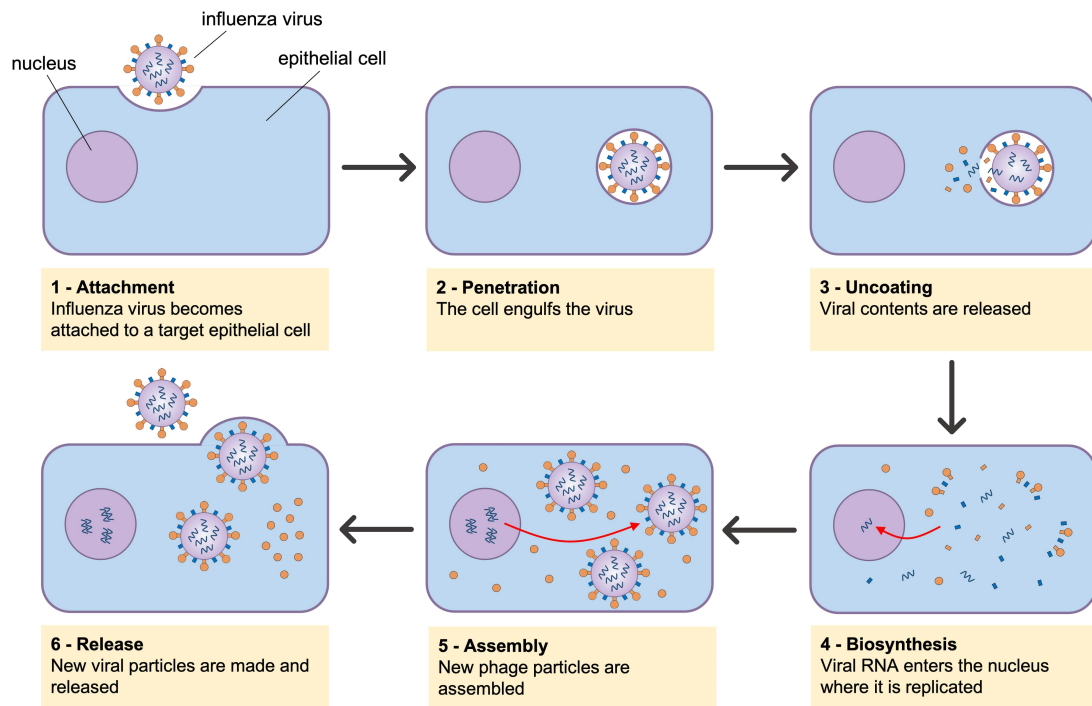


Figure 1.1: A diagram of influenza viral infection at the cellular level. Redrawn from Lumen Learning [234].

The immune response to viral infections can be divided into the innate and adaptive responses [349]. The innate immune response is non-specific, limiting pathogen replication and spread in the body and priming the adaptive immune response. The secretion of proinflammatory cytokines, particularly type I interferon, induces an antiviral state in susceptible cells and further activates signalling pathways and cellular responses. Inflammatory mediators in the innate immune response cause systemic symptoms which may persist even after the virus has been primarily cleared and viral load levels are low [87]. The innate response also includes immune cells such as Natural Killer cells [188].

The adaptive immune response is specific to the pathogen and can be characterised as either cellular or humoral. In humans it takes four to seven days for the adaptive immune response to take effect after being stimulated by a pathogen [192]. Cellular adaptive immunity is mediated by T-cells; $CD4^+$ cells which ‘help’ other immune cells by releasing cytokines, and cytotoxic $CD8^+$ cells which clear infected cells through several mechanisms, including lysis, the induction of apoptosis, and the production of pro-inflammatory cytokines [141, 183]. The humoral adaptive immune response controls infection via antibodies secreted by B-cells, which recognise and bind to specific epitopes on the virus. These antibodies (the most prominent being IgA, IgG, and IgM) bind to the HA protein of the influenza virus and inhibit the virion’s attachment and entry into a target-cell, thus neutralising the virion [269]. The immune response upon infection with influenza A is influenced by a multitude of factors such as age

[155, 216], and comorbidities including obesity [180] and heart disease [184]. A schematic of the immune response timeline is given in Figure 1.2.

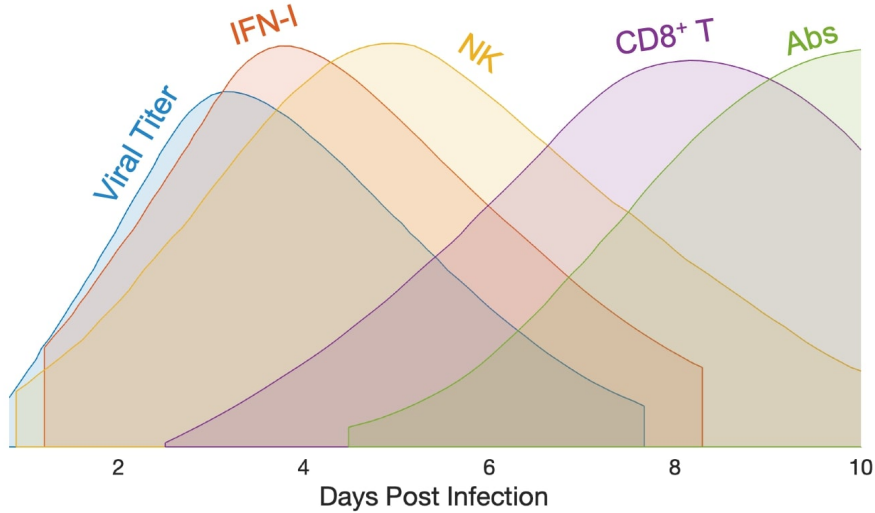


Figure 1.2: Influenza immune response timeline. The innate immune response, primary components being type I interferons (IFN-I, signalling proteins belonging to the family of cytokines) and Natural Killer (NK) cells, is first to act. The adaptive immune response, primarily driven by $CD8^+$ T-cells and antibodies (Abs) then helps to clear the infection. Redrawn from Boianelli et al. [45].

The time between infection with an influenza virus and the development of symptoms, known as the incubation period, is typically one to four days. However, viral shedding may occur before the onset of symptoms, resulting in an individual being infectious [144], and so, presymptomatic transmission may occur, leading to the undetected spread of the virus [291]. Shedding lasts for five to ten days in most individuals. However, viral shedding may persist in young children or immunocompromised individuals for weeks or even months [7].

When a novel pathogen, such as a new strain of influenza, is introduced into a population, there is a lack of immunity in the host population. Initially, due to the random nature of transmission, there is a chance that extinction will occur, that is, only a small number of individuals will become infected, before the pathogen dies out [8]. However, if an outbreak does occur and the number of infected individuals crosses a certain threshold, the number of cases will increase exponentially due to the large pool of susceptible individuals, in a manner analogous to the spread of the pathogen within the host. If no interventions are put in place, the number of cases will continue to rise until the peak of the epidemic, after which, herd immunity has been reached and the number of cases declines.

1.5 Infectious disease models

1.5.1 Population level epidemic models

One of the first epidemic models was derived in 1927 by Kermack and McKendrick [201]. The model describes pathogen transmission through a population, laying the foundation of grouping individuals into compartments based on their infection or symptom status. Many new models retain the same compartmental structure as we shall illustrate in this thesis.

Commonly, when modelling the spread of influenza, an ‘SEIR model’ is used to describe transmission in which individuals are either Susceptible, Exposed (i.e. infected but not yet infectious), Infectious, or Recovered [12]. A schematic of the model and an example of population level epidemic data which can be described by the model are given in Figure 1.3.

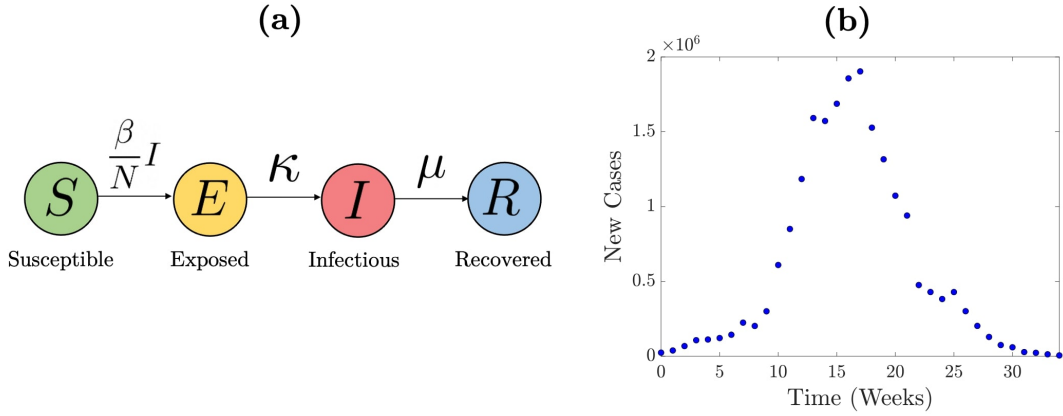


Figure 1.3: (a): Schematic of the SEIR model. (b): Example of population level epidemic data of the number of weekly cases during the 2009 H1N1 influenza epidemic in Japan [278], which we shall consider in Chapter 3.

The model, when formulated as a system of time-dependent ordinary differential equations (ODEs), is given by

$$\text{susceptible individuals} \quad \frac{dS}{dt} = -\frac{\beta}{N}SI, \quad (1.1)$$

$$\text{exposed individuals} \quad \frac{dE}{dt} = \frac{\beta}{N}SI - \kappa E, \quad (1.2)$$

$$\text{infectious individuals} \quad \frac{dI}{dt} = \kappa E - \mu I, \quad (1.3)$$

$$\text{recovered individuals} \quad \frac{dR}{dt} = \mu I, \quad (1.4)$$

where β is the transmission rate, $1/\kappa$ the latent period between being exposed and becoming infectious, $1/\mu$ the infectious period, and N the population size.

Infectious disease outbreak models, such as the SEIR model, assume that individuals have similar characteristics [37, 72]. More complex models account for differences between individuals. For example, in many studies that aim to determine optimal vaccination strategies,

populations are split into low-risk and high-risk groups [98, 125], and spatial heterogeneity can be incorporated by classifying individuals according to their location [230, 283]. Commonly, to account for age-dependent differences in contact rates and fatality rates, age-structured models are used to compartmentalise groups of individuals into distinct age-group classes [73, 85, 133, 208, 246, 364]. Models of influenza transmission that incorporate evolutionary and multi-strain dynamics [97, 166, 398], as well as phylodynamics (accounting for temporal changes in pathogen phylogenies) have also been considered [121, 143, 209]. Models in which populations are structured according to whether or not individuals carry pre-existing immunity can also be formulated by compartmentalising individuals based on their level of immunity [15, 243, 297, 320, 361].

Infectious disease models are often formulated as time-dependent systems of ordinary differential equations, as in equations (1.1)–(1.4). Differential equation models describe the mean-field dynamics of the underlying system (i.e. what would happen ‘on average’) and are deemed appropriate to use when they model a large number of infected individuals. Such an approach may be relevant when simulating a nationwide epidemic or pandemic, as we shall consider in Chapter 3 where we model an influenza epidemic in Japan. At the population scale when the number of infected individuals is large, stochastic effects underpinning transmission and disease progression do not dominate the dynamics of the outbreak.

Stochastic models can also be formulated to describe epidemics. They are commonly formulated as discrete- or continuous-time Markov chains or stochastic differential equations [8, 47]. Microscale stochastic agent-based models can even be used to describe interactions between individuals in a small system [299]. Most simply, stochastic branching process models can be used, where the number of infected individuals at each time is calculated as the sum of all the new ‘offspring’ infected individuals at previous times [9]. We consider this type of model in Chapter 2, where we investigate whether importing an infectious individual into a disease-free location will lead to an outbreak. Examples of stochastic models used to describe epidemics include [265, 278, 395] for influenza and [1, 167, 404] for COVID-19.

A key difference between stochastic and deterministic models is that stochastic models provide more information about the variation of potential outcomes of the epidemic. In deterministic models, an outbreak will always occur if the basic reproduction number, R_0 , the average number of secondary cases of disease arising from a primary case in an entirely susceptible population, is greater than one [170]. This is because each infected individual generates more than one other secondary case. However, importantly, as we shall explore in Chapter 3, in a stochastic model an outbreak can become extinct even if $R_0 > 1$ due to the randomness of the underlying epidemiological process. Thus, if modelling the early stages of an outbreak, or an outbreak in a small population, it may be necessary to use a stochastic model.

We show the results of simulations of an epidemic using an ordinary differential equation and stochastic discrete-time Markov chain SEIR model, modelling an epidemic in a small population of 1,000 individuals (Figure 1.4 (a)) and a large population of 1,000,000 individuals (Figure 1.4 (b)). In panel (a) we see significant variation in the stochastic trajectories, and in many cases the outbreak becomes extinct. In such cases using a stochastic model is important to capture the variation in the dynamics. However, in panel (b), because there is a large population and initially a relatively large number of infected individuals, there is little variation in the stochastic trajectories, they are all similar to the ODE trajectories.

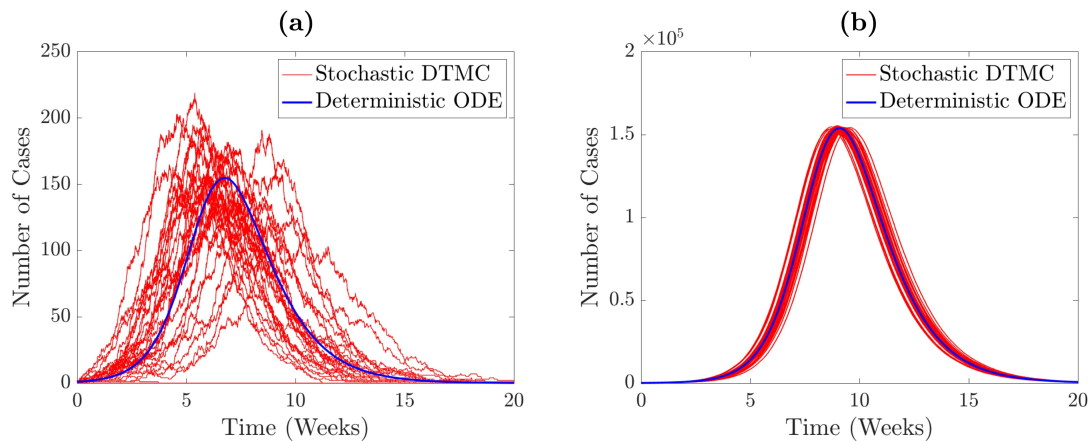


Figure 1.4: Comparison of SEIR epidemic models in small and large populations with different initial conditions. Simulations of an ordinary differential equation (ODE) model (equations (1.1)–(1.4)) and 50 sample paths of a stochastic discrete-time Markov chain (DTMC) model. Parameter values are $\beta = 2 \text{ weeks}^{-1}$, $\kappa = 0.5$ and $\mu = 1 \text{ weeks}^{-1}$ in each panel, with $N = 10^3$, $I_0 = 1$ in (a) and $N = 10^6$, $I_0 = 100$ in (b). The number of cases given in this figure defined as the number of currently infected individuals, I .

1.5.2 Within-host models

More recently, mathematical models have been used to describe the within-host viral dynamics of cells and virions in an analogous manner to population scale epidemic models, inspired largely by Nowak and May [282] and Perelson [298]. Cells can be classified as susceptible (uninfected), latently infected, productively infected, and removed (dead) at a within-host level in the same way that individuals can be classified at a population level. A schematic of such a model is presented in Figure 1.5. The model, which is known as the TEIV model, also accounts for the number of free virions. Also included in Figure 1.5 are typical viral titer data.

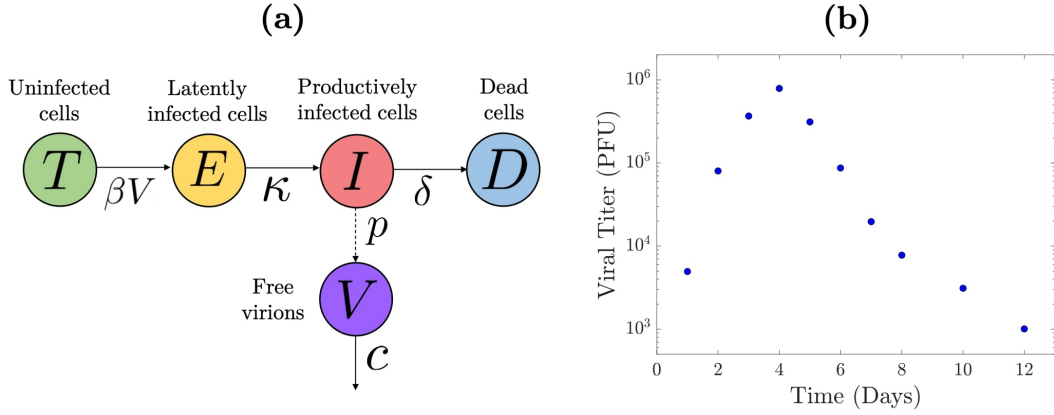


Figure 1.5: (a): Schematic of the TEIV model. (b): Example of within-host infection data of the viral titer (corresponding to V) in mice infected with influenza A [189], which we shall consider in Chapter 4.

The TEIV model is given by the set of ordinary differential equations:

$$\text{uninfected cells} \quad \frac{dT}{dt} = -\beta TV, \quad (1.5)$$

$$\text{latently infected cells} \quad \frac{dE}{dt} = \beta TV - \kappa E, \quad (1.6)$$

$$\text{productively infected cells} \quad \frac{dI}{dt} = \kappa E - \delta I, \quad (1.7)$$

$$\text{dead cells} \quad \frac{dD}{dt} = \delta I, \quad (1.8)$$

$$\text{free virions} \quad \frac{dV}{dt} = pI - cV, \quad (1.9)$$

where β is the infection rate, p is the virion production rate, $1/\kappa$ is the duration of the latent eclipse phase, $1/\delta$ is the duration of the productively infectious phase, and c is the clearance rate of free virions.

Within-host influenza modelling is a younger field compared to population scale influenza epidemic modelling. However, the works of Baccam et al. [19] which adopted the target-cell model, previously used to describe other viral infections such as HIV and hepatitis B [298, 323] has led to the rapid development of this field.

As we explain below, these models have improved our understanding of within-host epidemiological mechanisms, and can be used to describe and forecast a patient's disease and immunological dynamics, and to identify therapeutic targets to control infection. The full interplay between the virus and a host's immune response is not completely understood though.

Within-host epidemiological models can broadly be classified into two types. First, target-cell limited models, in which the infection is limited by the availability of susceptible cells [19, 91, 147, 152, 342]. Second, models can explicitly account for the immune response at varying levels of detail [44, 55, 58, 150, 172, 220, 252, 292, 330, 371]. Some models focus on one aspect of the immune response (for instance type I interferons [19, 330]), while others describe

the multitude of mechanisms involved in the immune response [220]. Due to the availability of data, the quantity of interest is typically viral load, although symptom dynamics have also been considered [55, 56, 57, 233, 240]. In Chapter 5 we investigate whether a simple differential equation model can describe the temporal changes in the severity of symptoms in different patients. Mathematical modelling has also been used to assess the efficacy of antiviral treatments such as neuraminidase inhibitors and adamantanes [32, 56, 92, 152, 194, 220, 259]. Models also exist which describe co-infection with influenza and another pathogen which may be another virus [303, 304, 305], or bacteria such as *Streptococcus pneumoniae* [67, 343, 347]. These models usually comprise ordinary differential equations, however stochastic Markov chain models have also been considered [23, 152, 168, 294, 305, 396]. In Figure 1.6, we present a schematic diagram which shows how cells and virions interact in the target-cell TEIV model. We also present typical model simulations showing the characteristic exponential increase and then decrease in viral load.

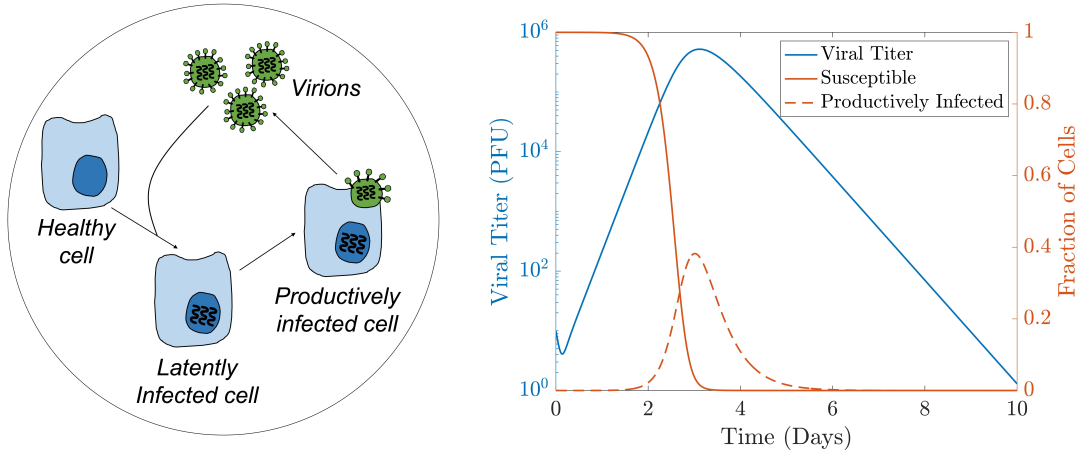


Figure 1.6: Left: Schematic of within-host infection. Right: Simulation of the TEIV target-cell model (equations (1.5)–(1.9)), displaying the viral titer, and fraction of susceptible cells and productively infected cells. Parameter values are $\beta = 1.5 \times 10^{-5} \text{ PFU}^{-1} \text{ days}^{-1}$, $p = 2 \times 10^{-3} \text{ PFU days}^{-1} \text{ cells}^{-1}$, $\delta = 2 \text{ days}^{-1}$, $c = 10 \text{ days}^{-1}$, $\kappa = 4 \text{ days}^{-1}$, $V_0 = 10 \text{ PFU}$, and $T_0 = 10^9$.

1.6 Data and parameter estimation

1.6.1 Data

To make forecasts and quantify infectious disease dynamics, it is important that models are informed by real-world data. Many different types of data may be utilised in epidemiological modelling studies. At the population level, data on the number of daily or weekly cases, hospitalisations, deaths, and vaccinations are often available. At the individual host level, viral titer data are commonly recorded, as well symptom severity and concentrations of components

of the immune response. Host level data, such as the duration of latent or infectious periods, can also inform population level models.

In Chapter 2, we utilise data on the number of vaccine doses administered in Israel and the Isle of Man to estimate the susceptibility to SARS-CoV-2 infection of each population [185, 286]. In Chapter 3, we use data relating to the 2009 H1N1 influenza epidemic in Japan, specifically the number of weekly cases seeking medical attention [278]. We use data relating to the viral titer, fraction of dead cells, and immune response of both immunocompetent and immunosuppressed mice infected with influenza in Chapter 4 [189]. In Chapter 5, we use data of the symptom scores, viral titer, and immune response in humans infected with influenza [120].

In each of Chapters 3–5, we use data first to calibrate models by estimating parameter values, and then to generate synthetic data. The generation of synthetic data is a valuable tool as we can describe how a model behaves in a range of different scenarios and epidemiological cases [360]. For example, in Chapter 4 we model within-host infection dynamics in hosts with different immunological characteristics. In Chapter 3 we model an epidemic spreading through a heterogeneous population, with different subpopulations having different levels of susceptibility. As data are generated from a model which has been well-parameterised, we can assume that they reflect real-world data to some extent. With synthetic data it is possible to control the quality of data used in modelling studies, and to make predictions about situations for which real-world data are currently unavailable. Large test and training datasets can be built using synthetic data, and it is often more convenient to carry out sensitivity analyses using synthetic data [4]. Studies have also used semi-synthetic datasets which combine real-world data with synthetic outbreaks [52]. Depending on the aim of the study, synthetic data can be used to replace real-world data, identify the data that need to be collected to resolve uncertainties, or be used as a proxy to accelerate research and produce findings when real-world data are scarce [137].

1.6.2 Parameter estimation

In order to make projections and understand epidemiological and immunological processes, parameter values must be estimated by fitting models to data. In this thesis, we use two methods to estimate parameter values using time series data: maximum likelihood estimation (MLE) which provides a point parameter estimate [6, 113], and Markov chain Monte Carlo (MCMC) estimation, a Bayesian approach which generates distributions of parameter values [130].

Both methods require a likelihood function. Under this assumption that the differences between the data and model forecasts are normally distributed and are independent of the

size of the data, this likelihood function takes the form

$$\mathcal{L}(y|\theta) = \prod_{j=1}^n \exp\left(\frac{-(z(t_j; \theta) - y_j)^2}{2\sigma^2}\right), \quad (1.10)$$

where $z(t_j, \theta)$ are model outputs generated using parameters θ , y_j are the observed data, and σ is a noise scaling parameter. By taking a logarithm of the likelihood function (1.10), the cost function can be derived,

$$J(\theta; y) = \sum_{j=1}^n A(y_j - z(t_j; \theta))^2. \quad (1.11)$$

In equation (1.11) the parameter A is a function of σ which we need not consider since it is constant and therefore does not influence the cost function, and can without loss of generality be set equal to 1. Hence, the MLE point estimate can be determined by solving the least squares problem

$$\hat{\theta} = \arg \min_{\theta \in \Theta} J(\theta; y), \quad (1.12)$$

where Θ is the feasible set of parameter values.

By treating the fitting parameters θ as random variables, Bayesian parameter inference can provide a probabilistic distribution of θ , given observed data y , known as the posterior distribution $f(\theta|y)$. Bayesian inference relies on Bayes' theorem, which states the relationship between the posterior distribution, prior distribution $\pi(\theta)$, and likelihood function

$$f(\theta|y) \propto \mathcal{L}(y|\theta)\pi(\theta). \quad (1.13)$$

We can exploit equation (1.13) to determine the posterior distribution using only the likelihood function and prior distribution. Prior information of the parameter distribution $\pi(\theta)$ may be incorporated by, for instance, knowing epidemiological constraints or using parameter estimates based on historical data. If no prior knowledge is known, we use a uniform prior, fixing $\pi(\theta)$ as constant. Often, due to the number of unknown parameters and computational cost, it is difficult to sample the posterior $f(\theta|y)$ directly. In such situations, Markov chain Monte Carlo (MCMC) numerical estimation methods can be used. These methods construct a Markov chain, with $f(\theta|y)$ as a stationary distribution. A sequence of random samples taken from the chain can be used to estimate the posterior distribution $f(\theta|y)$. Throughout this thesis, we use the Metropolis-Hastings algorithm to perform MCMC parameter estimation [160, 250, 285, 353].

In Figure 1.7 we present an illustrative example of an SEIR ordinary differential equation model (equations (1.1)–(1.4)) fitted to synthetic data, using both MCMC and MLE.

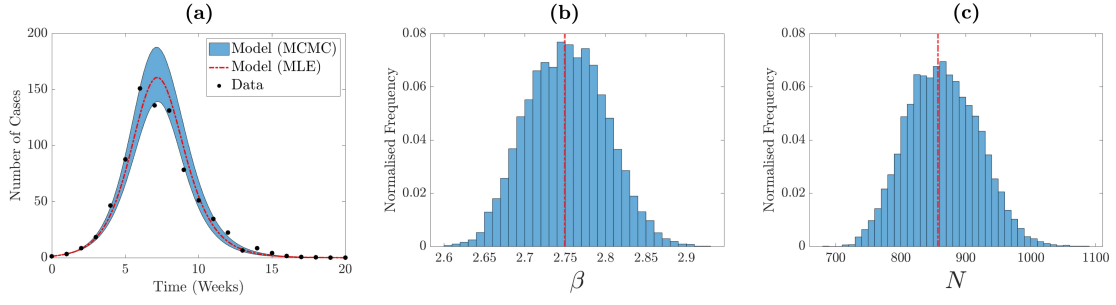


Figure 1.7: SEIR model (equations (1.1)–(1.4)) fitted to synthetic epidemic case data using MLE and MCMC. The shaded region in (a) indicates the central 95% credible intervals of model trajectories based on the posterior distributions of parameters, and the red solid line indicates the model trajectory with point estimate MLE parameters. Posterior fitting parameter distributions estimated using MCMC of β and N shown in (b) and (c), with MLE parameter estimates shown by vertical red lines. We performed 10^5 MCMC sampling iterations, discarding the first 10^4 iterations as the ‘burn-in’ period and recording every 100 iterations thereafter to reduce autocorrelation. Uniform priors $\beta \in [1, 10]$ and $N \in [100, 100,000]$ were used.

1.7 Thesis outline

This thesis addresses epidemiological modelling questions at the population and within-host scales. By focusing on problems relating to immunity and within-host immunological dynamics, we aim to establish how the accuracy of model projections of infectious diseases can be improved. The main body of work in this thesis is contained in Chapters 2 to 5. Chapters 2 and 3 are concerned with modelling the spread of pathogens, SARS-CoV-2 and influenza, at a population level, while Chapters 4 and 5 involve modelling influenza infection dynamics within a host. The work in Chapters 4 and 5 was done in conjunction with our industrial partner (Biosensors Beyond Borders) with the intention of investigating within-host models which describe infection dynamics in different groups of patients. An outline of the models, data, and pathogens considered in each chapter is given in Table 1.1.

In Chapter 2, we consider the spread of SARS-CoV-2. We use a stochastic model to determine the probability that infected individuals introduced into low prevalence settings initiate an outbreak. We consider Israel and the Isle of Man as case studies, since they have few COVID-19 cases currently, and are actively vaccinating their populations. Using a range of mathematical methods, including the ‘Numerical Outbreak Risk’ which we conceive in this chapter, we determine the risk of an outbreak occurring given that an infected individual enters the country. Our results suggest that when travel restrictions are relaxed in low prevalence settings, it will still be necessary to implement surveillance of incoming passengers to identify infected individuals quickly until global case numbers are reduced. This piece of work has been accepted for publication in *Communications Medicine* (<https://doi.org/10.1101/2021.05.21.21257589>).

Chapter	Scale	Model	Virus	Data	Model Output
2	Population	Stochastic branching process	SARS-CoV-2	COVID-19 vaccination doses in Israel and the Isle of Man	Outbreak risk
3	Population	Deterministic ODEs	Influenza A	2009 H1N1 influenza cases in Japan	Infected individuals
4	Within-host	Deterministic ODEs	Influenza A	Murine viral titer, dead cells, and immune response levels	Viral titer
5	Within-host	Deterministic ODEs	Influenza A	Human symptom scores, viral titer, and immune response levels	Symptom score

Table 1.1: A summary of the content of Chapters 2 to 5 in this thesis.

In Chapter 3, we focus on forecasting influenza epidemics. Previous exposure to influenza viruses confers cross-immunity against future infections with related strains. However, this effect is not always included in mathematical models used for forecasting during influenza outbreaks. We consider two infectious disease outbreak forecasting models: the ‘1-group model’ which does not account for cross-immunity and the ‘2-group model’ which does. We fit both models to estimated case notification data from Japan during the 2009 H1N1 influenza pandemic, and then generate synthetic data for a future outbreak by assuming that the 2-group model more accurately represents the epidemiology of influenza infections. In this scenario, the 1-group model only produces accurate outbreak forecasts once the peak of the epidemic has passed. The main contribution of this work is to show that more a epidemiologically realistic model, in which cross-immunity is represented explicitly, can generate more accurate projections during outbreaks than if cross-immunity is not included. This piece of work has been published in *Epidemics* (<https://doi.org/10.1016/j.epidem.2020.100432>).

We model influenza at the within-host rather than the population scale in Chapter 4. We consider a hierarchy of models which describe the kinetics of influenza infection: the ‘IR model’ explicitly accounts for the immune response to the virus, whereas the simpler, target-cell limited TEIV and TV models do not. We show that we can parametrise the IR model using available experimental murine data on viral titer, dead cells, and immune response levels, which substantiates the results of complementary structural and practical identifiability analyses. We use the IR model to generate synthetic data relating to infections in hosts who differ in the competence of their immune responses. We fit the TV model to these synthetic data, and show that, in many cases, it can describe the characteristic exponential increase and decay of viral load generated by the IR model. Furthermore, we demonstrate that the parameters of the TV model vary systematically as the immune response parameters in the

IR model used to generate the data are changed. This piece of work is being prepared for submission to the *Journal of Theoretical Biology*.

In Chapter 5, we extend the models from Chapter 4 by introducing an additional variable which describes the severity of symptoms experienced by a host, to formulate the ‘SD model’. We fit this model to pooled human data on viral titer, innate immune response levels, and symptom scores, estimating the model parameters in the process, and for the first time fitting a model to these three sets of data. We then generate synthetic data relating to both healthy patients and a hypothetical population of comorbid patients and assess the forecasting accuracy of a two-variable model, the ‘TS model’, that we derive from the TV model used in Chapter 4. The TS model describes how the fraction of susceptible cells and the severity of symptoms change during infection. We investigate our ability to use the TS model to make real-time forecasts which can be updated daily. We also investigate whether it is possible to classify patients by fitting the TS model to their symptom data. Finally, in Chapter 6, we discuss the contributions of the results obtained in the thesis and make suggestions for future work.

Chapter 2

The risk of SARS-CoV-2 outbreaks in low prevalence settings following the removal of travel restrictions

2.1 Introduction

Combinations of non-pharmaceutical interventions (NPIs) such as travel restrictions and national lockdowns have been introduced worldwide to counter the COVID-19 pandemic [69, 80, 83, 221, 227, 311]. The development and deployment of vaccines has also lowered transmission [157, 326, 336] and reduced the number of individuals experiencing clinical symptoms or severe disease once infected [21, 307, 379].

Effective NPIs have led to low levels of transmission in some locations, for example in Australia and New Zealand. With travel restrictions, short-term lockdowns, quarantine of inbound travellers, and contact tracing to identify infected contacts whenever cases have been discovered, the spread of the pathogen has been suppressed in these regions [25, 26, 181]. Rigorous interventions targeted at inbound travellers in low prevalence settings reflect the fact that imported cases can contribute significantly to local incidence [328], and potentially initiate outbreaks with substantial local transmission [171, 365]. At the time of writing (1st May 2021), Israel is the country that has vaccinated the largest proportion of its citizens, and the vaccination campaign there has been credited with reducing transmission [326], prompting some NPIs to be removed.

Despite the success of both NPIs and vaccines, the current overall picture is complicated. Vaccines do not prevent transmission entirely [157, 199], and vaccine uptake is incomplete, particularly in some ethnic groups and in underserved communities [318]. Current tentative estimates suggest that first doses of the Pfizer and AstraZeneca SARS-CoV-2 vaccines are around 60% effective at preventing infection, with the second doses leading to 65-85% efficacy against infection [199]. Furthermore, the appearance of novel SARS-CoV-2 variants makes eliminating transmission more challenging. For example, the Alpha variant (B.1.1.7, or VOC 202012/01) that first appeared in the United Kingdom in late 2020 has been found to be more

transmissible than the SARS-CoV-2 virus that originally emerged in China [86]. Although public health measures have had some successes, these concerns raise a key question: Can NPIs such as travel restrictions be removed without any risk in low prevalence settings where vaccines have been distributed widely, or might outbreaks occur initiated by SARS-CoV-2 reimportations from elsewhere?

Epidemiological models are often used to assess the risk of outbreaks in scenarios in which the potential for pathogen transmission is not changing. According to the mathematical theory of branching processes, the probability that cases introduced into a new host population generate an outbreak driven by sustained local transmission is given by

$$\text{Prob}(\text{outbreak}) = 1 - \left(\frac{1}{R}\right)^{I_0}, \quad (2.1)$$

in which R is the reproduction number of the pathogen and I_0 is the number of introduced infectious cases. This expression had been used to assess the risks of outbreaks of pathogens including the Ebola virus [11, 249, 366], before being applied early in the COVID-19 pandemic to assess outbreak risks outside China [365]. Equation (2.1) reflects the risk that introduced cases generate an outbreak, and is based on a simple underlying transmission model in which individuals mix homogeneously and infected individuals are infectious for exponentially distributed time periods. Equation (2.1) also involves an assumption that pathogen transmissibility is fixed at its current level. In other words, the value of R is implicitly assumed to remain constant over the initial phase of the potential outbreak. With a background of rapidly changing population immunity due to vaccination acting to reduce transmission, this assumption may not be accurate. To assess the risk of outbreaks if NPIs are removed during an ongoing vaccination campaign using branching process models, standard epidemiological modelling theory must be extended to account for temporally changing population immunity.

Here, we use branching processes to investigate whether or not an introduced case is likely to initiate an outbreak, accounting fully for temporal changes in population immunity due to an ongoing vaccination campaign. We use four metrics to assess the risk of an outbreak and consider two examples of vaccination campaigns from around the world, from the Isle of Man and the country of Israel. In both locations, vaccination is progressing quickly and prevalence is currently low. Given the relatively low numbers of cases in these locations during the pandemic, there is also likely to be a background of limited immunity from previous infections. We assess the risk of outbreaks in these places when travel restrictions and other NPIs are removed, considering scenarios in which NPIs are removed at different stages of the vaccination campaigns. Our goal is not to develop a detailed epidemiological model for estimating precise outbreak risk estimates. Instead, we use a simple model to show generally that, even when all vaccines have been deployed in low prevalence settings (i.e. a maximum vaccination proportion ν of individuals have been vaccinated, where the precise value of ν varies between locations), we project that the combination of incomplete vaccine uptake,

imperfect vaccination and variants of concern means that the risk of outbreaks due to imported cases will not be eliminated completely when NPIs are removed. This highlights the need for careful monitoring of imported cases until global prevalence is reduced to low levels. Until vaccines are rolled out worldwide, there is still a risk of local transmission arising in low prevalence settings initiated by importations from elsewhere.

2.2 Methods

2.2.1 Epidemiological model

We performed our main analyses using a stochastic branching process model that describes virus transmission in the initial stages of a potential outbreak. In this model, following the arrival of a case in the local population, new infections occur at rate $\beta(1 - \Lambda(t))$ and infected individuals have a mean infectious period of $1/\mu$ days (Figure 2.1 (a)). The times between successive events in the model are assumed to follow exponential distributions. The function $\Lambda(t)$ reflects the extent to which transmission has been reduced by vaccination, where a value of $\Lambda(t) = 0$ corresponds to an entirely unvaccinated population. We consider scenarios in which travel restrictions and other NPIs are removed, and transmission is only limited by vaccine-acquired immunity.

To model an ongoing vaccination campaign, we set $\Lambda(t) = \eta_1 V_1(t - \alpha) + \eta_2 V_2(t - \alpha)$, where $V_1(t)$ is the proportion of individuals in the population who have received a single vaccine dose at time t and $V_2(t)$ is the proportion of individuals in the population who have received two vaccine doses at time t (Figure 2.1 (b)). The parameters η_1 and η_2 reflect the effectiveness of the vaccine at preventing infection after one and two effective doses, respectively (specifically, η_1 and η_2 represent the multiplicative reduction in susceptibility following one or two effective doses, compared to an unvaccinated individual, so that $0 \leq \eta_1 \leq \eta_2 \leq 1$), and the parameter α represents the delay between a vaccine dose being administered and being effective in the recipient. In our main analyses, since we are modelling relatively low prevalence settings, we do not consider immunity due to prior infections, although we present a supplementary analysis in which we considered the robustness of our results to this assumption in Appendix A.

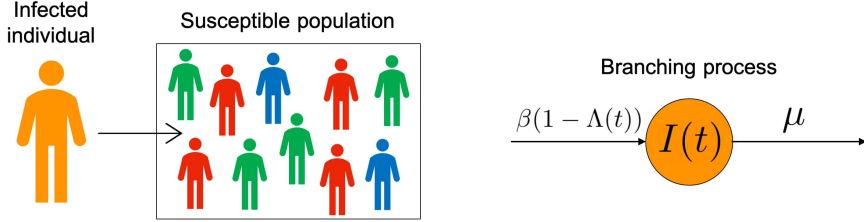
Under this model, the time-dependent reproduction number, accounting for any vaccines that have been administered and are effective at time t , is given by

$$R_V(t) = \frac{\beta(1 - \Lambda(t))}{\mu}. \quad (2.2)$$

The value of $R_V(t)$ represents the expected number of secondary infections generated by a single infected individual in the population at time t , under the assumption that no further vaccinations take place in future. It is therefore representative of the ‘instantaneous’ transmissibility at time t . Indeed, $R_V(t)$ is sometimes referred to as the instantaneous reproduction

number [118, 363, 385]. In the absence of vaccination, so that $\Lambda(t) = 0$, then $R_V(t)$ is equal to the basic reproduction number, R_0 .

(a) **Transmission model**



(b) **Vaccination**

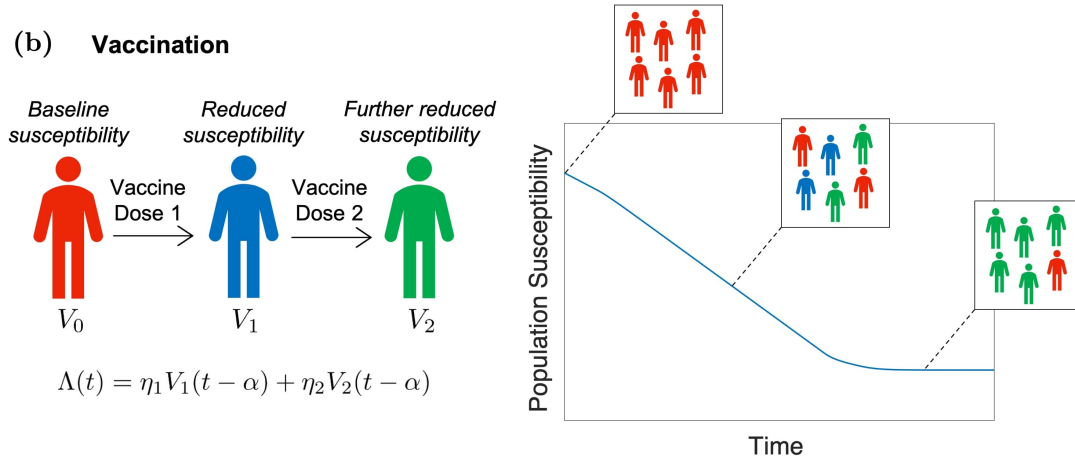


Figure 2.1: The epidemiological model used in our analyses. (a): Following the introduction of an infected individual into the host population, local transmission may happen with infections occurring at rate $\beta(1 - \Lambda(t))$, where $\Lambda(t)$ reflects the vaccination coverage in the local population. Infected individuals have a mean infectious period of $1/\mu$ days. (b): The vaccination process is modelled by setting $\Lambda(t)$ according to the proportion of individuals in the population who have been vaccinated with one (V_1) or two (V_2) doses. The first vaccine dose is assumed to have effectiveness η_1 and the second vaccine dose has effectiveness η_2 . Vaccine doses are effective α days after they are administered. This leads to declining population susceptibility as a vaccine is rolled out across the population.

2.2.2 Vaccination data

In the model, $V_1(t)$ and $V_2(t)$ were set based on vaccination data from the location under consideration (either the Isle of Man or Israel). Data describing the proportion of the total population who had received one or two vaccine doses were available for the periods up until 11th April 2021 (for the Isle of Man [185]) and 21st April 2021 (for Israel [286]).

To explore how the risk of outbreaks is likely to change in future, we projected the vaccine rollout forwards beyond these dates in the following way. We considered the total population size of the location under consideration (denoted N), as well as the numbers of individuals ($N_1(t)$ and $N_2(t)$) vaccinated with one or two doses, so that $V_1(t) = N_1(t)/N$ and $V_2(t) =$

$N_2(t)/N$. We assumed that a constant number of vaccine doses are available each day in future (denoted D), and that there is a target period of τ days between each vaccine dose. On any day in future, each available dose is assigned to an individual who has been vaccinated with their first dose at least τ days ago, with remaining doses then assigned to unvaccinated individuals. Resulting values of $N_1(t)$ and $N_2(t)$ were then converted to corresponding values of $V_1(t)$ and $V_2(t)$. To reflect the fact that vaccine uptake is imperfect, we assumed that a maximum proportion ν of the population can ever be vaccinated. Consequently, once the vaccination programme is complete, then $\Lambda(t) = \eta_2\nu$. The value of ν varies between locations, and depends on factors such as the age distribution of the host population. Since there is uncertainty in the vaccine uptake going forwards, we conducted supplementary analyses in which we considered a range of different values of ν (Appendix A). Values of the model parameters used in the analyses in the main text for the Isle of Man and Israel are shown in Table 2.1.

2.2.3 Outbreak risk metrics

We used four different metrics to assess the risk that an infected individual, introduced into the population at time t , initiates an outbreak driven by sustained local transmission. The values of each of these metrics vary temporally. Each of these metrics is used to estimate the risk that an imported case initiates an outbreak under the assumption that all travel restrictions and other NPIs are removed (in other words, pathogen transmissibility at time t is determined by $R_V(t)$, which represents an adjustment to the basic reproduction number R_0 due to vaccine-acquired immunity, as described above). Under the assumption of the branching process model, the distribution of epidemic sizes is bimodal for $R_V(t) > 1$, where either the disease dies out with few individuals ever becoming infected, or it becomes widespread in the population [366]. An outbreak describes the latter, where the disease becomes widespread, which we define to be 100 individuals infected at one time when a practical threshold is required, a value chosen which is a significant number of cases in relation to the populations of both the Isle of Man and Israel.

2.2.3.1 Instantaneous Outbreak Risk (IOR)

The first metric we considered is the Instantaneous Outbreak Risk (IOR). The IOR represents the risk that an imported case initiates an outbreak as opposed to the virus simply fading out, under the assumption that pathogen transmissibility is fixed at its value when the imported case enters the population. If the introduction occurs at time t , when the infection rate is $\beta(1 - \Lambda(t))$ and the time-dependent reproduction number is $R_V(t) = \beta(1 - \Lambda(t))/\mu$, then the outbreak probability can be derived by first denoting by $q_{I(t)}$ the probability that an outbreak does not occur starting from $I(t)$ infected individuals. Conditioning on whether the first event

Parameter	Description	Value	Source
β	Transmission rate parameter	Set so that $R_0 = \beta/\mu$ takes a prescribed value	Two scenarios were considered, with mean R_0 values of 3 (sim. to original SARS-CoV-2 variant [42]) and 5 (sim. to Alpha variant (B.1.1.7) [86])
$1/\mu$	Duration of infectiousness	5 days	[212]
α	Delay between a vaccine dose being administered and being effective	14 days	[62]
η_1	Multiplicative reduction in susceptibility of an individual vaccinated with a single effective dose (compared to an unvaccinated host)	0.6	[199]
η_2	Multiplicative reduction in susceptibility of an individual vaccinated with two effective doses (compared to an unvaccinated host)	0.85	[199]
ν	Vaccine uptake	0.8 (Isle of Man) and 0.7 (Israel). Other values are considered in Appendix A	[332]
N	Total population size	84,500 (Isle of Man) and 8,772,800 (Israel)	[391]
D	Number of vaccine doses available each day (in future projections)	1,000 (Isle of Man) and 56,000 (Israel)	Average number of doses administered each day in the previous 30 days (up to 11th April 2021 for Isle of Man [185] and 21st April 2021 for Israel [286])
τ	Target period between vaccine doses (in future projections)	84 days (Isle of Man) and 21 days (Israel)	[186, 387]

Table 2.1: Default parameter values used in our analyses.

is an infection or a removal event gives

$$q_{I(t)} = \frac{\beta(1 - \Lambda(t))}{\beta(1 - \Lambda(t)) + \mu} q_{I(t)+1} + \frac{\mu}{\beta(1 - \Lambda(t)) + \mu} q_{I(t)-1}. \quad (2.3)$$

In particular, starting from a single introduction at time t ,

$$q_1 = \frac{\beta(1 - \Lambda(t))}{\beta(1 - \Lambda(t)) + \mu} q_2 + \frac{\mu}{\beta(1 - \Lambda(t)) + \mu} q_0. \quad (2.4)$$

Since infections occur according to a branching process, $q_2 = q_1^2$. Furthermore, an outbreak will not occur if there are no infected individuals (i.e. $q_0 = 1$), so that

$$\beta(1 - \Lambda(t))q_1^2 - (\beta(1 - \Lambda(t)) + \mu)q_1 + \mu = 0. \quad (2.5)$$

Solving this quadratic equation, and taking the minimal non-negative solution (as dictated by Markov chain theory when calculating the probability that a branching process fades out [281]) leads to the IOR,

$$\text{IOR} = 1 - q_1 = \begin{cases} 0, & \text{for } R_V(t) \leq 1, \\ 1 - \frac{1}{R_V(t)}, & \text{for } R_V(t) > 1. \end{cases} \quad (2.6)$$

This quantity represents the expression shown in equation (2.1), with $R = R_V(t)$ and $I_0 = 1$. The IOR reflects the risk of an outbreak occurring starting from a single infected individual at time t , but under the assumption that the vaccine rollout does not continue after time t so that pathogen transmissibility is unchanged. While the IOR is straightforward to calculate, it does not reflect changing population immunity due to vaccination over the initial phase of the potential outbreak. This standard metric is often used to assess the risk of outbreaks in scenarios where pathogen transmissibility does not vary temporally [11, 49, 197, 249, 362, 365, 366].

2.2.3.2 Case Outbreak Risk (COR)

The second metric is the Case Outbreak Risk (COR). The COR is an extension of the IOR, accounting for changes in population susceptibility, $\beta(1 - \Lambda(t))$, due to vaccination over the initial phase of the potential outbreak. The COR has previously been used to assess outbreak risks using branching processes for models in which pathogen transmission varies periodically [17, 27, 49, 59].

We denote the probability that an outbreak does not occur starting from I infected individuals in the population at time t by $q_I(t)$. Then, starting from a single introduction into the population at time t and conditioning on the possible events in the next Δt days (where Δt represents a short time period, so that at most a single event can occur in the interval $[t, t + \Delta t)$) gives

$$q_1(t) = \beta(1 - \Lambda(t))\Delta t q_2(t + \Delta t) + \mu\Delta t q_0(t + \Delta t) + (1 - \beta(1 - \Lambda(t))\Delta t - \mu\Delta t)q_1(t + \Delta t). \quad (2.7)$$

Since infections occur according to a branching process, $q_2(t + \Delta t) = q_1(t + \Delta t)^2$. Furthermore, an outbreak will not occur if there are no infected individuals (i.e. $q_0(t + \Delta t) = 1$), so that

$$q_1(t) = \beta(1 - \Lambda(t))\Delta t q_1(t + \Delta t)^2 + \mu\Delta t + (1 - \beta(1 - \Lambda(t))\Delta t - \mu\Delta t)q_1(t + \Delta t). \quad (2.8)$$

Rearranging this expression, and taking the limit $\Delta t \rightarrow 0$, gives

$$\frac{dq(t)}{dt} = \beta(1 - \Lambda(t))q(t)(1 - q(t)) + \mu(q(t) - 1), \quad (2.9)$$

in which the variable $q_1(t)$ has been replaced by $q(t)$ for notational convenience. This equation can be solved numerically, after which the COR at time t (i.e. the probability that an outbreak occurs) is given by $1 - q(t)$.

Solving equation (2.9) numerically requires the value of $q(t)$ to be known at a single time-point. In scenarios in which the time-dependent reproduction number is above one when the vaccine programme is completed, we solve equation (2.9) backwards in time starting from the final condition $q(t^*) = \frac{1}{R_V(t^*)}$, where t^* is the time at which the vaccination programme ends. The rationale for this choice is that the COR then matches the IOR in scenarios in which virus transmissibility does not change in future (e.g. the end of the vaccination programme). For scenarios in which $R_V(t)$ is instead below one when the vaccination programme is completed, we set $q(\tilde{t})$ so that $1 - q(\tilde{t})$ matches the Numerical Outbreak Risk (NOR) (see below) at some time $t = \tilde{t}$ where $R(\tilde{t}) > 1$. We solve equation (2.9) backwards and forwards in time starting from $t = \tilde{t}$. In our analyses we set $\tilde{t}$ such that $R(\tilde{t}) = 1.05$. This value of $\tilde{t}$ was chosen such that this gives close agreement between $1 - q(t)$ and the NOR.

2.2.3.3 Simulated Outbreak Risk (SOR)

The third metric we considered is the Simulated Outbreak Risk (SOR). The SOR involves repeated simulation of the branching process model, using the direct version of the Gillespie stochastic simulation algorithm [131] adapted to account for temporally varying pathogen transmissibility (due to changing population immunity).

To simulate the branching process model, we follow the following steps:

1. Set the initial time t , and set $I(t) = 1$.
2. Calculate the time of the next event, $t + \tau$, using the expression

$$\int_t^{t+\tau} (\beta(1 - \Lambda(s)) + \mu)I(s) ds = -\ln(r_1), \quad (2.10)$$

where r_1 is a random number sampled from a uniform distribution on $(0, 1)$.

3. Determine whether the next event is an infection event or a removal event. To do this, sample a second random number (r_2) from a uniform distribution on $(0, 1)$. If

$$r_2 < \frac{\beta(1 - \Lambda(t + \tau))}{\beta(1 - \Lambda(t + \tau)) + \mu}, \quad (2.11)$$

then the next event is an infection event; set $I(t + \tau) = I(t) + 1$. If instead the inequality above is not satisfied, then the next event is a removal event; set $I(t + \tau) = I(t) - 1$.

4. Repeat steps 2–3 while the outbreak is still ongoing and the threshold value of $M = 100$ has not been hit (i.e. $I(t) > 0$ and $I(t) < M$).

Simulations were run starting with a single infected individual introduced into the population at time t . The SOR then corresponds to the proportion of simulations in which a local outbreak occurs; an outbreak is said to occur if the total number of individuals infected simultaneously exceeds a pre-defined threshold value, M . In our analyses, we set $M = 100$.

2.2.3.4 Numerical Outbreak Risk (NOR)

Finally, we consider the Numerical Outbreak Risk (NOR). The NOR is analogous to the SOR, but with the advantage that it can be calculated without performing model simulations. The NOR, therefore, also represents the risk that a single infected individual introduced into the population at time t initiates an outbreak in which at least $M = 100$ individuals are ever infected simultaneously. To calculate the NOR, we define the vector $p(t) = (p_0(t), \dots, p_{M-1}(t), p_M(t))^T$, where, for $i \leq M - 1$, $p_i(t)$ represents the probability that $I(t) = i$ and that the number infected has not reached threshold M by time t . The variable $p_M(t)$ is defined as the probability that $I(t)$ has reached threshold M by time t . As with the SOR, we again use a value of $M = 100$ in our analyses. Then, the Kolmogorov forward equations [198] are

$$\frac{dp(t)}{dt} = Q(t)p(t), \quad (2.12)$$

where

$$Q(t) = \begin{pmatrix} 0 & \mu & 0 & \dots & \dots & 0 \\ 0 & -(\beta(1 - \Lambda(t)) + \mu) & 2\mu & \dots & \dots & 0 \\ 0 & \beta(1 - \Lambda(t)) & -(2\beta(1 - \Lambda(t)) + 2\mu) & \dots & \dots & 0 \\ 0 & 0 & 2\beta(1 - \Lambda(t)) & \dots & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \vdots & (M-1)\mu & 0 \\ 0 & 0 & 0 & \vdots & -((M-1)\beta(1 - \Lambda(t)) + (M-1)\mu) & 0 \\ 0 & 0 & 0 & \vdots & (M-1)\beta(1 - \Lambda(t)) & 0 \end{pmatrix}. \quad (2.13)$$

In this matrix, the state $I(t) = M$ is an absorbing state (analogous to stopping the simulations as soon as $I(t)$ reaches the threshold M when calculating the SOR). The probability that there are at least M infected individuals at some stage following the introduction of a single infected host into the population at time t is then given by $\text{NOR}(t) = \lim_{s \rightarrow \infty} (p_M(s))$, where $p(t) = (0, 1, 0, \dots, 0, 0)^T$. The NOR is therefore calculated by solving the Kolmogorov forward equations numerically until the system converges to an equilibrium state.

We use the IOR to show the potential error in using a metric that only takes into account the instantaneous reproduction number without accounting for future vaccinations, and the

SOR, COR, and NOR to validate the results of the outbreak risk when temporal changes in the reproduction number are considered.

2.3 Results

2.3.1 Modelling the vaccination process and population susceptibility

As described in the Methods, we first generated projections of the number of vaccinated individuals in future for the Isle of Man (Figure 2.2 (a)) and Israel (Figure 2.2 (d)), based on past vaccination data in those locations. To explore the impact of vaccination on virus transmission, we calculated the time-dependent reproduction number ($R_V(t)$; equation (2.2)) throughout the vaccination campaigns. We considered two different scenarios. In the first, we set the mean value of R_0 (i.e. the reproduction number in the absence of vaccination) equal to 3, as was the situation early in the COVID-19 pandemic (Figure 2.2 (b),(e)). In the second scenario, we set the mean value of R_0 equal to 5 (Figure 2.2 (c),(f)) to reflect the fact that currently circulating SARS-CoV-2 variants are more transmissible than the original virus [86]. This second scenario is therefore likely to be a more realistic reflection of the current and future risk.

2.3.2 Modelling the risk of an outbreak

We then calculated the values of the four different outbreak risk metrics throughout the period considered (18th December 2020 to 20th August 2021) based on these vaccination projections (Figure 2.3). This involves a scenario in which NPIs are removed entirely, so that $R_V(t)$ is not reduced by interventions other than vaccination. These metrics then reflect the risk that a single case first entering into the population at the date of introduction shown initiates an outbreak driven by sustained local transmission, given that no NPIs are in place. For each metric, and each time during the vaccination programme, we calculated the outbreak risk by integrating over the full distribution for $R_V(t)$ shown in Figure 2.2. In practice, the precise value of $R_V(t)$ is not known exactly: the purpose of this integration is to account for this uncertainty in outbreak risk estimates. The resulting outbreak risk therefore represents a point estimate of the risk accounting for uncertainty in pathogen transmissibility.

The IOR represents the probability of sustained local transmission based only on the value of $R_V(t)$ at the precise instance that the virus is introduced into the population. When an introduction occurs while a vaccine is being deployed, there is a background of decreasing population susceptibility, and so $R_V(t)$ may in fact decrease over the initial stages of a potential outbreak. For that reason, it is unsurprising that the IOR sometimes overestimates the outbreak risk compared to the other risk metrics (see e.g. black lines in Figure 2.3 (a),(c) – and see Discussion). We note though that for large values of $R_V(t)$, for example $R_V(t) = 5$, either the outbreak occurs rapidly or the disease dies out early on, and so future vaccinations

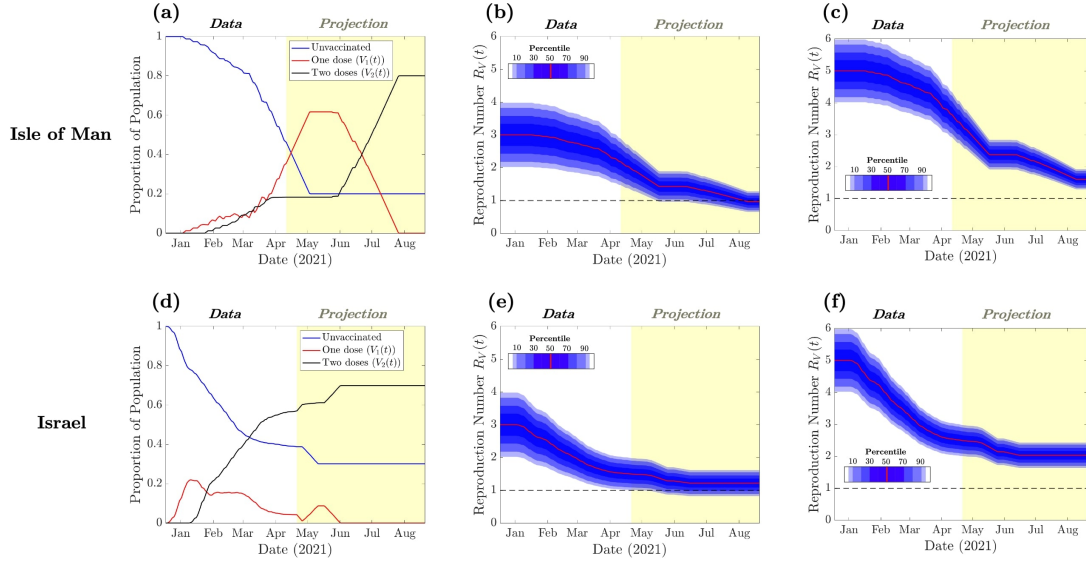


Figure 2.2: Effect of vaccination on population susceptibility. (a): The proportion of the population of the Isle of Man who are unvaccinated ($1 - V_1(t) - V_2(t)$), vaccinated with a single dose ($V_1(t)$) and vaccinated with two doses ($V_2(t)$). The period in which vaccination data were available is shown in white, and the period in which vaccination data were projected is shown in yellow. (b): The time-dependent reproduction number ($R_V(t)$) corresponding to the vaccination data in panel (a), starting from a mean initial value of $R_V(0) = R_0 = 3$. To account for uncertainty in the value of R_0 , a normal distribution was assumed about the mean value of R_0 with variance $\sigma^2 = 0.25$, which is reflective of the range of R_0 values estimated early in the COVID-19 pandemic [42]. (c): Analogous to panel (b) but starting from a mean initial value of $R_V(0) = R_0 = 5$. (d)-(f): Analogous to panels (a)-(c) but using vaccination data for Israel. In all panels, $t = 0$ days corresponds to 18th December 2020. Ticks on the x-axes refer to the starts of the months labelled. Parameter values are shown in Table 2.1.

do not have much of an effect on the outbreak risk, so the IOR and other outbreak metrics are in close agreement. It is when $R_V(t)$ is only slightly greater than 1, for example $R_V(t) = 1.1$, where it might take a number of weeks to reach 100 cases and so future vaccinations have more of an impact, where the IOR gives a significant overestimate of the outbreak risk.

In contrast, we found close agreement between the COR, SOR and NOR. Due to the high assumed vaccine uptake in the Isle of Man, the outbreak risk at the end of the vaccination programme there was calculated to be lower than when the vaccine rollout was completed in Israel (although we also considered supplementary analyses with different assumed vaccine uptake values – Figure A.1). In the first scenario that we considered (mean $R_0 = 3$), which is representative of the transmissibility of the original SARS-CoV-2 virus, the outbreak risk was projected to be low following the vaccination programme in the Isle of Man.

However, when the virus was assumed to be more transmissible (mean $R_0 = 5$), as is the case for variants of concern such as the Alpha variant (B.1.1.7), the outbreak risk was found to be substantial even following the projected end of the vaccination campaigns in both the Isle of Man and Israel. For example, when the assumed vaccine uptake values of

$\nu = 0.8$ and $\nu = 0.7$ had been achieved in the Isle of Man and Israel, respectively, the NOR took values of 0.37 (95% Credible Interval (CrI): [0.22, 0.48], calculating using the 95% CrI for $R_V(t)$ at the end of the vaccination rollout) and 0.51 (95% CrI: [0.39, 0.59]) in those locations. Given the uncertainty in any epidemiological modelling projection, the main take-home message is simply that these values are greater than zero (rather than the exact outbreak risk estimates themselves), suggesting that outbreaks can still occur when NPIs are removed following the completion of vaccination programmes. This result is robust to a range of assumptions surrounding virus transmission and the effectiveness of vaccines (see Appendix A).

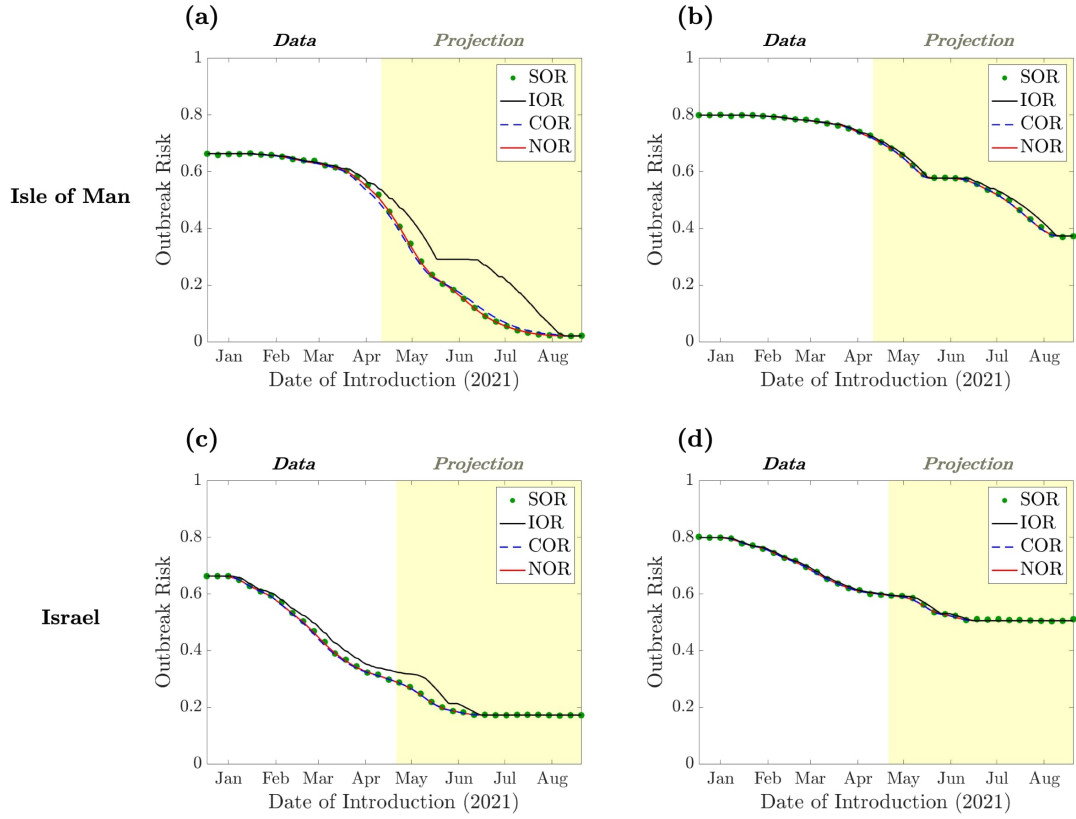


Figure 2.3: The risk that an infectious case introduced at each stage of the vaccination campaign initiates an outbreak if travel restrictions and other NPIs are removed. (a): The outbreak risk assessed using the four metrics (IOR – black; COR – blue dashed; SOR – green dots; NOR – red), based on vaccination data from the Isle of Man and assuming mean $R_0 = 3$, as in Figure 2.2 (b). The period in which vaccination data were available is shown in white, and the period in which vaccination data were projected is shown in yellow. (b): Analogous to panel (a) but with mean $R_0 = 5$ (see Figure 2.2 (c)). (c): Analogous to panel (a) but using vaccination data for Israel (see Figure 2.2 (e)). (d): Analogous to panel (a) but using vaccination data for Israel with mean $R_0 = 5$ (see Figure 2.2 (f)). Ticks on the x-axes refer to the starts of the months labelled. Parameter values are shown in Table 2.1.

2.4 Discussion

As vaccines are administered in countries around the world, attention has turned to the possibility that transmission will soon be reduced to the extent that travel restrictions and other NPIs can be relaxed. Here, we have investigated the impact of removing NPIs on the risk of outbreaks occurring in locations with low prevalence and a significant proportion of the population vaccinated. We used four metrics to estimate the risk that a case introduced at any stage in the vaccination rollout leads to an outbreak driven by local transmission, in a scenario in which NPIs are removed. We calculated temporal changes in the values of these metrics in the context of vaccination in the Isle of Man and in Israel, two locations with low prevalence and vaccination campaigns that have progressed quickly.

We found that vaccination is leading to a substantial drop in the potential for virus transmission in both the Isle of Man and Israel, as indicated by a decreasing value of the time-dependent reproduction number, $R_V(t)$ (Figure 2.2). However, even when the vaccine rollout is completed (with 80% of individuals fully vaccinated in the Isle of Man, and 70% in Israel; other values are considered in Appendix A), the combination of vaccines not preventing transmission entirely, incomplete vaccine uptake and the emergence of novel SARS-CoV-2 variants suggests that the risk of outbreaks initiated by infected individuals arriving from elsewhere will not be eliminated when NPIs are removed (Figure 2.3). This conclusion remains true unless the vaccine uptake is very high (Figure A.1). This suggests that, when NPIs such as travel restrictions are relaxed, it will still be advisable to be aware of the potential for local transmission. Ensuring that case numbers are reduced elsewhere (i.e. in locations that imported cases might travel from) will reduce the risk of importations, and strategies should be considered to suppress outbreaks quickly if importations occur.

One potential use of our modelling framework by policy-makers is to identify the dates on which travel restrictions can be relaxed, based on a maximum acceptable outbreak risk. As an example, if the maximum acceptable value of the NOR is 0.4, then our analysis with $R_0 = 5$ suggests that travel restrictions can be lifted on the Isle of Man at the end of July 2021 (Figure 2.3 (b); the first date on which the NOR falls below 0.4 is 29th July 2021). However, as described above, since the risk of local transmission following introductions remains, it will be necessary to continue to monitor inbound passengers in low prevalence settings to identify infected individuals, even once vaccination programmes are complete.

Our modelling approach for assessing outbreak risks was motivated by studies in which the potential for pathogen transmission varies periodically [17, 18, 27, 49, 59, 190, 207, 277], due, for example, to seasonal changes in weather conditions that affect transmission. In that context, Carmona and Gandon [59] describe a ‘winter is coming effect’, in which the risk of an outbreak is lower than current environmental conditions suggest if conditions become less favourable for transmission in the near future. If current environmental conditions promote pathogen transmission, then the IOR is expected to be high. This is because the IOR reflects

the outbreak risk based on the conditions at the precise instance when the virus is introduced into the population. However, if environmental conditions are expected to become less favourable for transmission in the near future, then the values of the other risk metrics are lower than the IOR, since those metrics account for changes in transmissibility over the initial phase of the potential outbreak. Although in general we found a close agreement between the four metrics that we considered, a background of decreasing population susceptibility can lead to a similar effect in which the IOR is larger than the COR, SOR or NOR (e.g. Figure 2.3 (a),(c)).

In this chapter, we used a simple branching process model to investigate the risk of outbreaks when NPIs are removed during a vaccination programme. This involved considering whether introduced cases are likely to lead to sustained local transmission or instead fade out without causing an outbreak. We made the standard branching process modelling assumption that population immunity is unaffected by infections in the earliest stages of potential outbreaks [11, 231, 171, 249, 365, 366]. In other words, infection-acquired immunity following the arrival of the pathogen in the host population is not considered. While this is reasonable when case numbers are low in the initial stages of potential outbreaks, a more detailed model is needed to explore other quantities, such as the eventual size of outbreaks. Following a vaccination programme, outbreaks are likely to be smaller than those that occur before vaccines are widely administered.

Another simplification of our model is that we only accounted for changes in population susceptibility due to the vaccine rollout. We did not account for prior immunity of some members of the population due to previous exposures to the virus. At the time of writing (1st May 2021), there have been 1,154 confirmed cases in the Isle of Man and 838,000 confirmed cases in Israel. Since these case numbers correspond to a relatively small proportion of the host population (representing 1.4% and 9.6% of the population in the Isle of Man and Israel, respectively), we do not expect this assumption to affect our key findings. Furthermore, immunity is likely to wane over time [53, 206], reducing the effect of previous exposures on the outbreak risk. To test the potential impact of infection-acquired immunity arising from cases occurring before May 2021, we also conducted a supplementary analysis in which the value of R_0 is reduced by 1.35% in the Isle of Man and 9.55% in Israel, and we found qualitatively similar results (Figure A.2): even in this ‘best case’ scenario, there is still a risk of outbreaks due to imported cases once the vaccination programmes are completed. Importantly, in other countries in which higher numbers of cases have occurred, prior immunity may play a larger role in reducing the risk of outbreaks compared to the low prevalence settings considered here. Understanding the extent of this effect, based on the rate at which immunity wanes, is an important target for further study.

In this research, we assumed that vaccines reduce transmission by lowering the probability that a vaccinated host becomes infected compared to an unvaccinated host. Two effective

vaccine doses were assumed to reduce the probability of infection substantially (by $\eta_2 = 0.85$), which is in line with estimates for the Pfizer vaccine used in the UK Government’s COVID-19 roadmap modelling [199]. However, there is uncertainty about the precise level of protection offered by different COVID-19 vaccines. For that reason, we also conducted a supplementary analysis (Figures A.3 and A.4) in which we considered a range of values of η_2 . In every scenario that we explored, the risk of outbreaks following the completion of vaccination programmes in low prevalence settings was greater than zero, with larger outbreak risks for lower values of η_2 .

In addition to reducing the risk that a host becomes infected as assumed here [312], vaccination may also reduce the risk of onwards transmission following infection [157], reduce the risk of severe disease developing, or a combination of these effects [136]. In principle, it would be possible to develop a more complex model that accounts for all of these effects separately. It would also be possible to develop a transmission model in which either vaccine- and/or infection-acquired immunity wanes. There is currently significant uncertainty about the extent and speed of waning immunity, with hope that antibody responses and/or cellular immune responses may provide long-lasting immune memory, at least in some individuals [359, 408]. Booster vaccination for vulnerable individuals is currently being considered in several high-income countries, given the possibility that vaccine-acquired immunity may wane. We note that any waning of immunity will act to enhance our main qualitative conclusion that the risk of outbreaks is unlikely to be eliminated completely in low prevalence settings by vaccination programmes.

Further considerations for future analyses include other factors affecting transmission and vaccination, for example the impact of population-structure on both vaccine efficacy [51] and transmission [48, 84], or the possibility that individuals’ behaviour may be different following the pandemic compared to beforehand, even when NPIs and travel restrictions are removed. For precise quantitative outbreak risk predictions to be made, it may be necessary to estimate R_0 in different locations. It would also be necessary to incorporate changes in R_0 due to further new variants in our analyses, although we note that any increases in transmissibility will act to strengthen our main conclusion that vaccine-acquired immunity alone is unlikely to be sufficient to prevent outbreaks. Another important consideration is the degree of heterogeneity in transmission between different infected individuals. For a range of pathogens including SARS-CoV-2, the majority of transmissions are generated by a relatively small proportion of infectious hosts [105, 228, 367]. We therefore also conducted a supplementary analysis in which we considered the dependence of outbreak risk estimates following the completion of vaccination programmes in the Isle of Man and Israel on the assumed individual-level variation in SARS-CoV-2 transmission. While a high degree of heterogeneity (i.e. a higher probability of ‘superspreading’ events) acts to lower the outbreak risk, in each scenario explored outbreaks

were still possible in low prevalence settings following vaccination programmes when NPIs are removed (Appendix A.5).

To summarise, our analyses have demonstrated the general principle that, even following complete vaccination programmes in low prevalence settings, the risk of outbreaks remains when NPIs are removed. Our intention is not to argue that travel restrictions and other NPIs should not be relaxed once vaccination programmes are sufficiently advanced, but rather that measures should be taken to ensure that any clusters of cases are suppressed quickly if they arise. A local outbreak requires two steps: first, the virus must be imported from elsewhere; second, local transmission must occur. The first step emphasises the need for a global approach to minimising transmission, since large case numbers at a potential source location translates into a higher importation risk. The second step emphasises that, while vaccination acts to reduce transmission substantially, continued surveillance of inbound passengers for infection, combined with isolation and testing of contacts of detected infected individuals, is important when travel restrictions and other NPIs are relaxed. These measures are necessary, since only once the global prevalence of SARS-CoV-2 infections is reduced substantially can the risk of outbreaks in low prevalence settings be eliminated.

Chapter 3

Accounting for cross-immunity can improve forecast accuracy during influenza epidemics

3.1 Introduction

In 2009, an H1N1 influenza pandemic was caused by a new strain generated by a reassortment of North American and Eurasian swine viruses. Believed to have originated in Mexico, the virus then spread rapidly worldwide [119, 372]. While the case fatality rate was lower than previous influenza pandemics, an estimated 11-21% of the global population contracted the virus [200].

As detailed in Chapter 1, previous exposure to an influenza virus confers cross-immunity to antigenically related strains of the original virus [117, 142, 210, 211, 273, 288]. A significant proportion of elderly individuals carried pre-existing immunity (due to cross-reactive anti-HA antibodies) to the influenza strain that was responsible for the 2009 pandemic [28, 140, 151, 156, 393]. This may be due to the similarities between the 2009 H1N1 virus and the 1918 Spanish Flu virus, as viral descendants of the 1918 Spanish Flu virus continued to circulate until the 1957 pandemic [341, 394]. The suggestion that immunity was due to previous infections that occurred decades earlier implies that this form of cross-immunity is likely to be lifelong [399]. The consequences of pre-existing immunity can be seen in the age distribution of infected individuals in Japan in the 2009 pandemic, where only a small proportion of the individuals who sought medical attention were elderly [258]. A number of studies in animal models have also shown that prior infection with seasonal H1N1 strains have been able to provide substantial protection against infection with the 2009 H1N1 pandemic virus [102, 103, 196, 217]. As well as the heterogeneity between hosts in infection risk and age mentioned in Chapter 1, models in which populations are structured according to whether or not individuals carry pre-existing immunity can also be formulated [15, 243, 297, 320, 361].

In this chapter, our attention is directed towards how cross-immunity (either homosubtypic or heterosubtypic) affects the predictability of major epidemics. We use mathematical models

to investigate whether or not it is necessary to account for cross-immune individuals in the population when forecasting the dynamics of future influenza epidemics of antigenically related strains. We consider two epidemiological models. In the first, cross-immunity is ignored (the ‘1-group model’). In the second, more epidemiologically realistic model (the ‘2-group model’), individuals with and without cross-immunity are accounted for explicitly.

First, we estimate the values of the parameters of each model (specifically, the transmission rate and the effective population size) using data from the 2009 H1N1 influenza epidemic in Japan. We then consider a synthetic future influenza outbreak of an antigenically similar strain, simulated using the more epidemiologically realistic 2-group model. We explore whether or not accurate forecasts of this epidemic can be obtained in real-time. If uninformative priors are used and parameters are estimated in real-time, even the more realistic 2-group model is unable to generate accurate forecasts of the remainder of the epidemic before the peak occurs. This motivates us to incorporate information from the 2009 epidemic to set informative priors. We show that forecasts made using the 1-group model in advance or at the start of a future epidemic are inaccurate, even if informative priors are used, because that model does not account for differing levels of cross-immunity between the starts of the 2009 epidemic and the future epidemic. We then use both information from the 2009 epidemic and data obtained as the future outbreak is ongoing to generate real-time forecasts using the 1-group and 2-group models. Early in the outbreak, only the 2-group model can provide accurate forecasts of the remainder of the epidemic. For that reason, cross-immunity should be included in epidemiological forecasting models whenever a large influenza outbreak is related to an antigenically similar strain that has previously caused a major epidemic.

3.2 Methods

3.2.1 Data

Our analysis involves case notification data from the 2009 H1N1 influenza epidemic in Japan comprising the estimated numbers of weekly cases seeking medical attention in that country. These nationwide data were based on sentinel data from 4800 hospitals, extrapolated to the total number of medical facilities in Japan [278, 284]. The data were acquired from Figure 1 of the analysis by Nishiura [278] using the data extraction tool <https://automeris.io/WebPlotDigitizer/>. The data represent the numbers of patients per week who sought medical attention and met at least one of the following criteria: (i) acute course of illness, (ii) fever higher than 38°C, (iii) cough, sputum or breathlessness (symptoms of upper respiratory infection), (iv) general fatigue, and (v) positive laboratory diagnosis.

It was estimated that 23.5% of the Japanese population were infected during the epidemic, and that 16.1% were infected and sought medical attention [258]. Therefore $(23.5 - 16.1)/23.5 = 31.5\%$ of infected individuals did not seek medical attention. We assume that

those infected individuals who did not seek medical attention suffered mild symptoms of influenza because they were cross-immune to the virus (see Discussion). Hence, extrapolating to the rest of the population and assuming that the susceptibility of hosts is unaffected by cross-immunity, we assume when fitting the 2-group model that 31.5% of the population were cross-immune to the virus initially and that 68.5% were immunologically naive (i.e. had not previously acquired cross-immunity to the 2009 H1N1 pandemic strain as they had not been infected by a previous antigenically similar strain of the virus). However, we also test the robustness of our results to this assumption by conducting analyses in which different proportions of the population are initially cross-immune (Appendix B.2).

3.2.2 Models

We consider two models characterising influenza outbreaks. In the first (the 1-group model), which is the commonly used SEIR model introduced in Chapter 1 [12, 66, 72, 256, 366], cross-immunity is neglected. In the second (the 2-group model), individuals who have been infected previously by a related strain are assumed to recover from infection more quickly than individuals who are immunologically naive. A schematic illustrating the compartmental structures of the two models is shown in Figure 3.1.

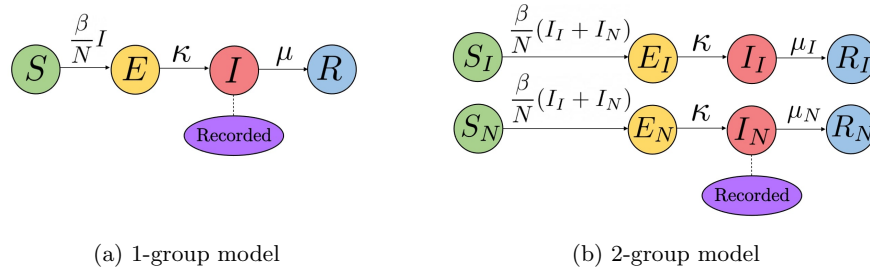


Figure 3.1: Schematic of the 1-group and 2-group models. In both cases, the population is compartmentalised into susceptible S , exposed (infected but not yet infectious) E , infectious I , and recovered R classes. The 2-group model distinguishes between cross-immune and immunologically naive individuals, and involves the assumption that only infected naive individuals are recorded in case notification data (with perfect reporting). The 1-group model does not distinguish between cross-immune and immunologically naive individuals, and involves the assumption that all infected individuals are recorded in case notification data (again with perfect reporting).

3.2.2.1 1-group model

The 1-group SEIR model is described by the following differential equations, in which individuals are either susceptible and available for infection, exposed (i.e. infected but not yet

infectious or symptomatic), infectious or, recovered:

$$\frac{dS}{dt} = -\frac{\beta}{N}SI, \quad (3.1)$$

$$\frac{dE}{dt} = \frac{\beta}{N}SI - \kappa E, \quad (3.2)$$

$$\frac{dI}{dt} = \kappa E - \mu I, \quad (3.3)$$

$$\frac{dR}{dt} = \mu I. \quad (3.4)$$

In this model, the infection rate is governed by the parameter β , the mean latent period is $1/\kappa$ weeks, and the mean infectious period is $1/\mu$ weeks. The basic reproduction number under the 1-group model is given by

$$R_0 = \frac{\beta}{\mu}. \quad (3.5)$$

Following Cintr3n-Arias et al. [76], the number of recorded cases in week j (recorded at the end of that week) where j is the integer number of weeks since the epidemic began, is given by

$$C(j) = \int_{j-1}^j \kappa E dt. \quad (3.6)$$

The constant value $S + E + I + R = N$ represents the effective population size. Since pathogens are most likely to be transmitted locally, individuals in distant locations are not available for infection and so N is expected to be smaller than the true population size [127, 308]. In Table 3.1, we list the parameters that appear in equations (3.1)-(3.4) and estimates of their values for the Japanese 2009 H1N1 epidemic (see also Section 3.2.3 and 3.3.1).

3.2.2.2 2-group model

The 2-group model is an extension of the standard SEIR model in which immunologically naive and cross-immune individuals are distinguished between. The 2-group model is given by the following system of differential equations:

$$\frac{dS_I}{dt} = -\frac{\beta}{N}S_I(I_I + I_N), \quad (3.7)$$

$$\frac{dE_I}{dt} = \frac{\beta}{N}S_I(I_I + I_N) - \kappa E_I, \quad (3.8)$$

$$\frac{dI_I}{dt} = \kappa E_I - \mu_I I_I, \quad (3.9)$$

$$\frac{dR_I}{dt} = \mu_I I_I, \quad (3.10)$$

$$\frac{dS_N}{dt} = -\frac{\beta}{N}S_N(I_I + I_N), \quad (3.11)$$

$$\frac{dE_N}{dt} = \frac{\beta}{N}S_N(I_I + I_N) - \kappa E_N, \quad (3.12)$$

$$\frac{dI_N}{dt} = \kappa E_N - \mu_N I_N, \quad (3.13)$$

$$\frac{dR_N}{dt} = \mu_N I_N. \quad (3.14)$$

The basic reproduction number R_0 of the 2-group model is given by

$$R_0 = \frac{\beta\nu}{\mu_I} + \frac{\beta(1-\nu)}{\mu_N}, \quad (3.15)$$

where the two terms represent the relative contributions from the cross-immune and immunologically naive groups respectively. The first term represents the expected number of infections generated by a single infected individual with previously acquired cross-immunity (assuming that no other individuals are infected at the time of introduction), scaled by the proportion of infected individuals who previously acquired cross-immunity. The second term represents the expected number of infections generated by a single infected individual who were previously immunologically naive, scaled by the proportion of infected individuals who were previously immunologically naive. See Appendix B.3 for details of the calculation.

There is evidence that cross-immunity in the elderly population reduced the disease severity of the 2009 H1N1 pandemic virus [65, 254]. We assume that cross-immune individuals experience less severe disease and therefore typically recover more quickly than immunologically naive individuals (i.e. $1/\mu_I < 1/\mu_N$). Cross-immunity is therefore partial: it does not reduce disease completely. To isolate this effect alone on the predictability of epidemics, in this model it is assumed that cross-immune and naive individuals are otherwise identical. In particular, we assume that susceptible cross-immune individuals and susceptible naive individuals are equally likely to be infected and have the same latent periods.

We assume that only cases of severe disease (i.e. infected individuals who were previously immunologically naive) report infection, so that the number of recorded cases in week j is given by

$$C(j) = \int_{j-1}^j \kappa E_N dt. \quad (3.16)$$

Since we assume that only immunologically naive individuals report infection, the infectious period of immunologically naive individuals in the 2-group model is assumed to be identical to the infectious period of individuals in the 1-group model (i.e. $1/\mu_N = 1/\mu$).

Denoting the proportion of individuals in the population that are cross-immune by ν , we have that $S_I + E_I + I_I + R_I = \nu N$ and $S_N + E_N + I_N + R_N = (1-\nu)N$, where $S_I + E_I + I_I + R_I + S_N + E_N + I_N + R_N = N$ is the total effective population size. In Table 3.1, we list the parameters that appear in equations (3.7)-(3.14) and estimates of their values for the Japanese 2009 H1N1 epidemic (see also Section 3.2.3 and 3.3.1).

3.2.3 Parameter estimation and forecasting

When fitting the models to each dataset in this chapter (either the data from 2009 in Japan or simulated future outbreak data), the transmission rate parameter, β , and effective population size, N , are estimated using Markov chain Monte Carlo (MCMC) with the Metropolis-Hastings algorithm [160]. All other parameters are assumed to be known. A likelihood function is used in which it is assumed that the differences between the data and model forecasts (where the

Parameter	Description	Values	Source
β	Transmission rate	1.644 Weeks^{-1}	Est. from data (Figure 3.3)
N	Effective population size	3.072×10^7	Est. from data (Figure 3.3)
$1/\kappa$	Latent period	4/7 Weeks	[373]
$1/\mu$	Infectious period	1 Week	[373]
$C(0)$	Initial number of recorded cases	24,073	[258]

(a): 1-group model

Parameter	Description	Values	Source
β	Transmission rate	1.947 Weeks^{-1}	Est. from data (Figure 3.3)
N	Total effective population size	4.660×10^7	Est. from data (Figure 3.3)
ν	Cross-immune fraction	0.3149	[258]
$1/\kappa$	Latent period	4/7 Weeks	[373]
$1/\mu_I$	Infectious period (cross-immune individuals)	3/7 Weeks	Est. based on [112]
$1/\mu_N$	Infectious period (naive individuals)	1 Week	[373]
$C(0)$	Initial number of recorded cases	24,073	[258]

(b): 2-group model

Table 3.1: Descriptions of parameters of the 1-group and 2-group models and estimates of their values for the Japanese 2009 H1N1 epidemic.

differences are due to noise not accounted for in the models) are normally distributed, and that this noise scales with the square root of the size of the data (i.e. the number of cases). We estimate this noise scaling parameter, σ , in the likelihood function (which we then fix throughout this chapter), by first fitting each model to data from the 2009 Japan epidemic using a least squares approach [76]. An analysis of the residuals (the scaled differences between each datapoint and the corresponding model values) is given in Appendix B.4, justifying the square root noise scaling assumption. In each MCMC simulation, we perform 2×10^5 sampling iterations, discarding the first 10^4 iterations as the ‘burn-in’ period and record every 100 iterations thereafter to reduce autocorrelation. Further details are given in Appendix B.5.

When making forecasts in real-time after $t = m$ weeks of the epidemic, we calibrate our model forecasts with the observed data for weeks $0, 1, \dots, m$ of the epidemic, estimating model parameters using the method described above. To generate forecasts, we use the estimated distributions of these model parameter values and project the models forwards, starting in week m with initial conditions estimated based on the number of cases observed in weeks $0, 1, \dots, m$ (for details, see Appendix B.6).

To compare different model forecasts throughout this chapter, we use the deviance information criterion (DIC) as stated in Spiegelhalter et al. [350]. Details and DIC values are given in Appendix B.7.

3.2.4 Modelling the size of the cross-immune population between epidemics

For the 2-group model, we assume that all immunologically naive individuals infected during an earlier epidemic acquire cross-immunity to antigenically related strains of the virus.

Individuals who were cross-immune at the beginning of the 2009 epidemic, as well as individuals who were infected during the 2009 epidemic, are therefore assumed to be cross-immune. Furthermore, we assume that cross-immunity to related strains is lifelong [399]. Using these assumptions, we can model the cross-immune fraction of the population over the years between the end of a first epidemic and the start of a future epidemic of a related strain of influenza accounting for population turnover (births and natural deaths). This is important, since the period between major influenza epidemics (or pandemics) is typically many years [202]. Further details are given in Appendix B.8. We seed the future epidemics by assuming there are 10,000 recorded cases in the first week from which we generate forecasts (three orders of magnitude smaller than the effective population sizes used in our models). When using the 2-group model to forecast future epidemics, we assume that the cross-immune fraction of the population is known exactly.

3.2.5 Approaches to forecasting a future epidemic

We use the 1-group and 2-group models (Figure 3.1) to explore the accuracies of three different potential forecasting approaches (see Figure 3.2 for a schematic illustrating these approaches). Specifically, we consider: generating a forecast using case notification data collected in real-time but without incorporating prior information from an earlier epidemic (Strategy 1); generating a forecast in advance of the epidemic, but using the previous epidemic as prior information (Strategy 2), and; generating a forecast using case notification data collected in real-time and also using the previous epidemic as prior information (Strategy 3).

3.3 Results

3.3.1 Fitting the models to the 2009 H1N1 influenza epidemic in Japan

We fit the 1-group and 2-group models to data from the 2009 H1N1 influenza epidemic (Figure 3.3 (a)-(b)). The shaded regions represent the 95% prediction credible intervals (CrIs) of epidemic trajectories based on the posterior distributions of fitting parameters β and N . Since the 1-group model does not account for infected but unrecorded individuals, the estimated parameters for the model represent a lower effective population size N than the for the 2-group model, while the estimated basic reproduction numbers R_0 are similar (Figure 3.3 (c)-(d)). The mean estimated values of R_0 for the 1-group and 2-group models are 1.644 (95% CrI [1.627, 1.661]) and 1.597 (95% CrI [1.580, 1.612]) respectively, comparable with the interquartile range of 1.30 to 1.70 based on 57 studies of the 2009 H1N1 pandemic strain [39]. The values for the estimated parameters and their corresponding CrIs are stated in Table 3.2. The numbers of recorded, unrecorded, and combined total weekly cases estimated using the 2-group model are shown in Appendix B.9.

We used the mean estimated transmission rate and effective population size from the 2-group model to generate synthetic data for future epidemics. We assumed that, if the

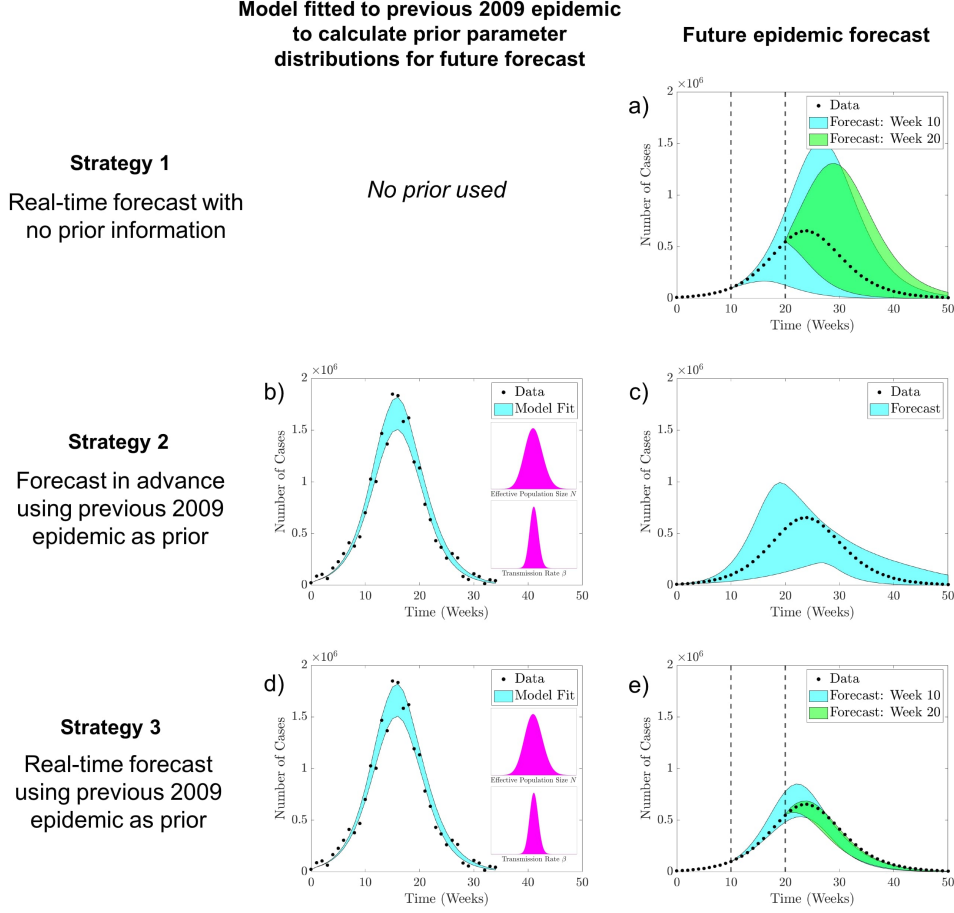


Figure 3.2: Schematic illustrating the modelling approaches considered for forecasting the dynamics of a future epidemic. The shaded regions indicate the 95% CrIs of epidemic trajectories and forecasts. (a) Strategy 1: Forecast in real-time, using data from the ongoing epidemic and assuming no prior information about the fitting parameters. Cyan and green forecasts calibrated using data up to week 10 and 20 of the epidemic, respectively. (b)-(c) Strategy 2: Forecast in advance of a future epidemic using prior parameter distributions estimated from fitting the model to data from a previous epidemic (priors are shown as panels in (b)). (d)-(e) Strategy 3: Forecast in real-time, using data from the ongoing epidemic as well as prior parameter distributions estimated from fitting the model to data from a previous epidemic (priors are shown as panels in (d)).

simulated epidemic takes place further in future, then the immune fraction of the population is lower due to deaths of cross-immune hosts and births of immunologically naive hosts (see Appendix B.8). In Appendix B.1 we show a sensitivity analysis of the latent and infectious period parameters in the 1-group and 2-group models.

3.3.2 Forecasting an epidemic in real-time without prior information

If a major influenza epidemic were to occur, real-time forecasts could be made using live data describing the numbers of cases each week to update predictions. We considered a future

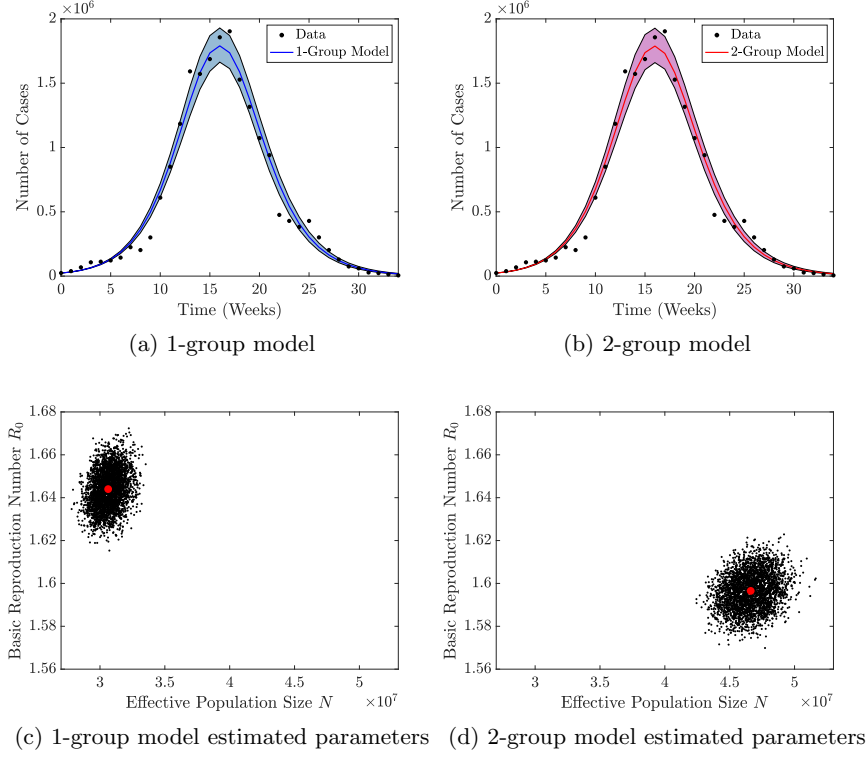


Figure 3.3: (a)-(b): The 1-group and 2-group models fitted to data describing the numbers of new recorded cases each week from the 2009 H1N1 influenza epidemic in Japan, using the transmission rate β and effective population size N as fitting parameters. Solid lines and shaded regions indicate the mean and 95% CrIs of epidemic trajectories based on the posterior distributions of the fitted parameters. (c)-(d): Scatter plots of the posterior distributions of R_0 (which is directly proportional to β) and N . Red dots represent the mean estimates of the parameters. DIC values are given in Table B.1. Estimated parameters along with their CrIs are given in Table 3.2.

Parameter	Mean	95% CrI
β	1.644	[1.627, 1.662]
N	3.072×10^7	$[2.905, 3.246] \times 10^7$

(a): 1-group model

Parameter	Mean	95% CrI
β	1.947	[1.927, 1.966]
N	4.660×10^7	$[4.388, 4.938] \times 10^7$

(b): 2-group model

Table 3.2: Mean values and CrIs of the fitting parameter posterior distributions, for the 1-group and 2-group models fitted to data from the 2009 H1N1 influenza epidemic in Japan.

epidemic due to a strain that is antigenically similar to the 2009 H1N1 virus occurring 25 years after the 2009 outbreak. We generated forecasts using the 1-group and 2-group models, at 10, 20, and 30 weeks after the first recorded cases. The results presented in Figure 3.4 show how the model forecasts change as case notification data become available during the epidemic. The uncertainty in the forecasts of both models is large when predictions are generated early in the epidemic. By contrast, if forecasts are generated later, the remainder of the epidemic is predicted accurately. However, for the more accurate predictions obtained in week 30, the peak of the epidemic has already passed, and so accurate forecasting may

be less practically useful. We conclude that it is challenging to generate practically useful forecasts of the remainder of an epidemic in real-time without prior information about the parameters governing pathogen transmission. This result motivates us to use information from previous epidemics when forecasting the dynamics of a future one; we investigate this further below.

3.3.3 Forecasting epidemics in advance

We considered informing forecasts of a future epidemic using gamma distributed estimates of the fitting parameters (N and β). The mean values of these gamma distributions, which were determined from the fit to the 2009 epidemic data, are given in Appendix B.10.

We used our models to generate forecasts of the dynamics of future major influenza epidemics in Japan arising from a related strain and occurring 25, 50, or 75 years after the 2009 epidemic (with no major epidemics occurring in each intervening period). The forecasts in Figure 3.5 show that, if the 1-group model is used, the dynamics of a future epidemic are predicted to be identical to those of the 2009 epidemic, regardless of when it occurs. By contrast, for the 2-group model, a large proportion of the population would be cross-immune after the 2009 epidemic, resulting in a lower basic reproduction number. Consequently, if an antigenically related strain were to emerge 25 years later, the epidemic would be smaller than in 2009 (Figure 3.5 (a)-(b)). If the next major epidemic instead occurred 75 years later, a large proportion of the population would be immunologically naive to a related strain of the 2009 virus (because of population turnover due to births and deaths), resulting in a greater basic reproduction number. Hence, if all other factors were similar to those in 2009, the future epidemic would be larger than in 2009 (Figure 3.5 (e)-(f)).

Assuming that the 2-group model reflects the underlying epidemiology more accurately than the 1-group model, we conclude that forecasts generated using the 1-group model may not predict the dynamics of a future epidemic in which cross-immunity is present closely. Forecasts using larger and smaller variances of the distributions of β and N are presented in Appendix B.11. The dependence of a range of parameters obtained using the 2-group model (the cross-immune fraction at the start of the future epidemic, the basic reproduction number R_0 , the total number of recorded cases, duration, and size and timing of the epidemic peak) on the time period between the 2009 epidemic and the future epidemic is shown in Appendix B.12. As seen in Figure 3.5, the further into the future the next epidemic of a related strain occurs then, all else being equal, the greater the total and peak number of cases and the shorter the duration of the epidemic. This is due to the smaller proportion of cross-immune individuals leading to a higher value of R_0 . We note that the impact of R_0 on the peak, final size, and duration of an epidemic has been explored previously for simple epidemiological models [362, 384].

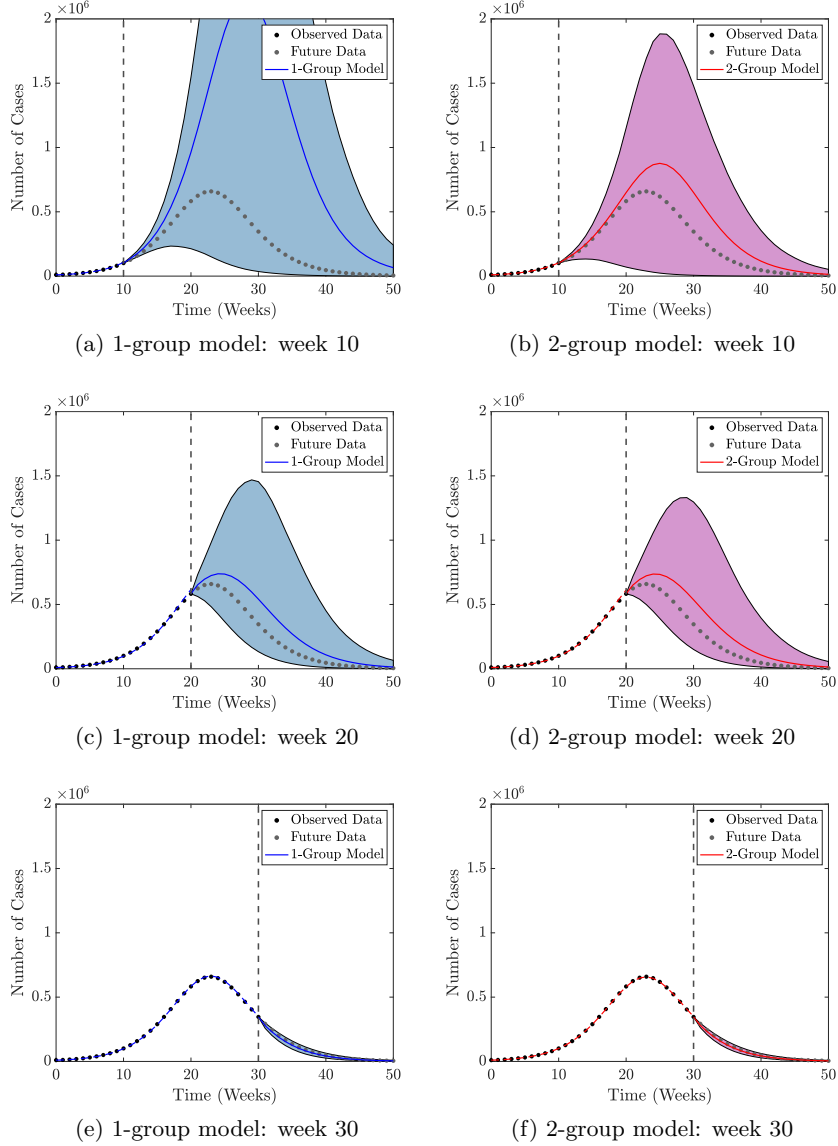


Figure 3.4: Forecast Strategy 1: Real-time forecasts obtained from the 1-group and 2-group models for a future epidemic occurring 25 years after the 2009 epidemic, calibrated by fitting the model parameters β and N to data describing the weekly numbers of new cases, using uninformative uniform priors. Forecasts were made in weeks 10, 20, and 30 of the epidemic, using the observed data up until those times to estimate model parameters. Dashed vertical lines separate the calibration and forecasting periods. Other dashed lines indicate the mean of the epidemic trajectories in the model calibration period. Solid lines and shaded regions indicate the mean and 95% CrIs of the forecasts, based on the posterior distributions of the parameters. The synthetic data were generated using the mean parameters of the 2-group model fitted to data from the 2009 epidemic. Uniform priors $N \in [1 \times 10^7, 12.8 \times 10^7]$ and $\beta \in [1, 3]$ were used in the estimation. DIC values are given in Table B.1.

3.3.4 Forecasting an epidemic in real-time with prior information

As shown above, real-time forecasts of a future influenza epidemic, without informative priors

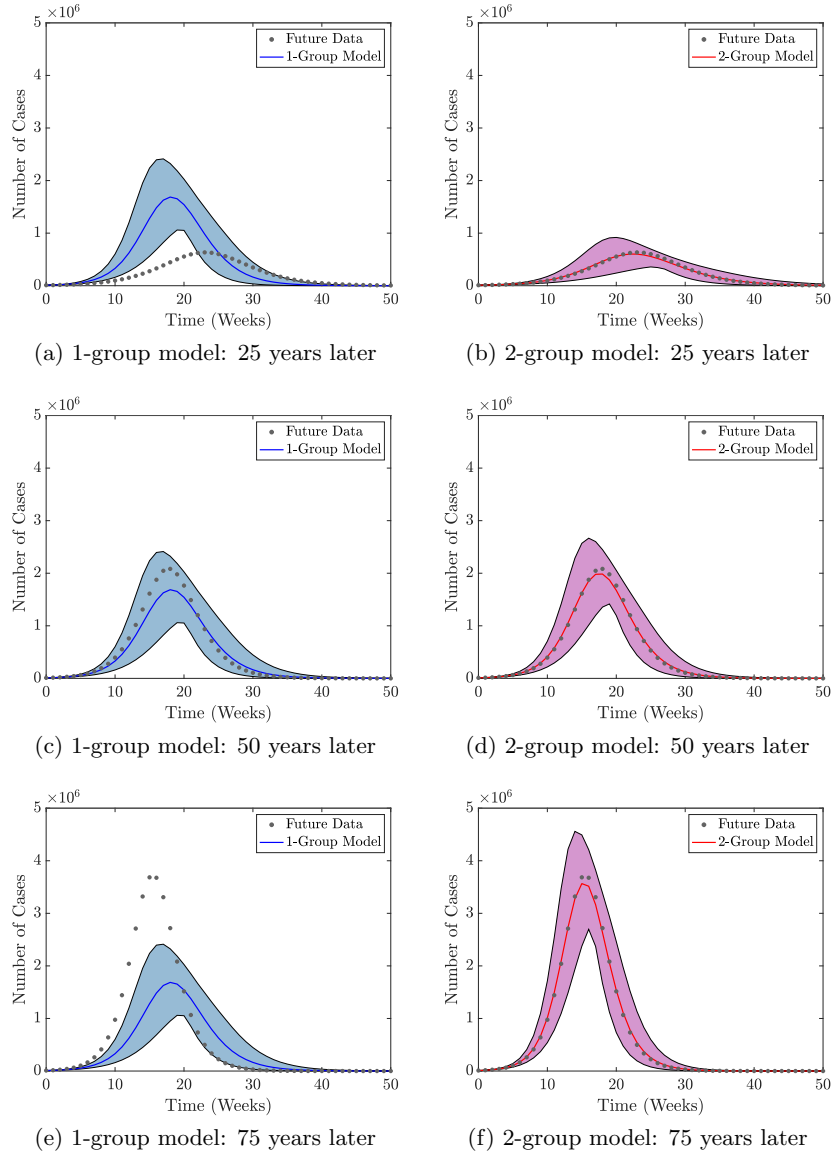


Figure 3.5: Forecasts Strategy 2: Forecasts obtained at the start of future epidemics using the 1-group and 2-group models, for epidemics occurring 25, 50, and 75 years after the 2009 epidemic. Forecasts were generated by assuming that the parameters β and N follow a gamma distribution, the parameters of which were determined by fitting the models to data from the previous 2009 epidemic (Appendix B.10). Solid lines and shaded regions indicate the mean and 95% CrIs of the forecasts. The synthetic data were generated using the mean parameters of the 2-group model fitted to data from the 2009 epidemic. DIC values are given in Table B.1.

for the fitting parameters, were typically very uncertain (e.g. Figure 3.4 (a)-(b)). We therefore also considered using priors to inform the fitted parameter values, so that real-time forecasts are based on both historical data (from the 2009 epidemic) and live data from the ongoing outbreak. Gamma distributed priors were set for β and N , with mean values based on the parameter estimates obtained using the data from the 2009 epidemic.

As before, we considered a scenario in which an epidemic occurs 25 years after the 2009 epidemic. In turn, predictions were made 0, 10, 20, and 30 weeks after the start of the future epidemic (Figure 3.6 (a)–(b), (c)–(e), (f)–(h), and (i)–(k), respectively).

We considered different widths of priors used to inform the epidemic forecasts (Figure 3.6 and Appendix B.13). Under the baseline variance considered, when the 1-group model was used, there was sometimes a discrepancy between the calibrated model trajectory and the observed epidemic data (e.g. left part of Figure 3.6 (f)). For that reason, we also show epidemic forecasts obtained using a wider prior (Figure 3.6 (d), (g), and (j)), so that the model fits the observed data more accurately. When the 1-group model is used, the forecast is either inaccurate (when a prior with low variance is used; e.g. Figure 3.6 (c)) or imprecise (when a prior with higher variance is used; e.g. Figure 3.6 (d)). In either case, as shown quantitatively by the DIC values in Table B.1, our main conclusion is unchanged: predictions are improved when the more epidemiologically realistic 2-group model is used.

3.4 Discussion

Influenza epidemics and pandemics place a significant burden on healthcare systems throughout the world [260]. Exposure to an influenza virus confers cross-immunity to antigenically related strains [211], however cross-immunity is not frequently included explicitly in influenza forecasting models [22, 313]. In this chapter, we have investigated whether or not it is necessary to account for cross-immunity when forecasting influenza epidemics and, moreover, whether data from previous epidemics can be used to improve forecasts.

We have considered two different mathematical models that describe pathogen transmission: a 2-group model in which immunologically naive and cross-immune individuals are differentiated between and a 1-group model in which all individuals are assumed to have similar characteristics. We parameterised each model using data describing the estimated numbers of infected individuals seeking medical attention per week in Japan during the 2009 H1N1 influenza pandemic [278]. We then considered a scenario in which we predict the dynamics of an epidemic of an antigenically related strain occurring 25 years after the 2009 pandemic. When no prior knowledge about the values of the transmission rate and effective population size is assumed, neither model can predict future dynamics of the epidemic in real-time early in the outbreak with any certainty. Previous research has also indicated that significant uncertainty exists when ordinary differential equation models are used to generate real-time forecasts early in influenza epidemics [149].

We then considered the effect of the time period between epidemics on the accuracy of epidemic forecasts made in advance of the future epidemic. Immediately after the 2009 epidemic, a large fraction of the population will have been cross-immune to strains that are related to the 2009 H1N1 virus. This fraction reduces as the time since the 2009 epidemic increases, as individuals (a large proportion of whom may have acquired cross-immunity) die

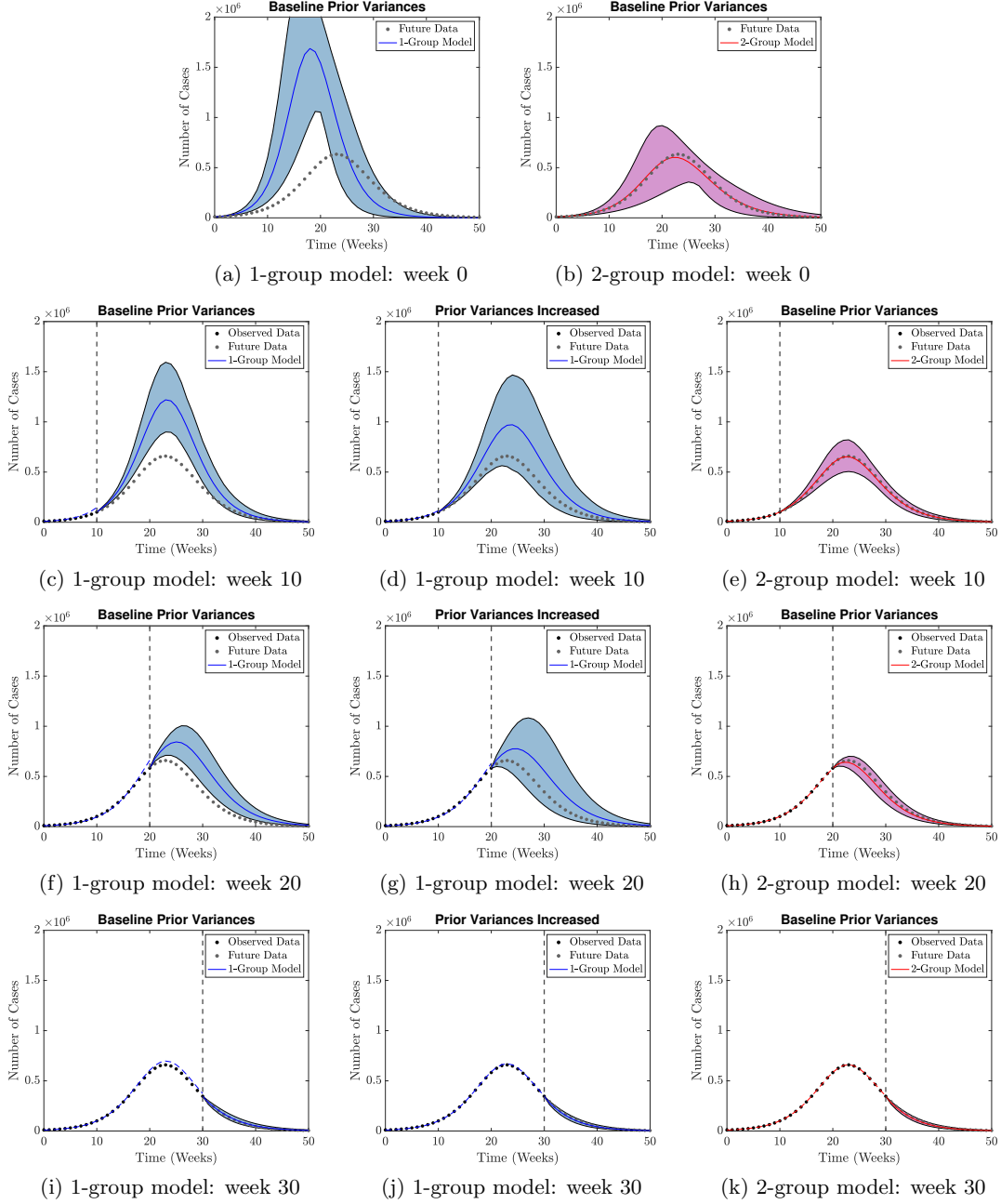


Figure 3.6: Forecast Strategy 3: Forecasts obtained using the 1-group and 2-group models during a future epidemic occurring 25 years after the 2009 epidemic, at the start of the epidemic (a)-(b), and using real-time case notification data obtained during the epidemic (c)-(k). Gamma distributed priors are prescribed on the fitting parameters β and N , the parameters of which were determined by fitting the models to data from the previous 2009 epidemic (Appendix B.10). Forecasts are generated in weeks 10, 20, and 30 of the epidemic, using the observed data up until those times to estimate model parameters. Dashed vertical lines separate the calibration and forecasting periods. Other dashed lines indicate the mean of the epidemic trajectories in the model calibration period. Solid lines and shaded regions indicate the mean and 95% CrIs of the forecasts, based on the posterior distributions of the parameters. The synthetic data were generated using the mean parameters of the 2-group model fitted to data from the 2009 epidemic. DIC values are given in Table B.1.

and immunologically naive individuals are born. Consequently, the expected size of a future epidemic of a related strain increases the further into the future it occurs. If we neglect cross-immunity by using the 1-group model, we predict the same epidemic dynamics as in 2009, regardless of when the future epidemic occurs. This is because changing numbers of cross-immune individuals are not accounted for in that model. This would either result in an overestimation of the size of a future epidemic, if the cross-immune fraction was greater than the baseline cross-immunity in 2009 (i.e. if the future epidemic occurred soon after 2009), or an underestimation of the epidemic size if the cross-immune fraction was lower than in 2009 (i.e. if the future epidemic occurred long after 2009).

Finally, we considered incorporating knowledge of parameters from the 2009 epidemic, combined with using case notification data obtained in real-time during the future epidemic, to update forecasts as the future epidemic is ongoing. When the 1-group model was used to make forecasts, we found that priors with significant uncertainty had to be used to calibrate the model trajectories to the data, especially if estimation was carried out using case notification data from before the epidemic peak alone. Although these priors, set based on the model fits to the 2009 epidemic data, led to improved forecasts compared to when no prior information was used, the forecasts were still inaccurate or uncertain when made before the peak of the epidemic. By contrast, the 2-group was able to generate more accurate forecasts. This is because cross-immunity is accounted for in the 2-group model, and the priors reflect the values of the true underlying parameters more accurately. We conclude that, to forecast the long-term dynamics of major influenza epidemics caused by strains that are antigenically related to those responsible for previous epidemics, a model that accounts for cross-immunity should be used, and data from related previous epidemics taken into account. Additionally, we note that short-term forecasts on the order of one to four weeks (a time horizon used by the US Centers for Disease Control when constructing seasonal influenza forecasts [40, 41, 61]) are also more accurate when made using the 2-group model (as shown in Figure 3.6).

We note that there may be instances in which the underlying epidemiology of an outbreak is not well known, or where a complex model cannot be parameterised due to poor surveillance or limited data availability [129, 367]. In such cases, using a simple model akin to the 1-group model used in this chapter may be the only possible option for making predictions. Although forecasts made using a simple model may lack precision, they could still be useful to policy makers during an epidemic. For example, the 1-group model was able to predict the duration of the epidemic accurately when using data from the first 10 weeks of the outbreak, even if it could not predict the trajectory of the entire epidemic perfectly. Simple models may be particularly useful for short-term forecasts [124]. By constantly updating model predictions using newly acquired data, and refining the modelling framework when it is unable to replicate observed data, simple models that do not describe the underlying epidemiology of an epidemic fully may still provide useful insights.

We investigated whether or not it is necessary to include a particular source of heterogeneity (i.e. cross-immunity) when constructing influenza epidemic forecasts. Our analysis used notification data from Japan collected during the 2009 influenza pandemic. Although the exact dataset underlying our analysis was not central to our main conclusion, it would be possible to replicate the analyses here based on any number of other influenza datasets [94, 99, 270, 309]. Going forwards, we could also consider how accounting for cross-immunity would affect forecasts of seasonal epidemics, in which case the particular strains driving dynamics in different seasons would need to be considered [301]. Consideration of whether and how forecasts of epidemics of pandemic influenza may be influenced by circulating seasonal strains [117] is a target for future work. We could extend our models to investigate whether or not other sources of heterogeneity should be accounted for. For instance, age-structure could be incorporated, taking into account the effects of immune imprinting [139, 140] and variable contact rates. Additionally, we could include spatial heterogeneity, by partitioning the population into distinct geographical regions or introducing a contact network structure [251, 255, 283, 377, 378]. Another potential avenue for further investigation may be to incorporate transmission models which account for cross-immunity into a phylodynamic framework to predict which strains of influenza may emerge and the dynamics of their respective epidemics [143].

In this chapter, we have considered the simple example of binary immunity (i.e. a class of cross-immune individuals and a class of immunologically naive individuals). In reality, the landscape of immunity within a population is much more complex, with individuals exhibiting varying levels of immunity based on their lifetime history of exposure to influenza viruses [139, 140, 217] and their age [155, 216]. Underlying health conditions such as obesity and heart disease may also impact individuals' immune responses [180, 184]. Nonetheless, even the simple framework that we considered allowed us to show that cross-immunity should be accounted for when making influenza forecasts. Future work could involve models containing numerous classes to account for population heterogeneities, or even a continuum level of immunity within the population based on the above-mentioned factors [15, 243].

In the 2-group model considered here, we assumed, as in Reichert et al. [319], that immunoprotection does not prevent infection, only its consequences. Different types of cross-immunity to related strains of a virus can be considered. For example, in Thompson et al. [361] it was assumed that cross-immunity decreases susceptibility, that is, the probability that an individual becomes infected each time they come into contact with an infectious host. This assumption has also been made in influenza transmission models [173]. Similarly, in Chapter 2, in the context of SARS-CoV-2, we made the assumption that vaccine-induced immunity reduces susceptibility. Moreover, infectivity and the duration of the latent period could also be assumed to differ between cross-immune and naive individuals. Future work could consider

whether or not the results of this chapter hold when different assumptions are made about the precise effects of cross-immunity.

In the 2-group model, in order to study the effects of cross-immunity in as simple a setting as possible, we assumed that infected individuals who did not seek medical attention had cross-immune protection to the virus. However, a range of factors (including age and behaviour) affect whether or not, and how quickly, an individual seeks medical attention [365]. This could be built into the underlying modelling framework considered here, although additional data would be needed to parameterise the resulting model. Other additions to the 1- and 2-group models could also be considered. Non-pharmaceutical interventions (NPIs) such as school closures were enforced during the 2009 H1N1 epidemic in Japan. The wide range of interventions that are implemented during outbreaks could be included in the models explicitly to determine their influence on epidemic dynamics and quantities such as transmission rates [20, 125, 229].

Nonetheless, our approach has demonstrated that including cross-immunity in influenza epidemic forecasting models can lead to more accurate forecasts. Cross-immunity due to previous infections has been shown to play a major role in the dynamics of influenza epidemics, with clear evidence emerging from analyses of data from the 1918 Spanish Flu pandemic [358] and the 2009 H1N1 influenza pandemic [151]. Cross-immunity will contribute to shaping the dynamics of future influenza epidemics. Consideration of cross-immunity by epidemiological modellers is therefore of clear public health importance.

Chapter 4

A target-cell limited model can reproduce influenza infection dynamics in hosts with differing immune responses

4.1 Introduction

While in Chapters 2 and 3 we considered epidemics at a population scale, in Chapters 4 and 5 we now consider influenza viral infection within a host. Influenza viruses infect epithelial cells in the respiratory tract of the host in uncomplicated cases [276]. Their kinetics in humans are characterised by exponential growth in viral load over one to three days followed by a decline [60]. Most individuals recover within one week, although the duration of infection may be longer in high-risk groups [375].

As stated in Chapter 1, the immune response to influenza infection can be divided into the innate and adaptive responses [349]. The innate immune response is non-specific to the pathogen, and is comprised of the secretion of cytokines and immune cells such as Natural Killer cells. The adaptive immune response is specific to the pathogen and is comprised of antibodies secreted by B-cells as well as T-cells which clear infected cells.

Mathematical models of influenza viral infection can be categorised into two types. First, target-cell limited models, in which the infection is limited by the availability of susceptible cells [19, 91, 147, 152, 342]. Second, models that explicitly account for components of the immune response at varying levels of detail [44, 55, 58, 150, 172, 220, 252, 292, 330, 371], in which the virus is predominately cleared by the immune system.

Model selection studies have been performed to compare models based on different biological mechanisms. In particular, comparisons of target-cell limited and immune response models, as we shall explore in this chapter, have been made in the context of infection dynamics in mice of different ages [172], drug treatment [152], and infection in immunocompromised and immunocompetent patients [58]. For a comprehensive review of within-host influenza models, see Beauchemin and Handel [31], Handel et al. [154], Smith [343], and Smith [344].

As described in Chapter 1, when constructing models to describe the within-host dynamics of viral infection, a compromise has to be made between model complexity and parameter identifiability (the ability to determine parameter values from the available data). In many studies, point-estimates for model parameters are obtained [19, 153, 175, 263, 330], but it is not always clear whether these parameter values can be uniquely determined or whether other parameter values can provide a similarly good fit. Parameter identifiability can be an issue especially for complex models, as often, only viral load data are available [253, 274]. As we investigate in this chapter, when additional data on the immune response are available (for example, antibody or cytokine levels), values of model parameters may be identified with tighter credible intervals than when only viral load data are available [93].

In this chapter, we address the question of what type of model with what level of biological detail should be used to model within-host influenza dynamics given different available data. Furthermore, we investigate whether a target-cell limited model can reproduce and differentiate between the viral load infection dynamics in different groups of patients within a population, and thus be used for classification purposes.

To address these questions, we consider a hierarchy of models which describe the kinetics of influenza infection. The IR (immune response) model which we adapt from Handel et al. [153], the target-cell limited TEIV model which can be derived from the IR model by removing the immune response terms, and the TV model that comprises only one class of cells. Motivated by the fact that often unknown parameters in within-host models cannot be estimated from partially observed data [252], we perform a structural identifiability analysis to assess if the parameters in theory can be uniquely identified based on the model structure, and then a practical identifiability analysis to determine parameter identifiability given noise-free synthetic data. We then fit each model to experimental murine data [189], first to the viral load kinetics from immunocompetent mice, and then fit the IR model to pooled time series data of the viral load, the fraction of dead cells, and the adaptive and innate immune responses from immunocompetent and immunosuppressed mice (they lack an adaptive immune response). We show that the IR model can be parameterised using the pooled data, then use it to generate synthetic viral load data relating to hosts which have differently functioning immune responses. We then fit the TV and TEIV models to the synthetic data, and show that, in many cases, they can reproduce the viral load data accurately. The parameters of the TV model are shown to depend on the immunological characteristics of the patient groups used to generate the data. Finally, by comparing the TV and IR models, we show that the TV model may be preferred when forecasting the viral load as it has fewer parameters and hence the parameter values can be estimated, which is not the case for the IR model.

4.2 Methods

4.2.1 Models

In this chapter, due to the availability of data relating to viral load, dead cells, and immune response quantities, we consider three models that describe the within-host dynamics of influenza infection. These models are termed the IR, TEIV, and TV models. A schematic of the models is shown in Figure 4.1.

4.2.1.1 IR model

The immune response (IR) model comprises six ordinary differential equations (ODEs) and is adapted from Handel et al. [153]. Uninfected epithelial cells, T , are infected by virions V at rate β and become latently infected cells, E . After a period of $1/\kappa$ days, latently infected cells become productively infected, I . Productively infected cells produce virions, at a maximum rate p (if the innate immune response is inactive). The lifespan of virions is $1/c$ days, while the lifespan of productively infected cells is $1/\delta$ days. The time-scale of the infection (one to two weeks) is much shorter than the average lifespan of uninfected epithelial cells (on the order of one year) [316], so we neglect their natural death and regrowth in our model.

Following infection, the innate immune response is activated by the identification of the virus by immune cells [356]. We consider the role of interferons, F , which inhibit viral replication [289] and are secreted by infected cells. Rather than describing the dynamics of interferons via a differential equation, we follow Handel et al. [153] and Hernandez-Vargas et al. [172], and account for their action via a prescribed function $F(t)$ which grows exponentially for $0 < t < t^*$ and decays exponentially for $t > t^*$ (see equation (4.7)). In equation (4.5), we assume that interferons reduce the rate at which productively infected epithelial cells produce virus particles, and introduce the parameter s to indicate the strength of this inhibition. As in Baccam et al. [19] and Handel et al. [153], a time delay τ is included to account for the time between the production and action of interferons. In practice, interferons initiate a signalling cascade that eventually induces an anti-viral state in cells, and thus a time delay between their production and action may be expected [126, 351].

The adaptive immune response is specific to the virus and takes longer to act than the innate response. Following Handel et al. [153], we assume that the adaptive response, X , is activated at a rate proportional to the viral load V (rate constant f), and undergoes clonal expansion at rate r/m when viral densities are high, and at rate r when viral densities are low. In most cases, $V \gg 1$, and so the ratio r/m governs the exponential growth rate of the adaptive response. It has been shown that the main action of anti-influenza antibodies is to neutralize the virus by removing free virions present in the extracellular space [287]. Accordingly, we assume free virions are neutralized by antibodies at rate k . We further assume that antibodies have a ‘lifespan’ of $1/\gamma$ days. Finally, following [19], we assume that the reduction in viral load due to target cell infection is negligible, and, thus, do not include

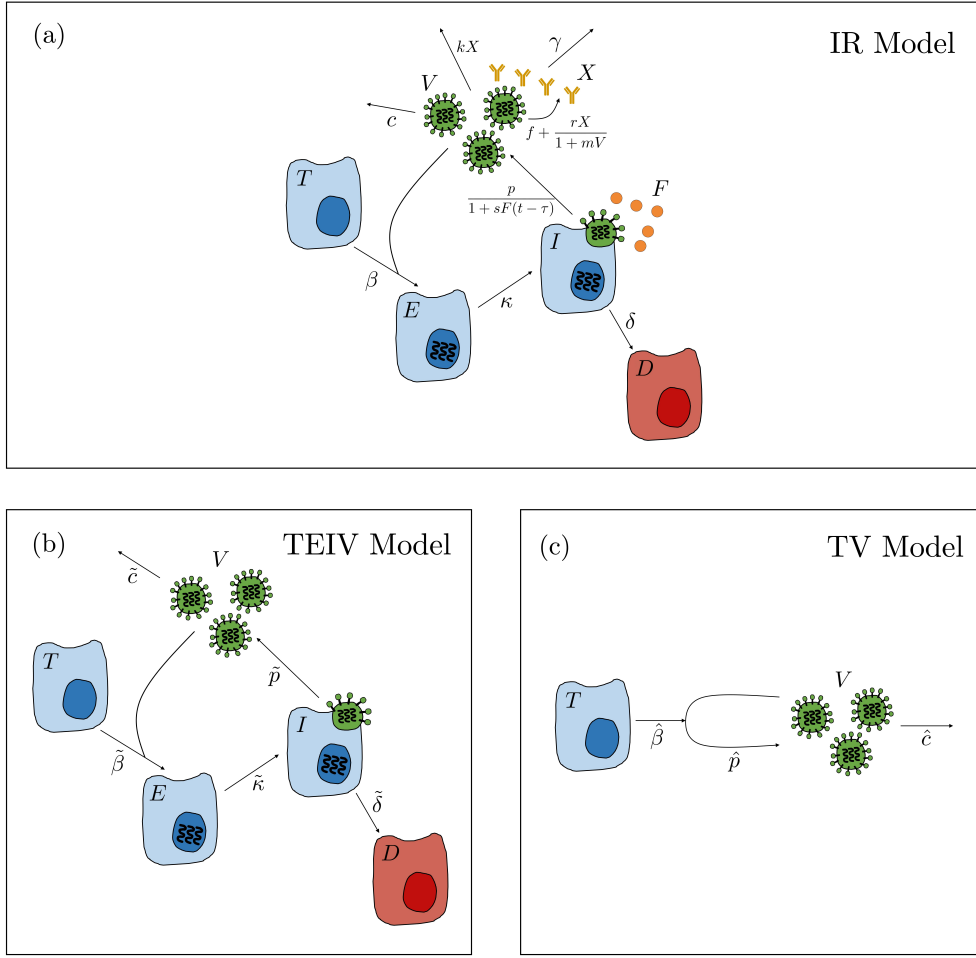


Figure 4.1: Schematic of the models. (a) IR Model: Uninfected epithelial cells, T , are infected by free virions, V , at rate β , and then become latently infected, E , until they become productively infected, I , at rate κ . Productively infected cells die (becoming dead cells, D) at rate δ . They also produce virions at a maximum rate p . Following infection, the innate immune response, F , becomes active after a time delay τ and reduces virion production with a strength parameter s . The strength of the adaptive immune response, X , increases at a rate which depends on the numbers of free virions and productively infected cells, governed by parameters f , r , and m . It decays at rate γ . Free virions are cleared by the adaptive immune response at rate k , and have a decay rate c . (b) TEIV Model: Analogous to the IR model, but without including the immune response explicitly. Hence obtained by setting the adaptive and immune response variables in the IR model to zero ($F = X = 0$). (c) TV Model: Simplest possible model of within-host infection dynamics in which free virions infect epithelial cells at rate $\hat{\beta}$ and $\hat{p}$ new virions are produced per infected cell.

it in the model. We note that in Handel et al. [153] this mechanism has a significant effect on the infection dynamics in some cases but postpone consideration of this effect for future work.

Combining the above assumptions, we arrive at the following system of equations which

describe the time evolution of the infection:

$$\text{uninfected cells} \quad \frac{dT}{dt} = -\beta TV, \quad (4.1)$$

$$\text{latently infected cells} \quad \frac{dE}{dt} = \beta TV - \kappa E, \quad (4.2)$$

$$\text{productively infected cells} \quad \frac{dI}{dt} = \kappa E - \delta I, \quad (4.3)$$

$$\text{dead cells} \quad \frac{dD}{dt} = \delta I, \quad (4.4)$$

$$\text{free virions} \quad \frac{dV}{dt} = \frac{pI}{1 + sF(t - \tau)} - cV - kVX, \quad (4.5)$$

$$\text{adaptive IR} \quad \frac{dX}{dt} = fV + \frac{rXV}{1 + mV} - \gamma X, \quad (4.6)$$

where the innate immune response is represented by the function

$$\text{innate IR} \quad F(t) = \begin{cases} 0, & \text{for } t \leq 0, \\ A_1 e^{\lambda_1 t} & \text{for } 0 < t \leq t^*, \\ A_2 e^{-\lambda_2 t} & \text{for } t > t^*. \end{cases} \quad (4.7)$$

As we neglect growth and natural death of uninfected cells, we can show that, by adding equations (4.1)–(4.4), the total number of cells $T + E + I + D$ is constant. Further, equation (4.4) decouples from the system. Throughout this study, and as in Handel et al. [153] in which the same murine data were analysed, we prescribe the functional form of $F(t)$ where the parameters A_1 , A_2 , λ_1 , λ_2 , and t^* , which govern its dynamics, are fixed. We justify this assumption by noting that the interferon dynamics in immunocompromised and immunocompetent mice are almost identical in the data we consider (see Figure 4.4 and Iwasaki and Nozima [189]).

In what follows, it will be convenient to introduce $C = D/T_0$, the fraction of dead cells, a term we will use when fitting the model to experimental data. We close equations (4.1)–(4.7) by imposing the following initial conditions: $T(0) = T_0$, $E(0) = I(0) = D(0) = 0$, $V(0) = V_0$, $X(0) = 0$, where T_0 is the initial number of uninfected cells and V_0 is the initial viral load. Throughout this chapter, the unit of V is viral titer measured in 50 percent plaque forming units (PFU) and C is dimensionless. We shall denote the unit of X , the titer of neutralizing antibodies, by u_X , and the unit of F , the titer of interferon, by u_F (see Iwasaki and Nozima [189] for further details on these units).

4.2.1.2 TEIV model

The TEIV model assumes that the infection is limited by the availability of uninfected target epithelial cells [19]. We derive the TEIV model by setting $F = X = 0$ in equations (4.1)–(4.7),

giving

$$\text{uninfected cells} \quad \frac{dT}{dt} = -\tilde{\beta}TV, \quad (4.8)$$

$$\text{latently infected cells} \quad \frac{dE}{dt} = \tilde{\beta}TV - \tilde{\kappa}E, \quad (4.9)$$

$$\text{productively infected cells} \quad \frac{dI}{dt} = \tilde{\kappa}E - \tilde{\delta}I, \quad (4.10)$$

$$\text{dead cells} \quad \frac{dD}{dt} = \tilde{\delta}I, \quad (4.11)$$

$$\text{free virions} \quad \frac{dV}{dt} = \tilde{p}I - \tilde{c}V, \quad (4.12)$$

where $\tilde{\beta}$, $1/\tilde{\kappa}$, $\tilde{\delta}$, $\tilde{p}$, and $\tilde{c}$ are the infection rate, duration of eclipse phase, infected cell clearance rate, virion production rate, and virion clearance rate, respectively. As before, equation (4.11) decouples. We close equations (4.8)–(4.12) by imposing the following initial conditions: $T(0) = \tilde{T}_0$, $E(0) = I(0) = D(0) = 0$, and $V(0) = \tilde{V}_0$.

4.2.1.3 TV model

The TV model does not distinguish between different types of infected cells [147]. Free virions are assumed to infect epithelial cells at rate $\hat{\beta}$, and $\hat{p}$ determines how many virions are produced by each infected cell. The model equations are given by

$$\text{uninfected cells} \quad \frac{dT}{dt} = -\hat{\beta}TV, \quad (4.13)$$

$$\text{free virions} \quad \frac{dV}{dt} = \hat{p}\hat{\beta}TV - \hat{c}V. \quad (4.14)$$

We note that although the three models contain similar functional forms, their interpretations differ. In particular, in the IR model, we fix δ and c , the inverse of the lifespan of infected cells and free virions, respectively. However, in the TEIV model, we view $\tilde{\delta}$ and $\tilde{c}$ as fitting parameters and $\hat{c}$ in the TV model as a fitting parameter. Descriptions and estimated values of parameters for all three models are given in Table 4.1.

4.2.2 Experimental data

Following Handel et al. [153], we use data from an experimental study by Iwasaki and Nozima [189] in which naive C3H/He mice were infected with the H1N1 influenza strain A/PR8. The data are from the tracheobronchial washings of immunocompetent mice, and immunosuppressed mice with suppressed antibody production. The data were collected over 12 days and consist of viral load titer, measured in 50 percent plaque forming units (PFU), the fraction of dead cells as measured by lung lesions, IgA, IgM, and IgG antibody titers, and combined type I and II interferon (IFN) levels. The antibody data represent the combined levels of IgA and IgM (IgG is neglected since it was not found to influence the infection dynamics [153, 189]). Further details of the experimental procedure are given in the original study [189]. Data were extracted from Iwasaki and Nozima [189] using the data extraction tool <https://automeris.io/WebPlotDigitizer/>.

Parameter	Description	Values (Units)	Source
β	Infection rate	Fitting parameter (PFU ⁻¹ days ⁻¹)	
p	Virion production rate	Fitting parameter (PFU days ⁻¹ cells ⁻¹)	
$1/\delta$	Lifespan of infected cells	2 days	[153]
$1/c$	Lifespan of free virions	10 days	[153]
$1/\kappa$	Duration of latent eclipse phase	4 days	[153]
τ	Innate IR action time delay	Fitting parameter (days)	
s	Strength of innate IR	Fitting parameter (u_F^{-1})	
r	Expansion rate of adaptive IR	Fitting parameter (PFU ⁻¹ days ⁻¹)	
f	Recruitment rate of adaptive IR	Fitting parameter (u_X PFU ⁻¹ days ⁻¹)	
m	Adaptive IR growth parameter	1	
γ	Decay rate of adaptive IR	1.159 days ⁻¹	[90]
k	Kill rate of adaptive IR	Fitting parameter (u_X^{-1} days ⁻¹)	
A_1	Innate IR dynamics parameter	0.8236 u_F	[153]
A_2	Innate IR dynamics parameter	2.128×10^6 u_F	[153]
λ_1	Innate IR dynamics parameter	1.241 days ⁻¹	[153]
λ_2	Innate IR dynamics parameter	1.712 days ⁻¹	[153]
t^*	Innate IR dynamics parameter	5 days	[153]
T_0	Initial no. of uninfected cells	7×10^9 cells	[153]
V_0	Initial viral load	Fitting parameter (PFU)	

(a): IR model - equations (4.1)–(4.7).

Parameter	Description	Values (Units)	Source
$\bar{\beta}$	Infection rate	Fitting parameter (PFU ⁻¹ days ⁻¹)	
$\bar{p}$	Virion production rate	Fitting parameter (PFU days ⁻¹ cells ⁻¹)	
$\bar{\delta}$	Clearance rate of infected cells	Fitting parameter (days ⁻¹)	[153]
$\bar{c}$	Clearance rate of virions	Fitting parameter (days ⁻¹)	[153]
$1/\bar{\kappa}$	Duration of latent eclipse phase	4 days	[153]
$\bar{T}_0$	Initial no. of uninfected cells	7×10^9 cells	[153]
$\bar{V}_0$	Initial viral load	Fitting parameter (PFU)	

(b): TEIV model - equations (4.8)–(4.12).

Parameter	Description	Values (Units)	Source
$\hat{\beta}$	Infection rate	Fitting parameter (PFU ⁻¹ days ⁻¹)	
$\hat{p}$	Virion production coefficient	Fitting parameter (PFU cells ⁻¹)	
$\hat{c}$	Clearance rate of virions	Fitting parameter (days ⁻¹)	[153]
$\hat{T}_0$	Initial no. of uninfected cells	7×10^9 cells	[153]
$\hat{V}_0$	Initial viral load	Fitting parameter (PFU)	

(c): TV model - equations (4.13)–(4.14).

Table 4.1: Descriptions and parameter values of the IR, TEIV, and TV models.

4.2.3 Structural identifiability analysis

We perform a structural identifiability analysis to determine whether in theory based on the model structure it is possible to uniquely recover model parameters, given data for one or more model outputs is observed [34]. In general terms, we wish to establish the structural identifiability of a model with unknown parameters θ , observable quantities $y(t; \theta) = [y_1(t; \theta), \dots, y_M(t; \theta)]$ and known initial conditions $y(0) = [y_1(0), \dots, y_M(0)]$. We employ the Taylor series approach [64, 71, 306], which assumes that the observable functions $y_j(t; \theta)$ for $j = 1, \dots, M$ and their time derivatives are unique analytic functions of time. Hence, it is possible to represent the observable functions via Taylor series expansions in a neighbourhood of

an initial state. Uniqueness of this representation in terms of the model parameters and initial conditions guarantees structural identifiability. The Taylor series expansion of the observable functions at time $t = 0$, is given by

$$y_j(t, \theta) = y_j(0; \theta) + y_j^{(1)}(0; \theta)t + y_j^{(2)}(0; \theta)\frac{t^2}{2!} + \dots + y_j^{(i)}(0; \theta)\frac{t^i}{i!} + \dots, \quad (4.15)$$

for $j = 1, \dots, M$, where

$$y_j^{(i)}(0, \theta) = \frac{d^i}{dt^i} y_j(0; \theta). \quad (4.16)$$

Let us define

$$a_j^i(\theta) = \lim_{t \rightarrow 0} y_j^{(i)}(t; \theta). \quad (4.17)$$

Then a sufficient condition for global structural identifiability is

$$a_j^i(\theta) = a_j^i(\theta^*) \text{ for } i = 1, \dots, i_{\max} \text{ and } j = 1, \dots, M \implies \theta = \theta^*, \quad (4.18)$$

where $i_{\max}$ is the smallest integer for which solutions for the parameters θ can be obtained in terms of the Taylor series coefficients $a_j^i(\theta)$ [71]. Hence, the problem reduces to establishing whether a particular system of algebraic equations for the parameters θ has a unique solution. A parameter is termed globally identifiable if a unique value of the parameter can be identified, and locally identifiable if the value of the parameter can be uniquely identified in a local region of parameter space.

4.2.4 Practical identifiability analysis

Structural identifiability depends on model structure and is independent of experimental data, and is a prerequisite for practical identifiability [314]. A parameter is deemed practically identifiable if it can be uniquely determined in a local region by fitting the model to data (either real-world or synthetic). We use a profile likelihood based approach to determine practical identifiability in which differences between the model output and data are calculated as parameters vary. This approach is widely used in biological models with large numbers of fitting parameters and does not impose any restrictions on the form of the model equations [314, 315].

Using the likelihood function (C.1) given in Appendix C.1, we define $\chi^2(\theta)$ by

$$\chi^2(\theta) = J(\theta; y) = \sum_{i=1}^M \sum_{j=1}^{N_i} \left(\frac{[-z_i(t_j; \theta) - y_{ij}]^2}{2\sigma_i^2} \right), \quad (4.19)$$

where $z_i(t_j, \theta)$ represent model outputs generated with parameters θ , y_{ij} represent data, and σ_i are noise scaling parameters. The subscripts $i = 1, \dots, M$ relate to the observables (e.g. viral titer, fraction of dead cells, etc.) and the subscripts $j = 1, \dots, N_i$ relate to the output time-points. The maximum likelihood estimated value of θ corresponds to

$$\hat{\theta} = \arg \min_{\theta \in \Theta} J(\theta; y). \quad (4.20)$$

Likelihood-based confidence intervals can be defined from a confidence region

$$\{\theta | \chi^2(\theta) - \chi^2(\hat{\theta}) < \Delta_\alpha\} \quad \text{with} \quad \Delta_\alpha = \chi^2(\alpha, df), \quad (4.21)$$

the boundaries of which represent confidence intervals [247], where Δ_α is the quantile of the χ^2 distribution with df degrees of freedom and a confidence level α . As in Raue et al. [314], we use $\Delta_{\alpha=0.68}$. We also calculate the profile likelihood (PL) χ_{PL}^2

$$\chi_{\text{PL}}^2(\theta_i) = \min_{\theta_{j \neq i}} [\chi^2(\theta_i)], \quad (4.22)$$

meaning the re-optimization of $\chi^2(\theta_i)$ with respect to all other fitting parameters $\theta_{j \neq i}$ (i.e. for each value of θ_i , the value of $\chi^2(\theta_i)$ is minimised using $\theta_{j \neq i}$ as fitting parameters). Following Raue et al. [314], a parameter estimate $\hat{\theta}_i$ is defined to be practically identifiable if the likelihood-based confidence region is not infinitely extended in increasing and/or decreasing directions of θ_i (i.e. $\chi_{\text{PL}}^2(\theta_i)$ crosses the threshold Δ_α for a desired confidence level, α , to the left and right of $\hat{\theta}_i$).

4.2.5 Parameter estimation

As described in Chapter 1, we use two methods to estimate model parameters: Markov chain Monte Carlo (MCMC), with the Metropolis-Hastings algorithm [160] (Figures 4.3, 4.4, and 4.9) and maximum likelihood estimation (MLE) (Figures 4.3, 4.6, 4.7, and 4.8). For each MCMC simulation, we perform 5×10^5 sampling iterations, discarding the first 10^4 iterations as the ‘burn-in’ period and recording every 100 iterations thereafter to reduce autocorrelation. See Appendix C.1 for further details on parameter estimation.

We use the deviance information criterion (DIC) as stated in Gelman et al. [128] and the corrected Akaike Information Criterion (AIC) [53] to compare different model forecasts (details given in Appendix C.2).

4.2.6 Global sensitivity analysis

Sensitivity analyses quantify the extent to which variability in model parameters contributes to variability in model outputs [402]. We focus on global sensitivity analysis, where parameters co-vary over a pre-defined region of parameter space. We use Sobol sensitivity analysis [348], and extended Fourier amplitude sensitivity analysis (eFAST) [333, 334] as they are commonly applied to biological models [242]. Both methods are based on variance decomposition techniques, the primary difference being the algorithm used to perform the multidimensional integration of the sensitivity indices. In eFAST, variance partitioning works by varying each parameter at a particular frequency using a sinusoidal function and Fourier analysis is used to measure the strength of each parameter’s frequency in the model output; Sobol sensitivity analysis uses a Monte Carlo integration method. Both methods output first-order and

total-order sensitivity indices between 0 and 1. The first-order sensitivity index of a parameter evaluates its fractional contribution to the output variance, while the total-order index accounts for interactions with other model parameters.

4.3 Results

4.3.1 Structural identifiability analysis

We performed structural identifiability analyses for the IR, TEIV, and TV models, the results are summarised in Table 4.2. Through nondimensionalisation (see Appendix C.4), we can show that for each model, only the grouping pT_0 can be identified, as we do not observe data of the number of uninfected or infected cells [19, 147, 153, 330]. Furthermore, if X is not observed in the IR model, then the grouping fk is identifiable but not each parameter independently.

Continuing to consider the IR model, if V and X or C and X are observed (Table 4.2, Cases 3 and 4), then all unknown parameters of the IR model, namely $(\beta, p, \tau, s, r, f, k, V_0)$, are at least locally identifiable if there is no time delay ($\tau = 0$ in equation (4.5)). If C and V are observed until at least time t where $t > t^* + \tau$, then all unknown parameters including τ are globally identifiable (Case 1a). Results are also presented which prove identifiability of certain parameter groups if only one or two sets of data are observed (Cases 2-7). For Cases 2-7 we assume certain parameter values in addition to Θ are known (e.g. we assume we know the value of k for Case 2) such that we could determine identifiability; with only Θ known, it was computationally infeasible to prove structural identifiability of all other parameters due to the high order of derivatives that needed to be computed symbolically.

For the TEIV and TV models, it is possible to show that all unknown parameters are identifiable (Cases 8 and 9), and hence can be estimated from viral load data. Hence from this analysis, we are at least in theory able to determine all fitting parameter values from the available data in this study [189]. See Appendix C.4 for details of structural identifiability calculations for the IR model given data for V , C , and X .

Model	Case	Observable Data	Globally Identifiable	Locally Identifiable	Known
IR	1a	V^*, C^*, X	$\beta, p, \tau, s, r, f, k, V_0$		Θ
	1b	V, C, X	$\beta, p, s, r, f, k, V_0$		$\Theta, \tau = 0$
	2	V, C	β, f, V_0	p, s, r	$\Theta, \tau = 0, k$
	3	V, X	r, f, k, V_0	β, p, s	$\Theta, \tau = 0$
	4	C, X		$\beta, p, s, r, f, k, V_0$	$\Theta, \tau = 0$
	5	V	V_0, f	β, p, s	$\Theta, \tau = 0, k, r$
	6	C	β, p, f, s		$\Theta, \tau = 0, k, r, V_0$
TEIV	7	X		$\beta, p, s, r, f, k, V_0$	$\Theta, \tau = 0$
	8	V	$\tilde{V}_0, \tilde{c}$	$\tilde{\beta}, \tilde{p}, \tilde{\delta}$	$\tilde{T}_0, \tilde{\kappa}$
TV	9	V	$\hat{V}_0, \hat{\beta}, \hat{p}, \hat{c}$		$\hat{T}_0$

Table 4.2: Results of the structural identifiability analysis of the IR, TEIV, and TV models given different observable data. For the IR model (Case 1a) V^* and C^* refer to data which are observed until at least time $t > t^* + \tau$. The set of parameters that is assumed to be known for the IR model is $\Theta = [A_1, A_2, \lambda_1, \lambda_2, t^*, \kappa, \delta, c, m, \gamma, T_0]$.

4.3.2 Practical identifiability analysis

Structural identifiability does not guarantee practical identifiability. Indeed, even if a parameter is structurally identifiable, it may not be possible to determine its value by fitting the model to data. Profile likelihood curves for Cases 1, 8, and 9 are presented in Figure 4.2 (see Appendix C.5 for Cases 2-7). Parameters are deemed to be practically identifiable if their profile likelihoods intersect the thresholds (black horizontal dotted lines) from above and below, creating a ‘U’ shape. When V , C , and X are observed then all unknown parameters in the IR model are identifiable (Case 1, Figure 4.2 (a)) However, if only one or two variables are observed, then most of the parameters are not practically identifiable. We note further that if X is not observed, the profile likelihood curves for f and k are identically zero for all values of f and k , which is consistent with our earlier observation that f and k are not structurally identifiable.

All four parameters of the TV model are practically identifiable when V is observed (Case 9, Figure 4.2 (c)). For the TEIV model, the parameters $\tilde{\beta}$, $\tilde{c}$, and $\tilde{V}_0$ are identifiable, while $\tilde{p}$ and $\tilde{\delta}$ are not (Case 8, Figure 4.2 (b)). In what follows though, as in Baccam et al. [19], we will continue to fit all five parameters of the TEIV model such that we can compare the TV model (four unknown parameters) to another target-cell limited model with a greater number of fitting parameters.

4.3.3 Fitting the models to experimental viral load data

We fit the IR, TEIV, and TV models to data of the viral load kinetics in immunocompetent mice (Figure 4.3). The shaded regions represent the 95% credible intervals (CrIs) of model trajectories based on the posterior distributions of fitting parameters calculated using MCMC. We also show the maximum likelihood model trajectory (dashed red line) in each case. Each model captures the exponential increase and decrease in viral load, with only small differences.

Model	Case	Observable Data	Identifiable	Non-Identifiable	Known
IR	1	V, C, X	$\beta, p, \tau, s, r, f, k, V_0$		Θ
	2	V, C	β, p, τ	s, r, f, k, V_0	Θ
	3	V, X	τ, s, r, f, V_0	β, p, k	Θ
	4	C, X	r	$\beta, p, \tau, s, f, k, V_0$	Θ
	5	V		$\beta, p, \tau, s, r, f, k, V_0$	Θ
	6	C		$\beta, p, \tau, s, r, f, k, V_0$	Θ
	7	X		$\beta, p, \tau, s, r, f, k, V_0$	Θ
TEIV	8	V	$\tilde{\beta}, \tilde{c}, \tilde{V}_0$	$\tilde{p}, \tilde{\delta}$	$\tilde{T}_0, \tilde{\kappa}$
TV	9	V	$\hat{\beta}, \hat{p}, \hat{c}, \hat{V}_0$		$\hat{T}_0$

Table 4.3: Results of the practical identifiability analysis of the IR, TEIV, and TV models given different observable data. Synthetic data were generated by simulating the relevant models using the parameter values given in Table 4.1, MLE values in Table 4.6 for the IR model, and MLE values in Table 4.4 for the TEIV and TV models, over a 12 day period. The set of parameters that is assumed to be known for the IR model is $\Theta = [A_1, A_2, \lambda_1, \lambda_2, t^*, \kappa, \delta, c, m, \gamma, T_0]$. Profile likelihood curves relating to Cases 1, 8, and 9 are shown in Figure 4.2, while curves for Cases 2-7 are given in Appendix C.5.

However, for the IR model, the posterior distributions of fitting parameters span many orders of magnitude (Figure 4.3 (b)), confirming that they are not identifiable given these data [338]. Although the posterior distributions of the TEIV model parameters are unimodal, they also span several orders of magnitude. By contrast, the TV model parameters have tighter credible intervals and their distributions have compact support. Estimated parameter values for each model are stated in Table 4.4. As the DIC values of the TEIV and TV models are lower than that of the IR model (Table 4.5), the TEIV and TV models should be preferred over the IR model when describing this particular viral load data. We also note that although the MLE fit of the IR model is more accurate than for the TEIV and TV models, as shown by the lower sum of squared residuals (SSR), the IR model's AIC value is larger due to the greater number of fitting parameters (Table 4.5), again confirming that the TEIV and TV models should be preferred.

4.3.4 Fitting the IR model to pooled experimental data

As in Handel et al. [153], we fit the IR model to pooled data of the viral load, fraction of dead cells, and adaptive immune response in immunosuppressed and immunocompetent mice (Figure 4.4 and Table 4.6). The same parameters are used to describe the immunocompetent and immunosuppressed mice, except that no adaptive immune response is elicited in the immunosuppressed mice (i.e. $f = r = 0$). The model captures the infection dynamics of both sets of mice. Initially the infection dynamics in both sets are similar. However, after approximately 4 days, the immunocompetent mice elicit an adaptive immune response, which clears the virus and limits the number of cells which die. By contrast in the immunosuppressed mice, the viral load resurges after reaching an initial peak because the innate immune response decays after approximately 5 days and they do not elicit an adaptive immune response. Consequently, viral production increases, causing all remaining cells in the immunosuppressed mice to become

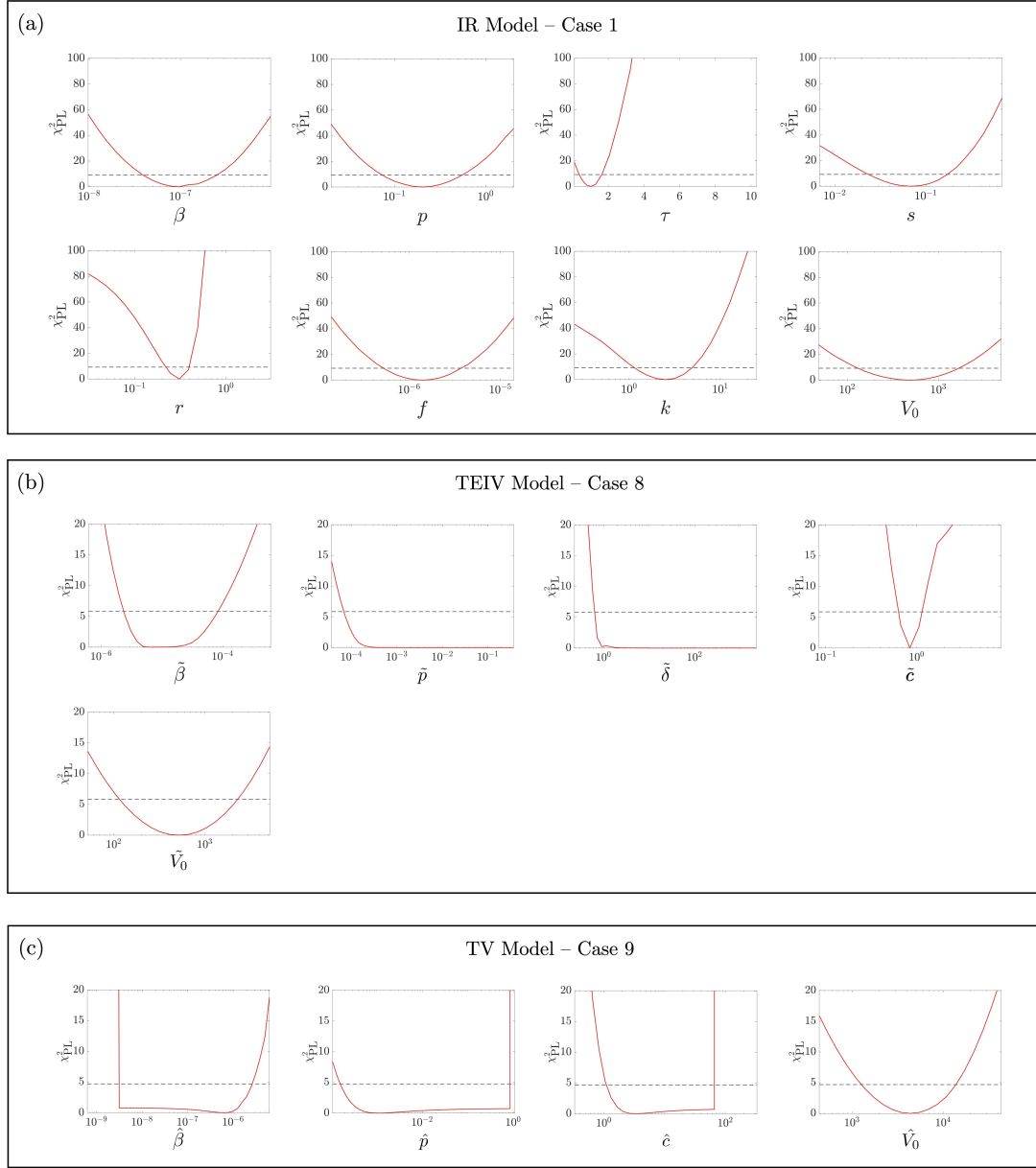


Figure 4.2: Profile likelihood (χ^2_{PL}) curves generated by fitting the IR, TEIV, and TV models to synthetic data for specific cases in Table 4.3: (a) IR model, Case 1, (b) TEIV model, Case 8, and (c) TV model, Case 9. Thresholds for confidence intervals relating to $\Delta_{\alpha=0.68}$ are given by horizontal dashed lines. Parameters are deemed practically identifiable if their profile likelihoods intersect the confidence interval thresholds from above and below. All parameters are practically identifiable for the IR model (Case 1) and the TV model (Case 9), whereas parameters $\tilde{p}$ and $\tilde{\delta}$ are not identifiable for the TEIV model (Case 8).

infected and then to die. As shown in Figure 4.4 (b), the distributions of all fitting parameters are unimodal, with credible intervals spanning approximately one order of magnitude, except for V_0 which has a larger credible interval. These results are consistent with the structural and practical identifiability results that all fitting parameters should be identifiable when these

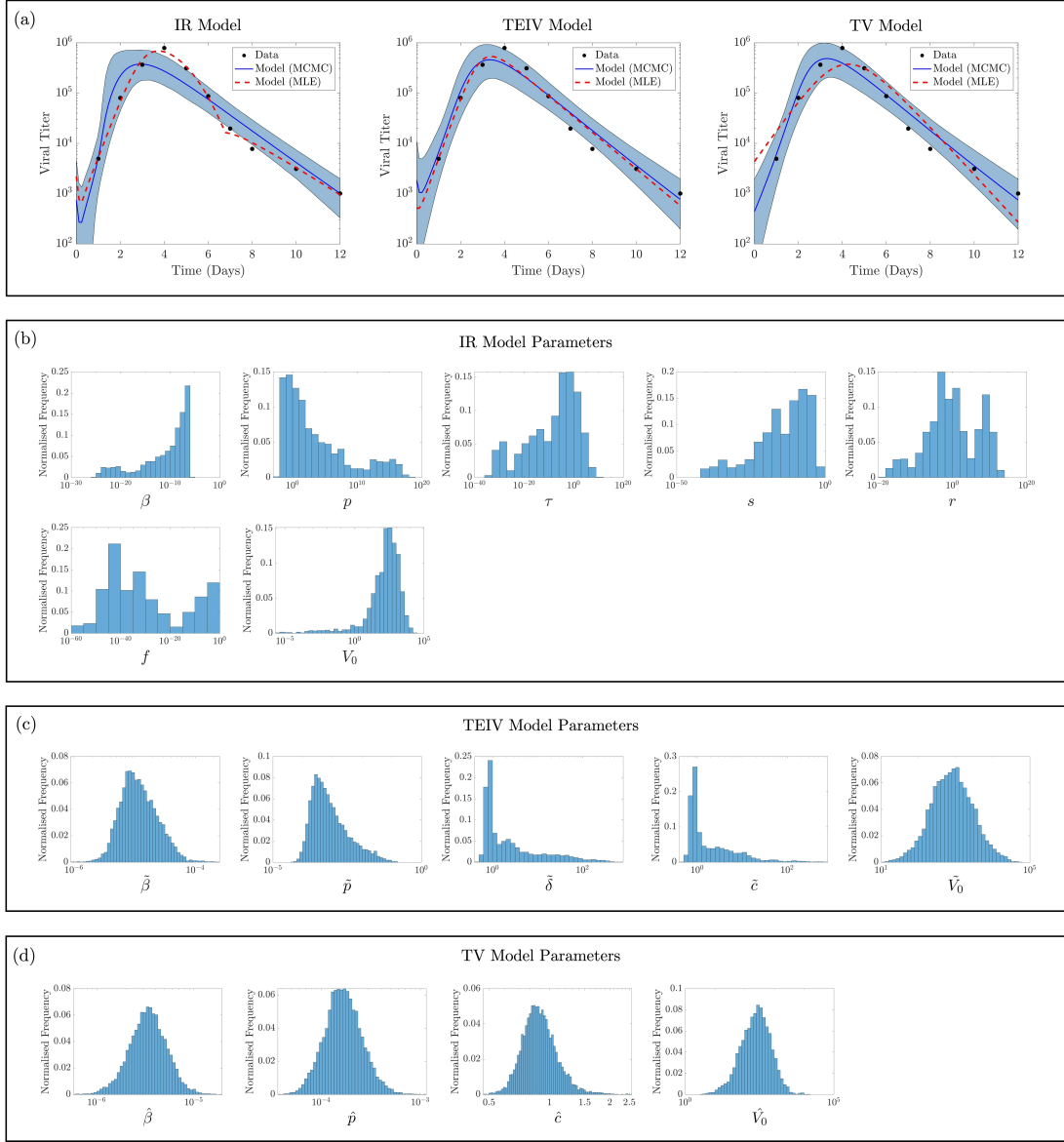


Figure 4.3: (a): The IR, TEIV, and TV models fitted to experimental data on the viral load of infected immunocompetent mice. Solid lines and shaded regions indicate the mean and 95% CrIs of model trajectories based on the posterior distributions of parameters. Red dashed lines indicate model trajectories relating to the maximum likelihood estimated parameter set. (b)-(d): Posterior distributions of fitted parameters for each model computed using MCMC parameter estimation. Normalised Frequency refers to the probability of selecting an observation within that bin interval. Log scale used on the x-axes to show the parameter distributions clearly. Fixed and estimated parameter values are stated in Tables 4.1 and 4.4 respectively.

data are observed. We note however that here we observe data for immunosuppressed and immunocompetent mice; we did not consider this case in the identifiability analyses. Estimated parameter values are given in Table 4.6.

Parameter	MLE
β	2.519×10^{-7}
p	5.464×10^{-2}
τ	1.757
s	8.251×10^{-3}
r	2.241×10^{-2}
f	1.271×10^{-5}
k	1.881
V_0	2.198×10^3

(a): IR model

Parameter	MCMC Mean	MCMC 95% CrI	MLE
$\tilde{\beta}$	1.546×10^{-5}	$[3.598 \times 10^{-6}, 4.979 \times 10^{-5}]$	8.059×10^{-6}
$\tilde{p}$	3.245×10^{-3}	$[1.168 \times 10^{-4}, 2.772 \times 10^{-2}]$	9.430×10^{-4}
$\tilde{\delta}$	13.254	[0.6486, 128.05]	5.939
$\tilde{c}$	7.816	[0.6427, 54.15]	0.8390
$\tilde{V}_0$	1.869×10^3	$[6.028 \times 10^1, 1.109 \times 10^4]$	6.181×10^2

(b): TEIV model

Parameter	MCMC Mean	MCMC 95% CrI	MLE
$\hat{\beta}$	3.710×10^{-6}	$[1.354 \times 10^{-6}, 7.945 \times 10^{-6}]$	1.085×10^{-6}
$\hat{p}$	1.836×10^{-4}	$[7.235 \times 10^{-5}, 4.080 \times 10^{-4}]$	4.764×10^{-4}
$\hat{c}$	0.9138	[0.6276, 1.387]	1.718
$\hat{V}_0$	4.392×10^2	$[1.974 \times 10^1, 1.942 \times 10^3]$	1.607×10^3

(c): TV model

Table 4.4: Parameter values and credible intervals (CrIs) of the IR, TEIV, and TV models estimated using MCMC and MLE respectively by fitting the models to data on the viral load of infected immunocompetent mice (Figure 4.3). MCMC parameter values are not stated for the IR model due to non-identifiability of parameters.

Model	DIC (MCMC)	AIC (MLE)	SSR (MLE)
IR	16.73	79.15	0.5586
TEIV	13.81	19.72	4.032
TV	13.57	20.25	10.25

Table 4.5: Deviance information criterion (DIC) relating to MCMC parameter estimation, and corrected Akaike information criterion (AIC) and the sum of squared residuals (SSR) relating to MLE, of the IR, TEIV, and TV models fitted to data on the viral load of infected immunocompetent mice (Figure 4.3).

4.3.5 Modulating the immune response in the IR model

In practice, the immune response may vary between hosts. Therefore, we investigate how the dynamics of the IR model change as parameters s , τ , k , and r , associated with the immune response, are varied. We introduce the following quantities to summarise the dynamics of the infection: (i) the area under the viral load curve (AUC), (ii) the duration of symptomatic infection, and (iii) the fraction of dead cells at the end of the infection. Following Dobrovolny et al. [91], we define the duration of symptomatic infection as the duration of time for which

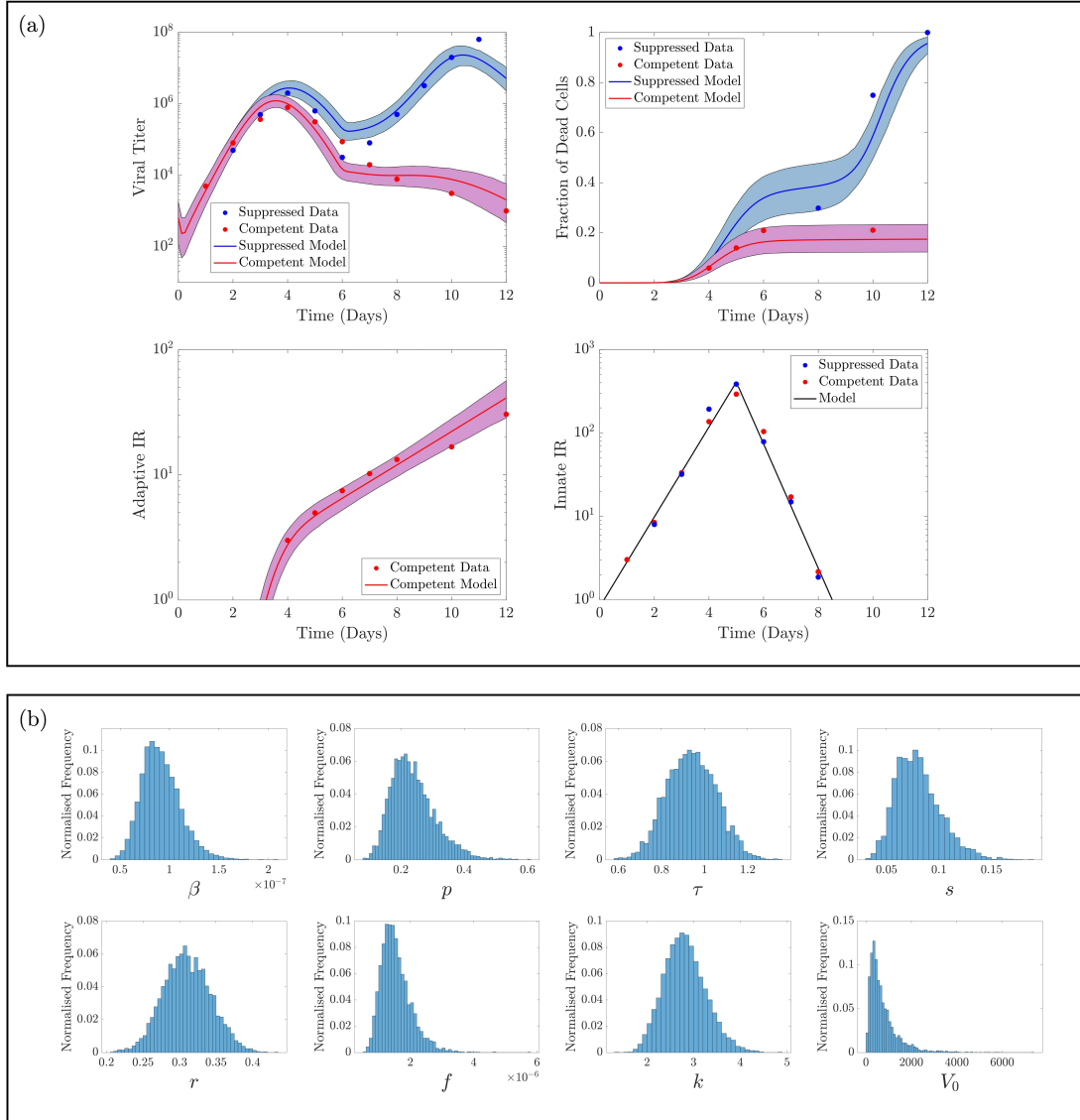


Figure 4.4: (a): The IR model fitted to pooled experimental data on the viral load, fraction of dead cells, and adaptive immune response in immunosuppressed and immunocompetent mice. Solid lines and shaded regions indicate the mean and 95% CRIs of model trajectories based on the posterior distributions of parameters. The innate immune response is assumed to be known precisely and is described by the same function as in Handel et al. [153]. (b): Histograms of the posterior distributions of estimated parameters. Normalised Frequency refers to the probability of selecting an observation within that bin interval. Linear scale used on the x-axes to show tight credible intervals. Fixed and estimated parameter values are stated in Tables 4.1 and 4.6 respectively.

$V > 6.9 \times 10^3$. Hence, letting t_V be the time the viral titer crosses this threshold from above, we define the AUC as $\int_0^{t_V} V(t) dt$, and the fraction of dead cells at the end of the infection as $C(t_V)$. We note that since only the parameter grouping fk affects these quantities, it suffices to vary k alone.

Parameter	MCMC Mean	MCMC 95% CrI	MLE
β	9.145×10^{-8}	$[5.803 \times 10^{-8}, 1.361 \times 10^{-7}]$	9.730×10^{-8}
p	0.2393	[0.1313, 0.4049]	0.2020
τ	0.9401	[0.7199, 1.167]	1.032
s	8.022×10^{-2}	$[4.502 \times 10^{-2}, 1.305 \times 10^{-1}]$	6.854×10^{-2}
r	0.3102	[0.2482, 0.3747]	0.3088
f	1.549×10^{-6}	$[8.382 \times 10^{-7}, 2.708 \times 10^{-6}]$	1.400×10^{-6}
k	2.815	[2.008, 3.786]	2.508
V_0	7.516×10^2	$[1.040 \times 10^2, 2.583 \times 10^3]$	4.870×10^2

Table 4.6: Parameter values and credible intervals of the IR model estimated using MCMC and MLE respectively by fitting the model to pooled data of viral load, fraction of dead cells, and adaptive immune response in immunosuppressed and immunocompetent mice (Figure 4.4).

We varied pairs of parameters relating to the innate immune response (s and τ : Figure 4.5 (a)-(c)) and adaptive immune response (k and r : Figure 4.5 (d)-(f)). In general, as the innate and adaptive immune response parameters are increased, the infection outcome of the infection is improved as the AUC, the final fraction of dead cells, and the infection duration all decrease. We note though that for fixed values of τ ($0 < \tau < 2$ days), the AUC and fraction of dead cells increase as the strength of the innate immune response s is increased. This is because, for this set of parameter values, a stronger initial innate immune response causes fewer cells to be infected early on. Consequently, when the innate immune response dies out, more susceptible cells remain for the virus to infect.

We performed a global parameter sensitivity analysis using the eFAST and Sobol methods (see Appendix C.3). The parameters β , p , and T_0 were found to have the largest influence on the three model outputs. However, the immune response parameters s , τ , r , and k , which may differ between hosts based on their immunological characteristics, significantly influence the dynamics of the IR model, which we shall explore further in the following sections.

4.3.6 Using the TEIV and TV models to reproduce viral load dynamics of the IR Model

We now investigate whether the TEIV and TV models can reproduce the dynamics of the IR model and, in particular, the viral load dynamics. We generate synthetic data from the IR model for different values of the parameters associated with the immune response (i.e. r , k , s , and τ in equations (4.1)–(4.7)). We view $\tilde{\beta}$, $\tilde{p}$, $\tilde{\delta}$, and $\tilde{c}$ as fitting parameters for the TEIV model, and $\hat{\beta}$, $\hat{p}$, and $\hat{c}$ as fitting parameters for the TV model (see Table 4.1). We fix $\tilde{V}_0 = \hat{V}_0 = 100$ PFU (a relatively small starting value) as using the initial viral load as an extra fitting parameter did not improve model fits (results not shown).

For each parameter in the IR model that we vary, there is a region of parameter space where the TEIV and TV models accurately replicate the viral load dynamics (with an SSR of less than 15). In each case, the best-fit values of the TEIV and TV models vary systematically as the respective IR model parameter is increased. With only three fitting parameters, these

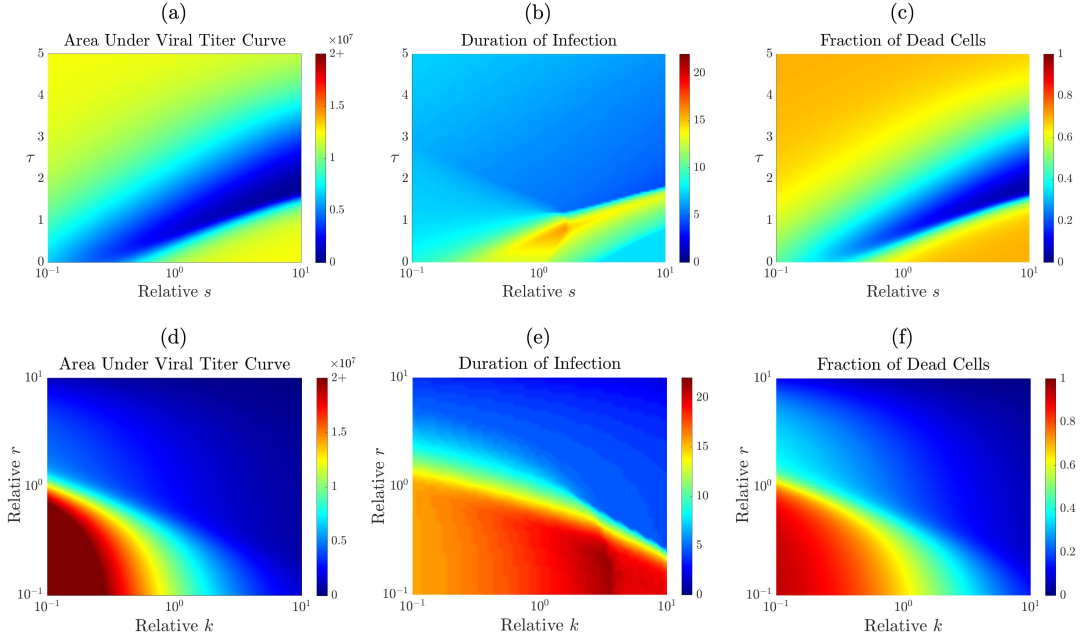


Figure 4.5: The change in the three infection-related quantities of interest as the IR model innate immune response parameters s and τ are varied ((a)-(c)), and the adaptive immune response parameters k and r are varied ((d)-(f)). Scale factors (SF) refer to the parameter values normalised by those given in Table 4.6. The unit of the area under the viral load curve is (PFU days), the unit of the duration is (days), while the fraction of dead cells is dimensionless.

variations are clearer in the TV model. For instance, as the value of s is increased, the best-fit values of $\hat{\beta}$ in the TV model increase, while the values of $\hat{p}$ and $\hat{c}$ decrease (Figure 4.7 (h)). For illustrative purposes, we also show the synthetic data and best-fit models (Figures 4.6 and 4.7 right columns) for parameter values that are 0.1, 0.5, 2, and 10 times the values given in Table 4.6. In general, the TEIV and TV models can accurately replicate dynamics for which the viral load increases exponentially to a peak value and then decreases exponentially (see, for example, the bottom left panels in Figures 4.6 (f) and 4.7 (f)). However, neither model can reproduce dynamics for which the viral load exhibits a resurgence and second peak (see, for example, the top left panels in Figures 4.6 (l) and 4.7 (l)).

When comparing the TEIV and TV model fits, we observe that the SSRs are similar (Figures 4.6 and 4.7 left columns). To directly compare the model fits, we computed the mean SSR for each varied value of the IR parameter (see Table 4.7). In each case, the simpler TV model has a lower mean SSR, suggesting that the extra fitting parameter in the TEIV model does not improve the model fitting. We conclude that the TV model should be used to describe synthetic viral load data generated from the IR model rather than the TEIV model.

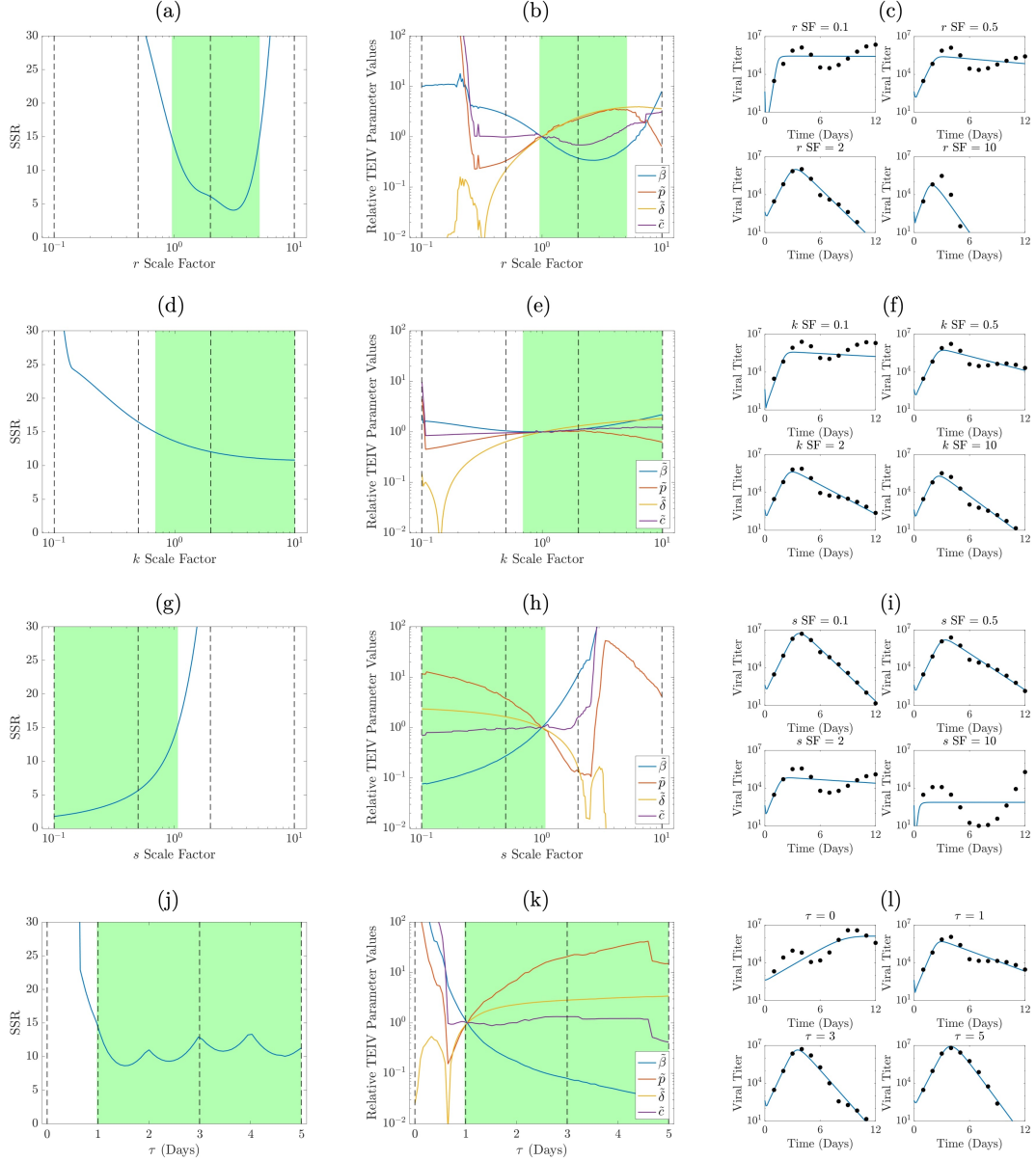


Figure 4.6: The TEIV model fitted to synthetic data as the immune response parameters in the IR model used to generate the data are individually varied. Left column ((a), (d), (g), (j)): The sum of squared residuals (SSR) of the best-fit TEIV model fitted to the synthetically generated data from the IR model as individual parameters of the IR model are varied. Central column ((b), (e), (h), (k)): Best-fit parameter values when the TEIV model is fitted to the synthetic data generated from the IR model. The relative TEIV parameter values refer to parameter values normalised by those given in Table 4.4. Right column ((c), (f), (i), (l)): The synthetic data (black dots) and best-fit TEIV model (blue lines) for different values of the IR model parameters. Green shaded regions represent regions of parameter space where $SSR < 15$ as this indicates a good fit. The parameter values of the IR model are defined in Tables 4.1 and 4.6, unless otherwise stated. Scale factors (SF) refer to IR parameter values normalised by those given in Table 4.6. Vertical dashed lines in the left and central columns refer to scale factors of 0.1, 0.5, 2, and 10 respectively as used to generate synthetic data in the right column. The TEIV model is fitted to the synthetic data using MLE, with $\tilde{V}_0 = 100$ and $\tilde{T}_0 = 7 \times 10^9$.

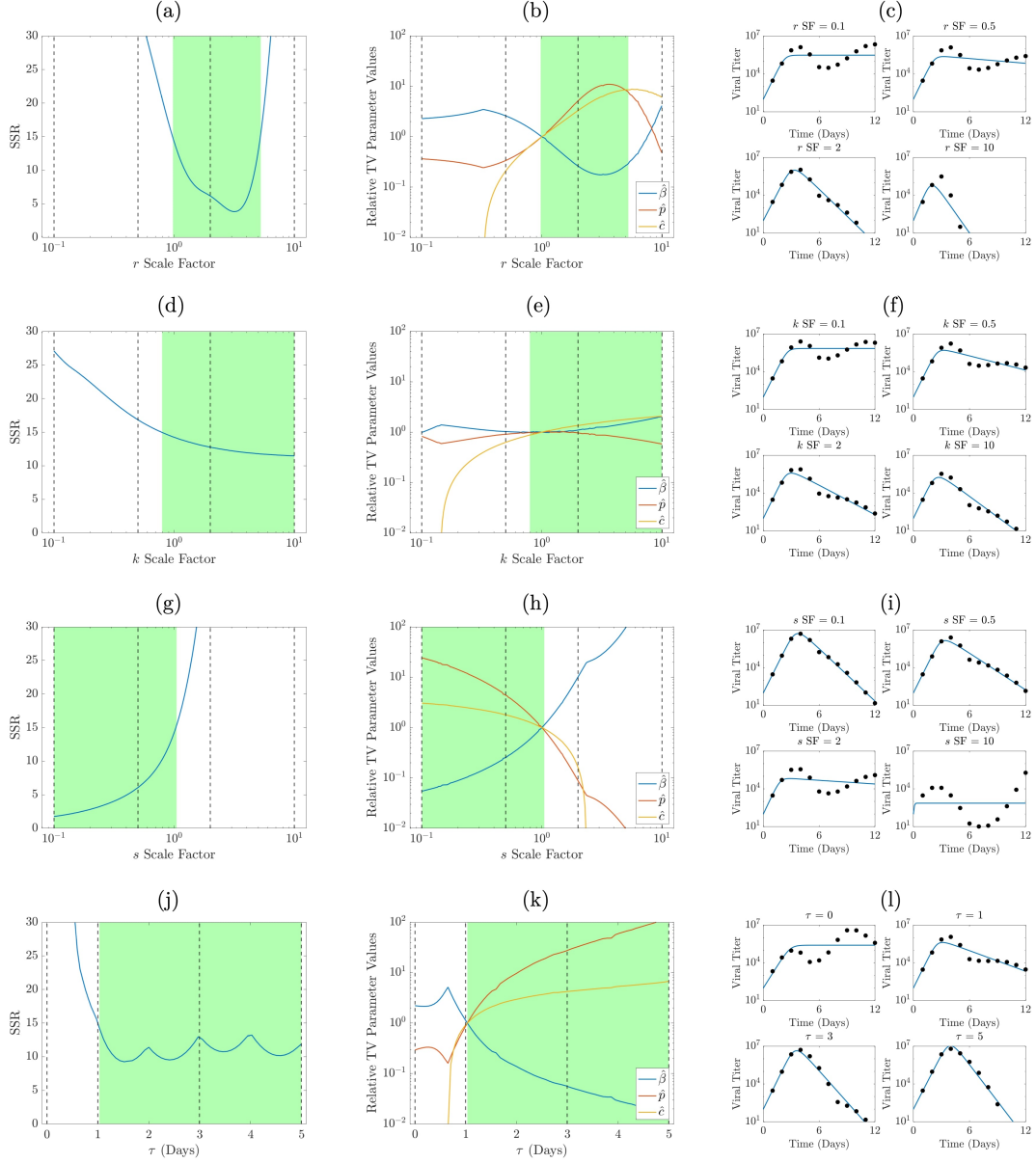


Figure 4.7: The TV model fitted to synthetic data as the immune response parameters in the IR model used to generate the data are individually varied. Left column ((a), (d), (g), (j)): The sum of squared residuals (SSR) of the best-fit TV model fitted to the synthetically generated data from the IR model as individual parameters of the IR model are varied. Central column ((b), (e), (h), (k)): Best-fit parameter values when the TV model is fitted to the synthetic data generated from the IR model. The relative TV parameter values refer to parameter values normalised by those given in Table 4.4. Right column ((c), (f), (i), (l)): The synthetic data (black dots) and best-fit TV model (blue lines) for different values of the IR model parameters. Green shaded regions represent regions of parameter space where $SSR < 15$ as this indicates a good fit. The parameter values of the IR model are defined in Tables 4.1 and 4.6, unless otherwise stated. Scale factors (SF) refer to IR parameter values normalised by those given in Table 4.6. Vertical dashed lines in the left and central columns refer to scale factors of 0.1, 0.5, 2, and 10 respectively as used to generate synthetic data in the right column. The TV model is fitted to the synthetic data using MLE, with $\hat{V}_0 = 100$ and $\hat{T}_0 = 7 \times 10^9$.

IR Model Parameter	Mean SSR	
	TEIV	TV
r	30.53	28.44
k	16.17	16.11
s	61.89	60.50
τ	21.43	19.33

Table 4.7: Mean SSR values as assumed with the fits of the TEIV and TV models to the viral load data generated from the IR model shown in the left columns of Figures 4.6 and 4.7 respectively.

4.3.7 Using the TV model to reproduce viral load dynamics in hosts with different immunological characteristics

In the previous section we showed that the TEIV and TV models could be fit to viral load data generated from the IR model when individual parameters were varied. In practice, the strength of the innate and adaptive immune responses can vary greatly between hosts, and depends on age and the presence of comorbidities such as obesity [155, 180, 184, 216]. We now determine whether the TV model can reproduce viral load dynamics when multiple immune response parameters of the IR model are varied simultaneously. We consider four hypothetical groups of hosts; ‘baseline’, ‘elderly’, ‘obese’, and ‘partially immune’, and use the IR model to generate synthetic viral load data relating to hosts in each group. Although the IR model is parameterised with murine data, we use this parameterisation to explore the qualitative differences that might exist in humans. We assume that the immune response parameters s , τ , and k vary between and within groups. Accordingly, we prescribe gamma distributions for these three parameters (while assuming the remaining model parameters are fixed) to characterise the variations between hosts within each group. The baseline group refers to healthy hosts (i.e. hosts who are not considered to be elderly, obese, or partially immune, and whose parameters have mean values as in Table 4.6).

Influenza severity is strongly linked with age due to the gradual deterioration of the immune system during natural aging [132, 155, 216]. Both the innate and adaptive immune responses are altered during viral infections in the elderly. For example, it has been shown that influenza A infected monocytes from older hosts have impaired antiviral interferon production, and, specifically that IFN- β levels in the monocytes of elderly hosts were half those in similarly infected monocytes of young hosts [302]. The antibody response is also reduced in older individuals, with virus-specific antibody titers shown to be lower in elderly compared with young hosts [43, 123]. Based on these results, we shall assume that in elderly hosts values of the adaptive IR kill rate k and the strength of the innate IR s are half the baseline values given in Table 4.6.

Similarly, obesity is linked with influenza severity as it impairs the host’s innate and adaptive immune responses [241]. For example, when obese hosts are infected with influenza,

their expression of type I interferons is delayed and reduced compared to non-obese hosts [180]. Based on these findings, we characterise obese hosts as having values of s and τ which are half and double their respective baseline values.

We also consider hosts who are partially immune to the virus (due for example to previous infection with an antigenically related strain or vaccination), and, as a result, have stronger immune responses [210, 211, 329]. We assume that their values of k and s are double the baseline values. The mean values of s , τ , and k for each group are summarised in Table 4.8.

We found that the TV model accurately replicates the infection dynamics of individuals in each group (Figure 4.8 (a)). We then generated 10^5 viral load datasets for each group and fit the TV model to each dataset using MLE. Noting that previously an SSR value below 15 corresponded to a good fit, Figure 4.8 (b) shows that the TV model describes the synthetic data accurately for each group. In Figure 4.8 (c)-(k) we show how the distributions of the IR model parameters s , τ , and κ for each group map to the TV model parameters $\hat{\beta}$, $\hat{p}$, and $\hat{c}$. The histograms and scatter plots of the TV model parameters show that for each group, the fitted parameters cluster into different regions of parameter space. We observe that the infection rate $\hat{\beta}$ is smaller for elderly and obese hosts than for the baseline group, whereas the virion production coefficient $\hat{p}$ is larger than the baseline for these two groups. The virion clearance rate $\hat{c}$ is greater than the baseline in the partially immune and obese groups. We conclude that the TV model can accurately reproduce the viral load dynamics of influenza infection in hosts with different immunological characteristics, with different model parameter values for each group.

Group	s	τ	k
Baseline	1	1	1
Elderly	0.5	1	0.5
Obese	0.5	2	1
Partially Immune	2	1	2

Table 4.8: Mean parameter values of the gamma distributions of s , τ , and k for the different groups of hosts, normalised by values given in Table 4.6. These distributions are sampled from for each simulation of the synthetic data to represent variability in parameters between hosts within a population. The assumed standard deviations for each distribution are 0.1 times the mean values. Distributions are shown explicitly in Figure 4.8.

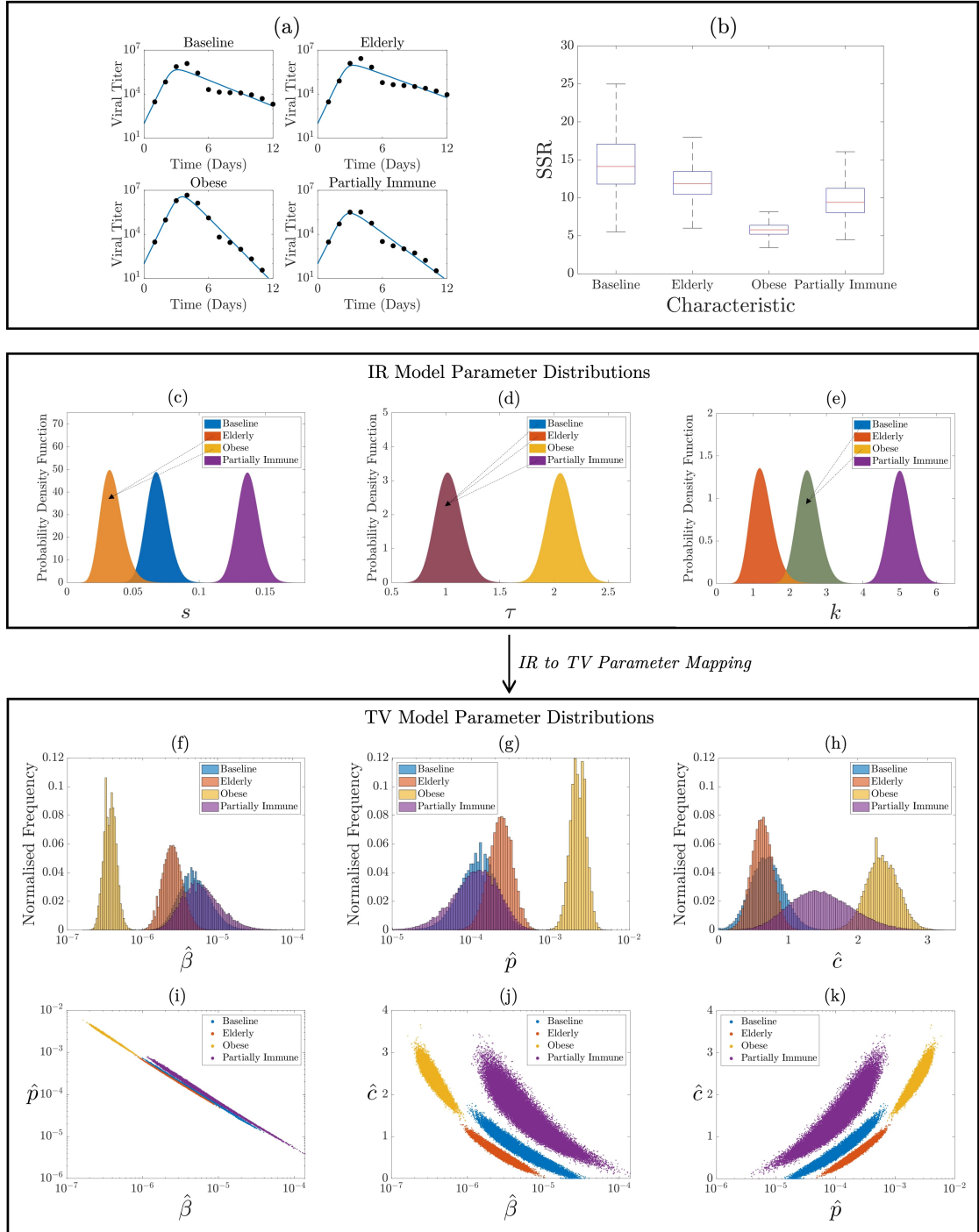


Figure 4.8: The TV model fitted to synthetic data relating to different patient groups. (a): TV model fitted to viral load data relating to the four groups using MLE, with mean values of s , τ and k used to generate the data. (b): Boxplots of the SSR of the TV model fitted to synthetic data relating to hosts from each group. (c)-(e): Distributions of s , τ , and k for the different groups in the IR model. (f)-(k): Histograms and scatter plots of the fitted TV model parameters for each group. For each group, 10^5 synthetic datasets were generated. The TV model was fitted to the synthetic data using MLE, with $\hat{V}_0 = 100$ and $\hat{T}_0 = 7 \times 10^9$. The synthetic data were generated using the IR model with the default parameter values given in Tables 4.1 and 4.6.

4.3.8 Forecasting influenza viral load dynamics

We have used synthetic data to show that the TV model can reproduce the viral load dynamics in different host groups. We now present a practical application of a scenario where the TV model would be preferred to the more complex IR model. Specifically, we consider making real-time forecasts of patient-specific influenza infection viral load dynamics.

In Figure 4.9 we show forecasts made of viral load progression using the TV and IR models. Model parameters were estimated by fitting the models to six viral load training datasets. The training data, along with the test data, were generated from the IR model using parameter values relating to obese hosts as this group had the lowest TV model SSR (Figure 4.8 (b)).

Both models generate accurate forecasts in advance if the parameters are calibrated using training data (Figure 4.9 (a)-(b)). However, the TV model forecast is significantly more accurate when the forecast is updated in real-time at day 6 using the new test data up until that time (Figure 4.9 (c)-(d)); the TV model yields a DIC value of 1252 compared to 3585 for the IR model. This is due to the high dimension of the eight fitting parameters and the fact that the parameters are non-identifiable in the IR model, making it infeasible to update parameters using MCMC estimation. However, as the TV model contains only three fitting parameters and they are all identifiable, it can be used to update forecasts in real-time to improve forecast accuracy (Figure 4.9 (c)). Furthermore, computing forecasts with the TV model is less computationally expensive than using the IR model. Hence, when only observing data for viral load, in this scenario, the simpler TV model should be used to make forecasts. We investigate within-host forecasting further in Chapter 5.

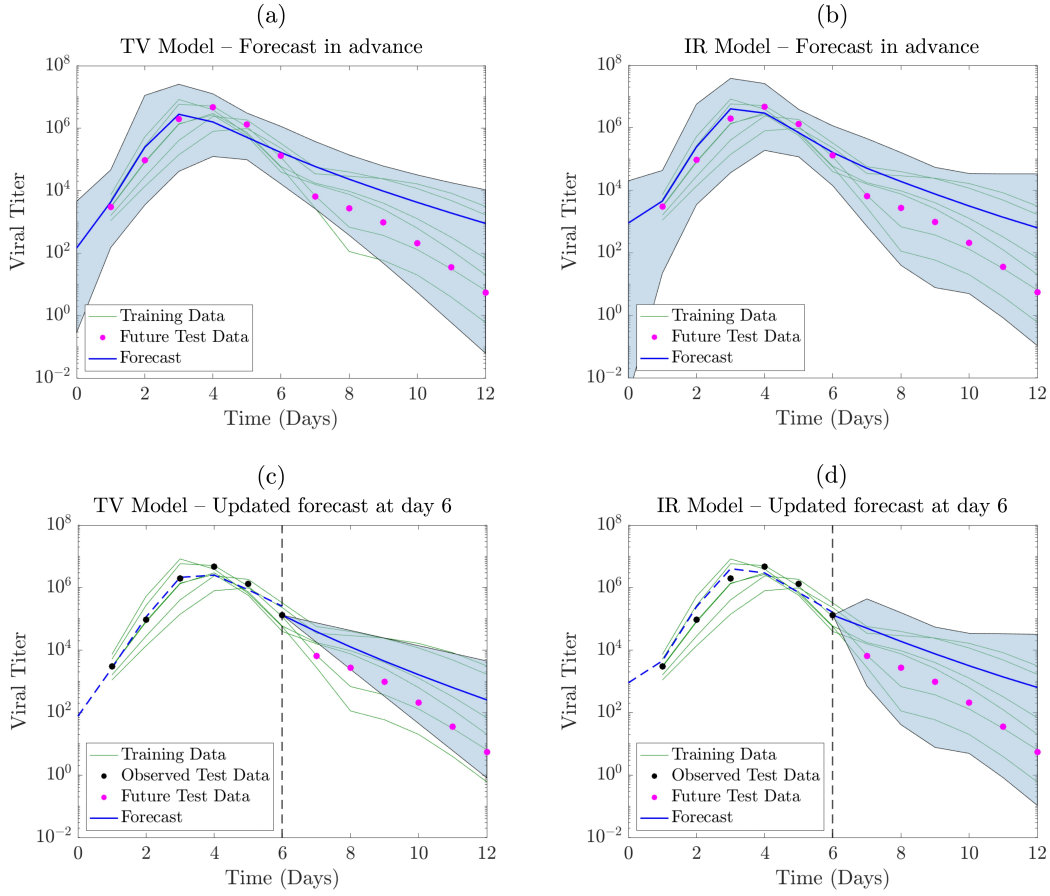


Figure 4.9: Comparison of viral load forecasts obtained using the TV and IR models. Training data from six hosts were used to calibrate the models. Dashed blue lines indicate the mean of the model trajectories during the calibration period. Solid blue lines and shaded regions indicate the mean and 95% CrIs of the forecasts, based on the posterior parameter distributions. Synthetic training and test data were generated using the IR model by sampling obese parameter values (Table 4.8) and fixing the remaining parameters at the default values given in Tables 4.1 and 4.6. DIC values are: (a) 5281, (b) 4050, (c) 1252, and (d) 3585.

4.4 Discussion

Influenza infection is a complex biological process involving multiple interactions between the immune response, cells, and the virus [95, 122]. Consequently, when developing mathematical models for understanding the within-host mechanisms driving viral infections, compromises must be made between biological realism and analytical tractability.

In this chapter, we have considered a hierarchy of ordinary differential equation models which describe influenza infection; the IR model which explicitly models the innate and adaptive immune responses, and the simpler TEIV and TV models which do not. We first performed a structural and practical identifiability analysis, showing that, given perfect data, the unknown parameters of the IR model can be identified if data on the innate and adap-

tive immune responses are available, while the TV model parameters (and some of the TEIV model parameters) can be identified using viral load data alone. We then fit the models to experimental data from Iwasaki and Nozima [189], showing that the TV and TEIV model parameters can be estimated using viral load data from immunocompetent mice, and that the IR model parameters can be estimated using pooled data from immunocompetent and immunosuppressed mice that detail viral load, fraction of dead cells, and immune response levels.

Using the IR model with these estimated parameters, we then generated synthetic viral load data, and showed that the TV model can reproduce the data with accuracy in cases where the data rises to a single peak and then decays. Also, we determined that when considering this synthetic data, the TV model should be preferred over the TEIV model due to having fewer parameters but equivalent model flexibility. Furthermore, the parameters of the TV model that reproduce the data vary systematically as the immune response parameters in the IR model used to generate the data are changed (e.g. as s in the IR model is increased, $\hat{\beta}$ in the TV model increases, while $\hat{p}$ and $\hat{c}$ decrease). Thus, changes in parameter values in the TV model can be used to compare the viral load dynamics of different hypothetical groups of hosts who have different functioning immune responses. Finally, we presented an example of viral load forecasting for which the simpler TV model may be preferred over the IR model. Although we considered influenza, our results may also generalise to other viral infections which can be modelled using target-cell limited and immune response models, such as SARS-CoV-2 [96, 107], West Nile virus [29], Adenovirus [225], Respiratory Syncytial Virus [138], yellow fever virus [261], and Zika virus [36] among others.

The way in which the immune response is modelled is driven by the available data. Thus, to advance our understanding of the immunological processes that affect infection dynamics, it is key that modellers and experimentalists work together in a cycle of data-driven mathematical modelling and model-driven experimental design [268, 343, 344]. Here the roles of interferons in the innate immune response and antibodies in the adaptive immune response are included in the IR model due to experimental data on interferon and antibody levels being available. These mechanisms are incorporated such that interferons reduce viral load, and antibodies clear free virions [153]. However, other mechanisms may be at play and could be incorporated into the IR model. For example, interferons may promote resistance to infection in target epithelial cells, effectively moving them to a refractory class where they are no longer susceptible [172]. Furthermore, although interferon levels may be used as a proxy for the entire innate immune response [152], cellular components such as neutrophils and macrophages may also cause infected cell death and/or clear free virions [159, 239]. Regarding the adaptive immune response, T-cells are also known to kill infected cells [35, 101, 235]. Different expansion and activation mechanisms of the innate and adaptive immune responses could also be considered. In particular, a mechanistic differential equation model which describes the innate

immune response could be formulated to replace equation (4.7), which although attempted in this study as in Handel et al. [153], could only describe the response in the immunocompetent mice and not the immunosuppressed mice that elicited an innate immune response but lacked an adaptive immune response.

Due to the availability of data, viral load has been used as the predominant infection related quantity we have been describing. However, influenza pathogenesis is linked to both viral load and elevated cytokine levels [87], a prime example of the latter being the 1918 influenza pandemic in which it is hypothesised that ‘cytokine storms’ led to severe disease and mortality [296]. Metrics related to cytokine and chemokine expression may be better proxies for disease severity [162, 213, 226]. Thus more work could address the problem of modelling influenza severity and pathogenesis more directly [55], for example by modelling a host’s ‘symptom score’ which is based on the number of symptoms an infected host is suffering from (see Chapter 5). Although the TV model replicated the viral load dynamics well in many of the cases we considered, it could not reproduce complex dynamics such as when the viral load rose to a second peak. Therefore, to describe more complex viral kinetics, a model incorporating the immune response may be needed, along with sufficient data to be able to parameterise it. Different sampling intervals could also be considered, which may affect the ability of the TV model (or any other model) to describe the data. Furthermore, data could also be generated from a stochastic model to account for the inherent randomness in the infection process [23, 152].

We generated synthetic viral load data pertaining to different hypothetical groups of hosts; baseline, elderly, obese, and partially immune. By fitting the TV model to these data, we showed that we could cluster hosts into different groups. If suitable real-world data on different patient groups were available, future work could examine whether or not it is still possible to separate patients into clusters based on the TV model parameters. The goal of this study though is not tightly linked to the specific infection dynamics of any group in particular, but serves to show that a simple model can be used to describe the infection dynamics in groups with sufficiently differently functioning immune responses.

Model selection should be guided by the problem being addressed. If the aim is to make accurate forecasts given past data without infection-mechanism interpretability, then machine-learning [381] or phenomenological mathematical models which do not attempt to describe the biology [82] have their advantages due to their flexibility and ease of use. If the aim is to understand the mechanisms affecting the infection, then more complex mathematical models would be more appropriate. For instance, it was shown in Cao and McCaw [58] that the assumed mechanism of viral control (either target-cell limitation or immune-mediated control) markedly altered the viral kinetics of infection under antiviral treatment and, as such, an immune response model was more suitable for predicting infection dynamics and the effect of antiviral treatment in immunocompetent hosts.

In summary, we have explored which models should be used to describe within-host influenza infection dynamics. Parameter identifiability is often an issue in within-host modelling studies [253, 274]. Given sufficient data relating to the immune response, we showed that the IR model could be parameterised. However, these data are often not commonly collected (especially for infections in humans), and thus only data pertaining to the viral load are typically available. When this is the case, we found that simple target-cell limited models such as the TV model should be used to describe the kinetics of infection [147]. Moreover, we showed that the TV model can differentiate between the infection dynamics in hosts which have differently functioning immune responses. Hence, although target-cell limited models do not fully describe the underlying immunology of infection, their simplicity and parameter identifiability make them a useful tool with regards to describing and predicting infection dynamics, whilst still maintaining biological interpretability and facilitating an understanding of the processes that drive infection.

Chapter 5

Modelling and forecasting the symptom dynamics of influenza

5.1 Introduction

In Chapter 4, we formulated the IR model to describe influenza infection, and showed that a simple TV model can reproduce its dynamics. There we focused on the time evolution of viral load. In practice, however, viral load is typically only measured in a clinical setting, and as previously stated there may be better proxies for the disease severity of influenza. Here, we extend the work from Chapter 4 to account for the severity of influenza symptoms.

Upon infection, individuals often suffer from malaise and fever, followed by respiratory symptoms, headaches, and muscle fatigue. In adults, fever and systemic symptoms usually last approximately three days, while respiratory symptoms, including coughs, may persist for one to two weeks [164]. However, symptoms and patient outcomes vary; young and elderly individuals and those with underlying health conditions are at greater risk of severe illness and mortality [389].

The biological mechanisms that cause influenza symptoms are not fully understood [162]. However, it is thought that respiratory symptoms are caused by cytopathic effects (structural changes in host cells caused by viral invasion) on respiratory epithelial cells, and that systemic symptoms are due to cytokine production [100, 327]. There is likely overlap between these mechanisms, with both the virus and cytokines driving local and general effects [163]. Cytokines that may be responsible for influenza symptoms include IFN- γ , IL-6, IL-10, and TNF- α [148, 244, 400], although it remains to be established which has the biggest effect on symptoms.

In experimental studies, symptoms can be quantified using a ‘symptom score’. Here an individual self-reports their symptoms on a scale, and their total symptom score is the sum of contributions from each symptom [78, 120, 162, 165]. Mathematical models describing influenza symptom scores have been formulated and fit to data. In Canini and Carrat [55], a model is formulated to describe symptom dynamics, and then extended to distinguish between systemic and respiratory symptoms [57]. Models which incorporate symptom dynamics have

also been used to study treatment with neuraminidase inhibitors (NAIs) which suppress viral reproduction [56, 290]. Moreover, models have been fitted to symptom score data for different strains of influenza [240] and to both viral titer and symptom score data [233]. Previous studies have also captured inter-individual variability of virus kinetics and symptoms [55, 233].

While there has been considerable development in within-host viral dynamics modelling [31, 154, 343, 344], little attention has been directed towards real-time forecasts of disease progression at a host level, even though population level forecasts have been widely studied. This may be due to difficulties in collecting quantitative data on disease severity, and the short time-scales of influenza and similar respiratory diseases. In the same manner that population level projections aid in controlling the spread of an epidemic, being able to predict the evolution of symptoms of an individual may help improve patient outcomes. Individuals who may become critically ill could be identified earlier by measuring their symptoms and then referred to receive more specialist treatment or other medical assistance. However, more complex within-host models which describe different aspects of the immune system are less amenable to making real-time forecasts due to problems with parameter identifiability [252], as explored in Chapter 4. Thus, as we shall investigate in this chapter, simple models of host disease progression may be preferred for forecasting.

In this chapter, we investigate whether a two-variable ordinary differential equation model, the ‘TS model’, can be used to forecast the symptom score of an individual infected with influenza in real-time. First, we extend the IR model from Chapter 4 and parameterise a symptom dynamics model, the ‘SD model’, using pooled human data of symptom scores, viral titer, and cytokine levels. We then use the SD model to generate synthetic data relating to symptom scores in groups of healthy and comorbid influenza patients. We investigate whether the progression of symptoms of a host can be predicted in real-time during an infection as this may contribute towards improving patient outcomes, by, for instance, informing treatment methods. We show that if data from previous patients are available, the TS model can accurately predict the symptom scores of a patient in real-time if sufficient data are observed, even if the time of infection is unknown. There is also significant variability in the symptoms displayed by influenza patients which, in part, may be due to underlying health conditions [389]. Similarly to Chapter 4, we find that by analysing the distributions of parameters, the TS model can be used to classify patients into different groups (e.g. healthy and comorbid). Thus, we show how a simple ODE model can be used to forecast symptom dynamics, as a first step towards further work on host level infectious disease forecasting.

5.2 Methods

5.2.1 Models

In this chapter we consider ordinary differential equations models that describe the symptom dynamics of influenza infection. A schematic of the models is shown in Figure 5.1.

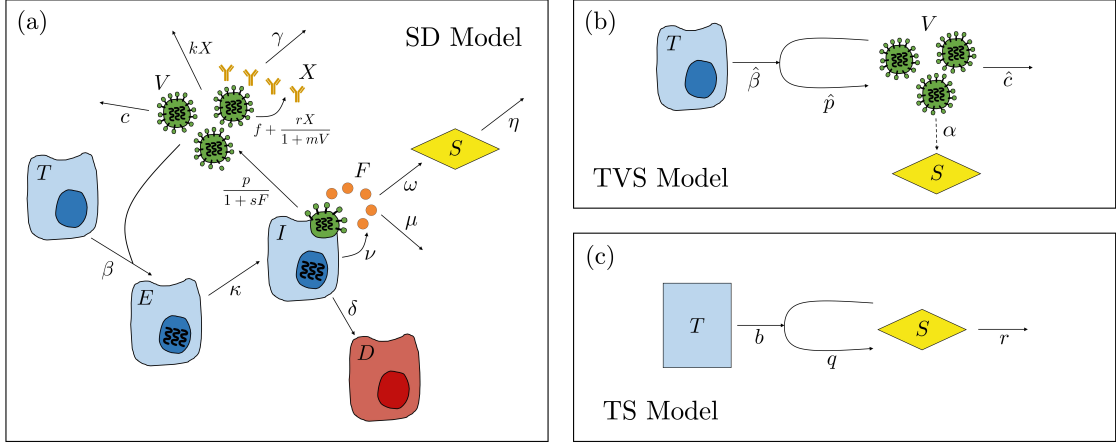


Figure 5.1: Schematic of the models. (a) SD Model: Uninfected epithelial cells, T , are infected by free virions, V , at rate β , and then progress to the eclipse phase (latently infected), E . Latently infected cells become productively infected I , at rate κ . Productively infected cells produce virions at a maximum rate p , which is modulated by the innate immune response, F , with a strength parameter s . Infected cells die at rate δ . Infected cells also activate the innate immune response, F , at rate ν . This response decays at rate μ . The strength of the adaptive immune response, X , increases at a rate which depends on the numbers of free virions and productively infected cells, governed by parameters f , r , and m . It decays at rate γ . Free virions are cleared by the adaptive immune response at rate k , and have a decay rate c . Symptoms, S , are induced by the innate immune response, F , at rate ω , and dissipate at rate η . (b) TVS Model: Free virions, V , are assumed to infect epithelial cells at rate $\hat{\beta}$, and are produced at a rate which is proportional to the rate at which target-cells are infected (constant of proportionality $\hat{p}$). The symptom score, S , is assumed to be proportional to the viral load, V . The viral load is cleared at rate $\hat{c}$. (c) TS Model: A semi-nondimensional model, where the fraction of uninfected cells T , is infected at a rate which is proportional to the symptom score, S , governed by parameters b and q . Symptoms dissipate at rate r .

5.2.1.1 SD model

First, we formulate a model to describe the symptom dynamics of an influenza infection, which we term the SD model. This model extends the IR model considered in Chapter 4, to include differential equations for the symptom dynamics variable, S , and the innate immune response, F (which in Chapter 4 was described by an algebraic function as a differential equation could not describe the innate immune response data for both the immunocompetent and immunosuppressed mice).

As in the IR model, we assume that uninfected epithelial cells, T , become infected by virions at rate β and then become latently infected, E . After a period of $1/\kappa$ days, latently infected cells become productively infected, I , and produce virions, V , at rate p (if the innate immune response is inactive). The lifespan of virions is $1/c$ days, while the lifespan of infected cells is $1/\delta$ days.

Following infection, the innate immune response, F , is activated. We again focus on the role of cytokines, which may be viewed as a proxy for the entire innate immune response

[153]. Infected cells are assumed to activate the innate immune response, F , at rate ν , and F is assumed to decay at rate μ . We suppose that the innate immune response inhibits the rate of production of the virus; the parameter s determines the strength of this reduction. The adaptive immune response, X , (see equation (5.5)) is modelled via the same differential equation as in the IR model.

In the SD model, the symptom score S measures the magnitude of symptoms experienced by a host. We assume that the innate immune response induces symptoms as the kinetics of immunopathology are determined by pro-inflammatory cytokines [368]. Thus, as in Canini and Carrat [55], we assume that S increases at a rate proportional to F at rate ω , and that symptoms dissipate via first order kinetics, at rate η . Combining the above assumptions we arrive at the following system of equations which define the SD model:

$$\text{uninfected cells} \quad \frac{dT}{dt} = -\beta TV, \quad (5.1)$$

$$\text{latently infected cells} \quad \frac{dE}{dt} = \beta TV - \kappa E, \quad (5.2)$$

$$\text{productively infected cells} \quad \frac{dI}{dt} = \kappa E - \delta I, \quad (5.3)$$

$$\text{free virions} \quad \frac{dV}{dt} = \frac{pI}{1+sF} - cV - kVX, \quad (5.4)$$

$$\text{adaptive IR} \quad \frac{dX}{dt} = fV + \frac{rXV}{1+mV} - \gamma X, \quad (5.5)$$

$$\text{innate IR} \quad \frac{dF}{dt} = \nu I - \mu F, \quad (5.6)$$

$$\text{symptom score} \quad \frac{dS}{dt} = \omega F - \eta S. \quad (5.7)$$

Descriptions and estimated values of model parameters are given in Table 5.1 (a). Throughout this chapter, the unit of V is viral titer measured in TCID₅₀/ml, the unit of F is cytokine concentration measured in pg/ml, and S is dimensionless. As in Chapter 4 we denote the unit of X by u_X . Initially, the immune response and symptom score variables are zero such that $X(0) = F(0) = S(0) = 0$. We note that although each of the individual mechanisms in the model have been previously used in the literature, the combination of these mechanisms to describe symptom dynamics makes this model novel.

5.2.1.2 TS model

Our aim is to describe and forecast the symptom dynamics S of a viral infection. In what follows, the SD model will be used to generate synthetic symptom score data but due to the larger number of unknown parameters it would be infeasible to fit the SD model to data and make forecasts (if only symptom score data are observed). Hence we consider an adaptation of the TV model in Chapter 4 in which symptom output scores S is proportional to the viral

load. Hence, in dimensional form, the ‘TVS model’ is given by

$$\frac{dT}{dt} = -\hat{\beta}TV, \quad T(0) = T_0, \quad (5.8)$$

$$\frac{dV}{dt} = \hat{p}\hat{\beta}TV - \hat{c}V, \quad V(0) = V_0, \quad (5.9)$$

$$S = \alpha V. \quad (5.10)$$

Due to our focus on the symptom score S , we nondimensionalise V and T , writing $T = T^*\tilde{T}$ and $V = V^*\tilde{V}$, where $\tilde{T}$ and $\tilde{V}$ are dimensionless, and T^* and V^* are the respective dimensional scaling parameters. We set $T^* = T_0$ and $V^* = 1/\alpha$, relabel $b = \hat{\beta}/\alpha$, $q = \hat{p}\hat{\beta}T_0$, $r = \hat{c}$, and assume $S_0 = \alpha V_0$, which gives $S = \tilde{V}$. After dropping the tildes for ease of notation, we write the model as

$$\frac{dT}{dt} = -bTS, \quad T(0) = 1, \quad (5.11)$$

$$\frac{dS}{dt} = qTS - rS, \quad S(0) = S_0. \quad (5.12)$$

We shall refer to this model as the TS model, and henceforth use it to forecast symptom scores. We note that as we are only considering S we need not explicitly model the viral load, but $V = S/\alpha$ so viral load can be easily determined. Descriptions of model parameters for the TVS and TS models are given in Tables 5.1 (b) and (c) respectively. We note the TS model is mathematically equivalent to the TV model from Chapter 4, but with $S = V$ and the number of cells nondimensionalised.

5.2.2 Experimental data

We focus on an experimental study in which eight human volunteers were infected with influenza A/Texas/36/91 (H1N1) and data on their viral titers, cytokine levels, and symptom scores were collected over a 7-day period [120]. The data for each quantity are the mean values of eight volunteers from a placebo group. The volunteers assessed their symptoms on a 4-point scale with scores of 0 and 3 corresponding to no symptoms and severe symptoms respectively. The symptoms assessed were nasal stuffiness, runny nose, sore throat, cough, sneezing, earache/pressure, breathing difficulty, muscle aches, fatigue, headache, feverish feeling, hoarseness, chest discomfort, and overall discomfort. The individual symptom scores were divided into three groups (systemic symptoms, upper respiratory symptoms, and lower respiratory symptoms). We sum the scores from each symptom group to arrive at a total symptom score. Levels of cytokines IL-6, IFN- γ , TNF- α , and IL-10 were measured in pg/ml. We sum their concentrations to arrive at a total cytokine concentration (a similar approach was adopted by Hernandez-Vargas et al. [172]). Viral titer was measured in TCID₅₀/ml. Further details of the experimental procedure are given in the original study [120]. Data were extracted from Figures 1, 3, and 5 in Fritz et al. [120] using the data extraction tool <https://automeris.io/WebPlotDigitizer/>.

Parameter	Description	Values (Units)	Source
β	Infection rate	Fitting parameter $((\text{TCID}_{50}/\text{ml})^{-1} \text{ days}^{-1})$	
p	Virion production rate	Fitting parameter $(\text{TCID}_{50}/\text{ml} \text{ days}^{-1} \text{ cells}^{-1})$	
$1/\delta$	Lifespan of infected cells	2 days	[153]
$1/c$	Lifespan of free virions	10 days	[153]
$1/\kappa$	Duration of latent eclipse phase	4 days	[153]
s	Strength of innate IR	0.0686 ml/pg	Chapter 4
r	Expansion rate of adaptive IR	0.309 $(\text{TCID}_{50}/\text{ml})^{-1} \text{ days}^{-1}$	Chapter 4
f	Recruitment rate of adaptive IR	$1.40 \times 10^{-6} \text{ u}_X (\text{TCID}_{50}/\text{ml})^{-1} \text{ days}^{-1}$	Chapter 4
m	Adaptive IR growth parameter	1	
γ	Decay rate of adaptive IR	1.159 days^{-1}	[90]
k	Kill rate of adaptive IR	$2510 \text{ u}_X^{-1} \text{ days}^{-1}$	Est. based on Chapter 4
ν	Expansion rate of innate IR	Fitting parameter $(\text{pg}/\text{ml} \text{ days}^{-1} \text{ cells}^{-1})$	
μ	Decay rate of innate IR	Fitting parameter (days^{-1})	
ω	Expansion rate of symptom score	Fitting parameter $(\text{ml}/\text{pg} \text{ days}^{-1})$	
η	Decay rate of symptom score	Fitting parameter (days^{-1})	
T_0	Initial no. of uninfected cells	7×10^9 cells	[153]
V_0	Initial viral load	1 $\text{TCID}_{50}/\text{ml}$	[120]

(a): SD model - equations (5.1)–(5.7).

Parameter	Description	Values (Units)
$\hat{\beta}$	Infection rate	- $((\text{TCID}_{50}/\text{ml})^{-1} \text{ days}^{-1})$
$\hat{p}$	Virion production rate	- $(\text{TCID}_{50}/\text{ml} \text{ cells}^{-1})$
$\hat{c}$	Clearance rate of virions	- (days^{-1})
α	Symptom score parameter	- (virions^{-1})
T_0	Initial no. of uninfected cells	- (cells)
V_0	Initial viral load	- $(\text{TCID}_{50}/\text{ml})$

(b): TVS model - equations (5.8)–(5.10).

Parameter	Description	Values (Units)
b	Infection rate $(= \hat{\beta}/\alpha)$	Fitting parameter (days^{-1})
q	Symptom growth rate $(= \hat{p}\hat{\beta}T_0)$	Fitting parameter (days^{-1})
r	Symptom decay rate $(= \hat{c})$	Fitting parameter (days^{-1})
S_0	Initial symptom score	0.01

(c): TS model - equations (5.11)–(5.12).

Table 5.1: Descriptions and parameter values of the SD, TVS, and TS models.

5.2.3 Parameter estimation and forecasting

We perform parameter estimation and forecasting using the methods used in Chapters 3 and 4. In brief, when fitting models to data, parameters are estimated using Markov chain Monte Carlo (MCMC) with the Metropolis-Hastings algorithm [160]. We also determine the maximum likelihood estimation (MLE) parameter estimates when fitting the SD model to data from Fritz et al. [120]. See Appendix D.1 for further details.

When making real-time forecasts, we calibrate our model forecasts using observed symptom score data for days $0, 1, \dots, m$ of the infection, estimating model parameters using the MCMC method described above. We use the estimated distributions of these model parameter values to project the model forwards, starting at day m with initial conditions based on

the symptom score at day m .

To compare different model forecasts throughout this study, as in Chapter 4, we use the deviance information criterion (DIC) stated in Gelman et al. [128]. DIC values are given in Appendix D.5.

5.2.4 Generating synthetic data

Throughout this chapter we generate synthetic symptom score data using the SD model. We consider two groups of patients, ‘healthy’ and ‘comorbid’. To account for intra-group variation, we prescribe gamma distributions for the parameters β , p , s , r , f , k , ν , μ , ω , and $\tilde{\eta} = \eta/\omega$ to characterise the variations between hosts within each group. When fitting the SD model to real-world data we will show that the parameter grouping $\tilde{\eta} = \eta/\omega$, rather than η , has a larger influence on the model kinetics (see Appendices D.2 and D.3 for further details). For healthy patients, the mean values of these gamma distributions are given in Tables 5.1 and Table 5.2 (MLE parameters). Comorbid patients are characterised by values of $\tilde{\eta}$ which are half those of patients in the healthy group, which results in a larger peak symptom score. We fix the standard deviations of the gamma distributions for each parameter to be 0.3 times the mean parameter value. This ensures that the all parameters have the same relative variability regardless of their absolute values.

To add realism to the synthetic data, we add normally distributed noise, scaling noise levels with the square root of the amplitude of the data. The noise in the synthetic data are representative of the level of noise in the experimental data [120]. The synthetic data are outputted every 12 hours over a period of 10 days.

5.2.5 Classification

Here we outline the method used to classify patients as either healthy or comorbid by comparing the posterior distributions of the TS model parameters. Let us define the combined matrix of posterior distribution parameter values relating to the TS model fitted to previously observed healthy and comorbid patient synthetic symptom score data by θ_H and θ_C respectively. These consist of three-dimensional vectors θ_{Hi} and θ_{Ci} relating to parameter values b , q , and r . The index i relates to each previously observed patient, where in this study $i = 1, \dots, 12$. We decompose the three-dimensional parameter space of b , q , and r into an $R \times R \times R$ grid of ‘subcubes’. We define the density values m_{Hj} and m_{Cj} where $j = 1, \dots, R^3$ as the number of parameter value vectors θ_{Hi} and θ_{Ci} which lie in each respective subcube.

Now let us define a matrix of posterior distribution parameter values relating to the TS model fitted to a new patient as θ_P and calculate its density values m_{Pj} for the same $R \times R \times R$ grid of subcubes. We can then determine the similarity between θ_P and θ_H , and θ_P and θ_C ,

respectively, by calculating

$$D_H = \frac{\sum_{j=1}^{R^3} m_{Hj} \times m_{Pj}}{\sum_{j=1}^{R^3} m_{Hj}}, \quad D_C = \frac{\sum_{j=1}^{R^3} m_{Cj} \times m_{Pj}}{\sum_{j=1}^{R^3} m_{Cj}}. \quad (5.13)$$

If $D_H > D_C$, we classify a patient as healthy, and if $D_C > D_H$, we classify them as comorbid. We use $R = 32$ in this study as this decomposes the parameter spaces finely enough to differentiate between the two groups of patients in general, but is still small enough such that the classification is computationally feasible.

5.3 Results

5.3.1 Fitting the SD model to pooled experimental data

We fit the SD model to pooled data of viral load, the innate immune response (cytokine levels), and symptom scores from Fritz et al. [120]. To the best of our knowledge, this is the first time a host level influenza model has been fit to these three measurements simultaneously. The results presented in Figure 5.2 show that the model describes the data accurately, capturing the initial increase and subsequent decrease in viral load, as well as the growth and decay of the innate immune response and symptom scores. In model simulations, the viral load peaks after 1.5 days, while the innate immune response and symptoms peak after approximately 2.5 days. We fit the six unknown parameters β , p , ν , μ , η , and ω , using both MCMC and MLE; the posterior distributions calculated using MCMC are presented in Figure 5.2 (b). The distributions are unimodal, suggesting parameter identifiability [338], and those for β , p , ν , and μ have tight credible intervals (CrI; see Table 5.2). The credible intervals for ω and η , which govern the symptom dynamics, span several orders of magnitude. However, they are highly correlated (see Appendix D.2). Therefore we shall define $\tilde{\eta} = \eta/\mu$ and write equation (5.7) as $\frac{dS}{dt} = \omega(F - \tilde{\eta}S)$ for the remainder of this study, such that we can generate synthetic data which display a wider range of dynamics.

Parameter	MCMC Mean	MCMC 95% CrI	MLE
β	1.203×10^{-3}	$[4.602 \times 10^{-4}, 2.752 \times 10^{-3}]$	1.164×10^{-3}
p	7.074×10^{-4}	$[1.895 \times 10^{-4}, 1.704 \times 10^{-3}]$	5.868×10^{-4}
ν	7.250×10^{-7}	$[5.679 \times 10^{-7}, 8.992 \times 10^{-7}]$	7.211×10^{-7}
μ	0.7675	[0.6059, 0.9502]	0.7876
ω	22.20	[0.7180, 152.40]	1.606
η	478.6	[15.49, 3278]	33.92

Table 5.2: Mean and 95% CrIs of estimated parameters using MCMC parameter estimation, and estimated parameters using MLE, of the SD model fitted to data for viral load, innate immune response (cytokines) and symptom score from Fritz et al. [120] (Figure 5.2).

5.3.2 Fitting the TS model to synthetic symptom score data

We generate synthetic data of symptom scores using the SD model for two groups of patients. The infection dynamics of healthy patients are simulated using parameter values estimated

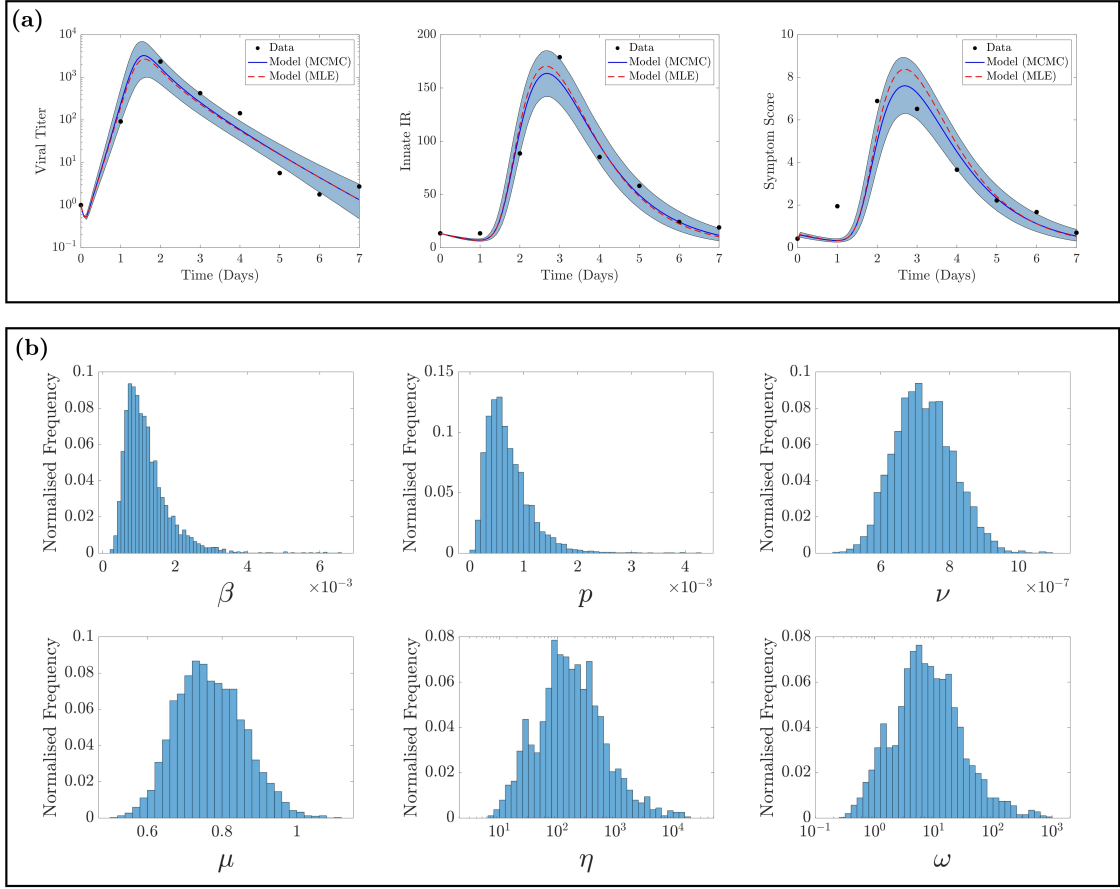


Figure 5.2: (a): The SD model (equations (5.1)–(5.7)) fitted to experimental data on the viral load, innate immune response (cytokines), and symptom score from Fritz et al. [120]. Solid blue lines and shaded regions indicate the mean and 95% CrIs of model trajectories based on the posterior distributions of parameters. Red dashed lines indicate the model trajectory relating to the maximum likelihood estimated parameter set. (b): Posterior distributions of fitted model parameters computed using MCMC parameter estimation. Normalised Frequency refers to the probability of selecting an observation within that bin interval. Fixed and estimated parameter values are stated in Tables 5.1 and 5.2 respectively.

from the real-world experimental data and stated in Table 5.2. We suppose that comorbid patients are assumed to recover more slowly than healthy patients by assuming that the value of their symptom recovery rate $\tilde{\eta}$ is half that of healthy patients, while other parameter values are the same. In Figure 5.3 we fit the TS model to synthetic symptom score data from a healthy and comorbid patient. In both cases, the TS model fits the data well. Moreover, the estimated model parameters are shown to be identifiable and to localise in distinct regions of parameter space (Figure 5.3 (b)-(c) and Table 5.3); the comorbid patient has a larger symptom growth rate q while a lower infection rate b and symptom decay rate r , compared to the healthy patient. This indicates that for this set of data, the TS model can distinguish symptom dynamics in these two patient groups, with parameter values that lie in distinct regions of parameter space.

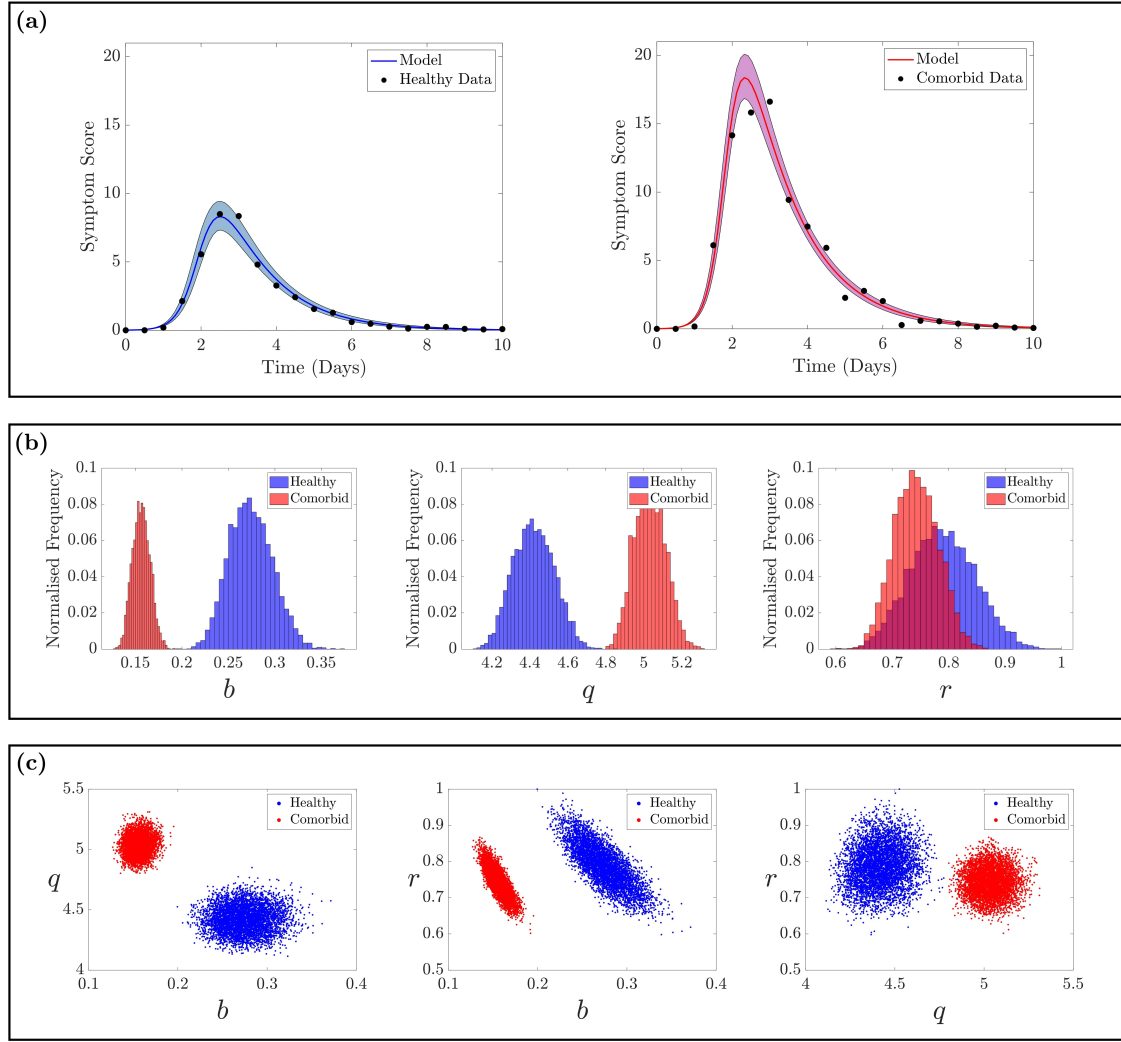


Figure 5.3: The TS model (equations (5.11)–(5.12)) fitted to synthetic symptom score data generated from the SD model for a healthy and comorbid patient. (a): Solid lines and shaded regions indicate means and 95% CrIs of model trajectories based on the posterior distributions of parameters. (b)–(c): Histograms and scatter plots of estimated parameter values of the TS model for the healthy and comorbid patients. Normalised Frequency refers to the probability of selecting an observation within that bin interval. The synthetic data were generated from the SD model using the parameter values stated in Table 5.1 and the mean MCMC estimated parameters in Table 5.2, with normally distributed noise added.

5.3.3 Forecasting symptom scores in real-time using the TS model

We have shown that the TS model can describe symptom score data accurately when the entire infection is observed. We now attempt to predict future symptom scores, in real-time, and update our forecasts as new data becomes available. First, we forecast a patient's symptom score in real-time, without using prior information about the symptom progression from previous patients. Before the peak of symptoms, forecasts are uncertain, and it is difficult to predict when the peak will occur (Figure 5.4 (a)–(b)). At 3 days, just after the peak (Figure

Parameter	Mean	95% CrI
b	0.2741	[0.2313, 0.3210]
q	4.421	[4.223, 4.632]
r	0.7892	[0.6800, 0.9052]

(a): Healthy synthetic symptom score data.

Parameter	Mean	95% CrI
b	0.1564	[0.1387, 0.1753]
q	5.033	[4.878, 5.195]
r	0.7426	[0.6696, 0.8185]

(b): Comorbid synthetic symptom score data.

Table 5.3: Mean and 95% CrIs of estimated parameters using MCMC parameter estimation of the TS model fitted to synthetic symptom score data generated from the SD model for a healthy and comorbid patient (Figure 5.3).

5.4 (c)-(d)), forecasts remain uncertain. While forecasts made at day 4 are more accurate as demonstrated by the smaller DIC values (see Table D.1), they are less useful as the peak has passed and symptoms are diminishing.

In order to investigate whether incorporating data from previous patients improves forecasting predictions, we generated synthetic data of the entire course of infection for 24 patients (12 healthy and 12 comorbid shown explicitly in Figure D.3). We used these data to compute informative priors by fitting the TS model to these data. In this case, the TS model can predict the future symptom scores of the new patients accurately, even before the peak of their symptoms (Figure 5.5 (a)-(b)). As expected, precision and accuracy of forecasts improve as more data are observed (Figure 5.5 (c)-(f) and Table D.1).

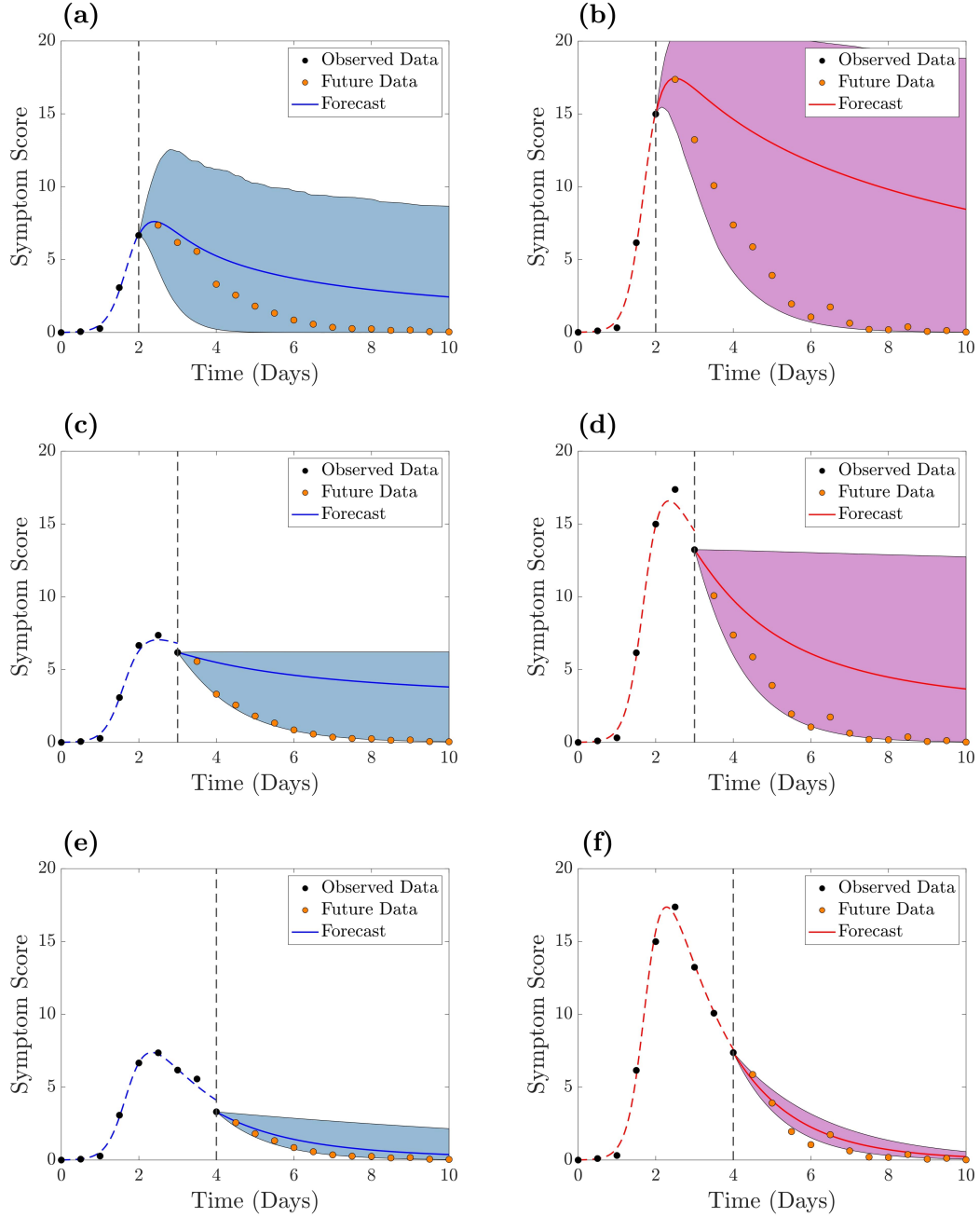


Figure 5.4: Real-time forecasts obtained from the TS model of the symptom score dynamics of a healthy and comorbid patient using no prior information from previous patients (i.e. uninformative priors). Forecasts were made at days 2, 3, and 4 after infection, using the observed data (black dots) up until those times to estimate model parameters. Dashed vertical lines separate the calibration and forecasting periods. Other dashed lines indicate the mean of the symptom score trajectories in the model calibration period. Solid lines and shaded regions indicate the mean and 95% CrIs of the forecasts, based on the posterior distributions of the parameters. The synthetic data were generated from the SD model using parameter values stated in Tables 5.1 and 5.2, with normally distributed noise added. DIC values are given in Table D.1.

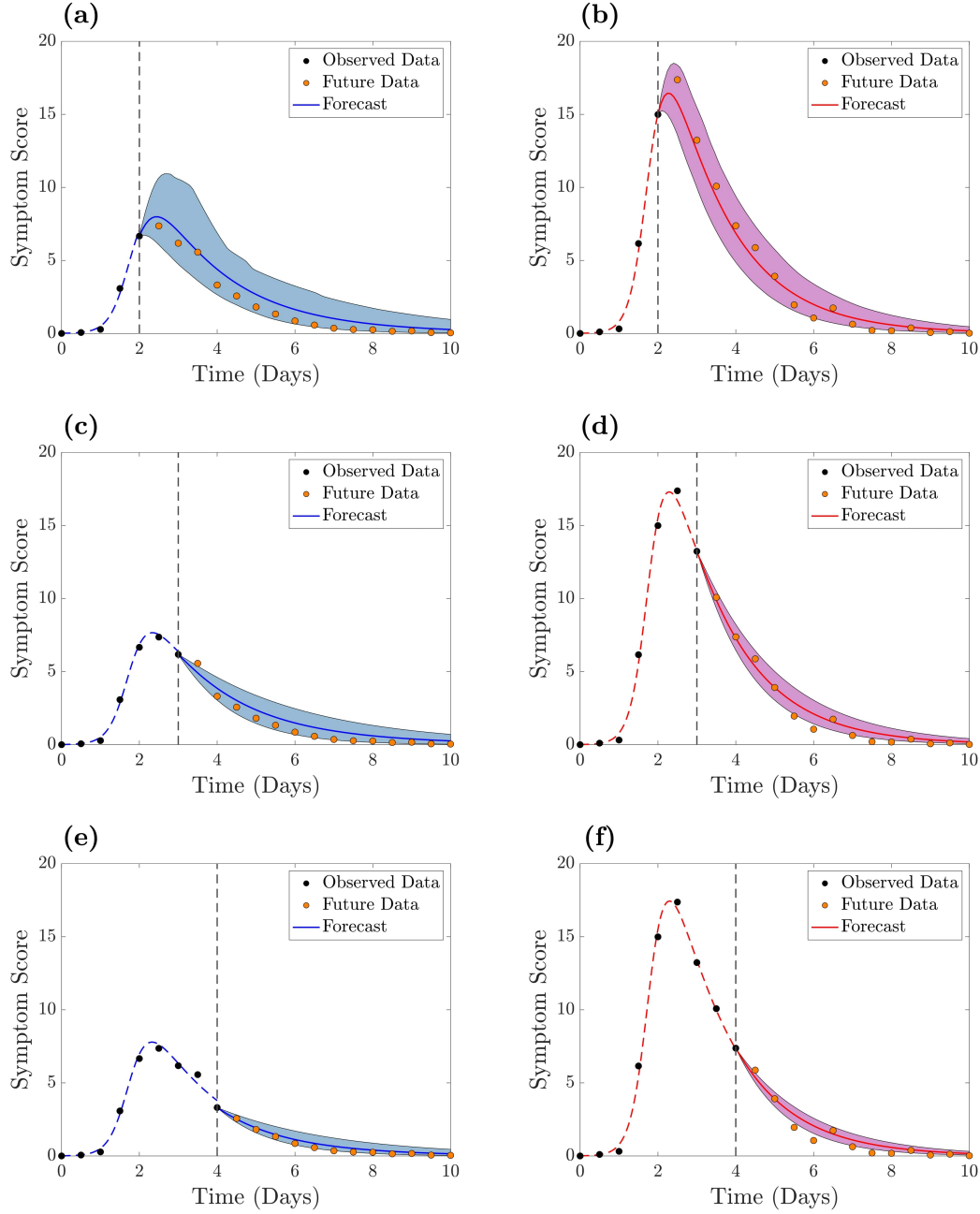


Figure 5.5: Real-time forecasts obtained from the TS model of the symptom score dynamics of a healthy and comorbid patient using prior information of symptom scores from 24 previous patients (i.e. informative priors are set on the fitting parameters b , q , and r). Forecasts were made at days 2, 3, and 4 after infection, using the observed data up until those times to estimate model parameters. Dashed vertical lines separate the calibration and forecasting periods. Other dashed lines indicate the mean of the symptom score trajectories in the model calibration period. Solid lines and shaded regions indicate the mean and 95% CrIs of the forecasts, based on the posterior distributions of the parameters. The synthetic data were generated from the SD model using parameter values stated in Tables 5.1 and 5.2, with normally distributed noise added. DIC values are given in Table D.1.

So far we have made forecasts using observed data from the beginning of the infection, where the time of infection was known. However, patient data are often only collected later in infection [406]. We now consider the scenario when the first symptom score datapoint is observed at some unknown time $t_0 = 1, 2$, or 3 days after infection respectively. We fit the TS model to the observed data, now using t_0 as a fitting parameter along with b , q , and r . As the true value of t_0 increases from 1 (Figure 5.6 (a)-(b)), to 2 (Figure 5.6 (c)-(d)), to 3 (Figure 5.6 (e)-(f)), the simulations made backwards in time from the time of the first observed datapoint to the time of infection become more uncertain. If only data for after the peak are observed, it is challenging to predict when the peak occurred and hence make forecasts of the unobserved previous symptom dynamics. In each case though, as shown by the posterior distributions in each panel, the true value of t_0 falls well within the 95% CrIs of the distributions.

Following on, we next consider making real-time forecasts where the time of infection is unknown. As before, we consider three cases where the true time of infection is 1, 2, and 3 days before the first observed datapoint. Now though, to investigate the capability of using the model to make forecasts with limited data over a short time-span, we only observe three datapoints over a period of one day. Thus this forecasting study is directly comparable to Figure 5.5, the difference here though being that we do not observe the initial data and do not know the time of infection. As in Figure 5.5, we use synthetic symptom score data from 24 previous patients to inform forecasts. Despite only observing three datapoints, the TS model accurately predicts the future course of the symptom dynamics (Figure 5.7), with the forecasts only slightly less accurate than when the initial course of the infection was also observed (see Figure 5.5 and Table D.1 for comparison of DIC values).

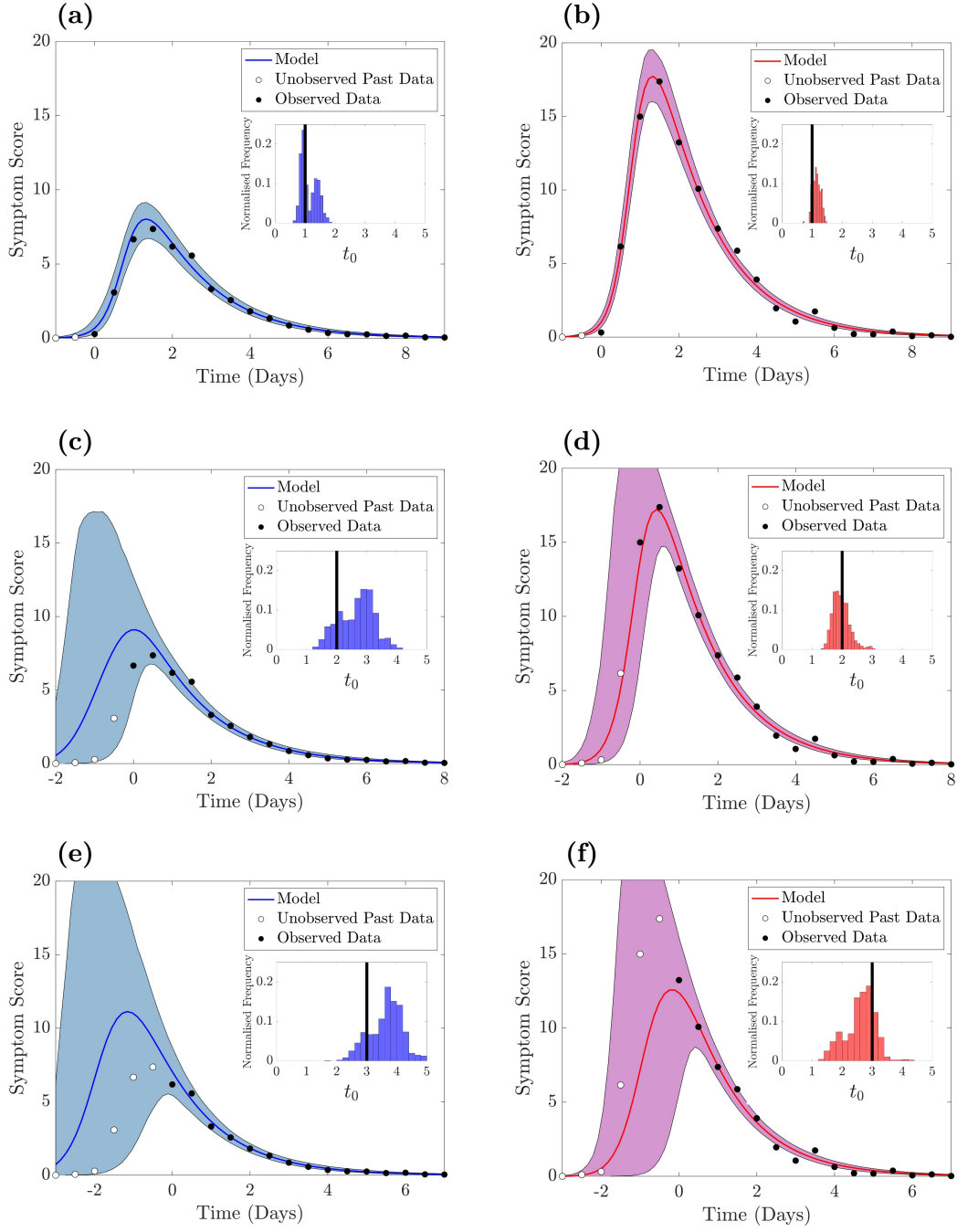


Figure 5.6: The TS model fitted to the symptom score data of healthy and comorbid patients, when the first symptom score datapoint is observed at some unknown time t_0 after infection. Data at earlier times are not observed (white dots); the model is fit to the observed data (black dots), and t_0 , b , q , and r are fitting parameters. The posterior distributions of t_0 are shown within each panel, where the true values of t_0 are 1 (a)-(b), 2 (c)-(d), and 3 days (e)-(f) respectively, shown by vertical black lines. Solid lines and shaded regions indicate the mean and 95% CrIs of model trajectories based on the posterior distributions of parameters. The synthetic data were generated from the SD model using parameter values stated in Tables 5.1 and 5.2, with normally distributed noise added. DIC values are given in Table D.1.

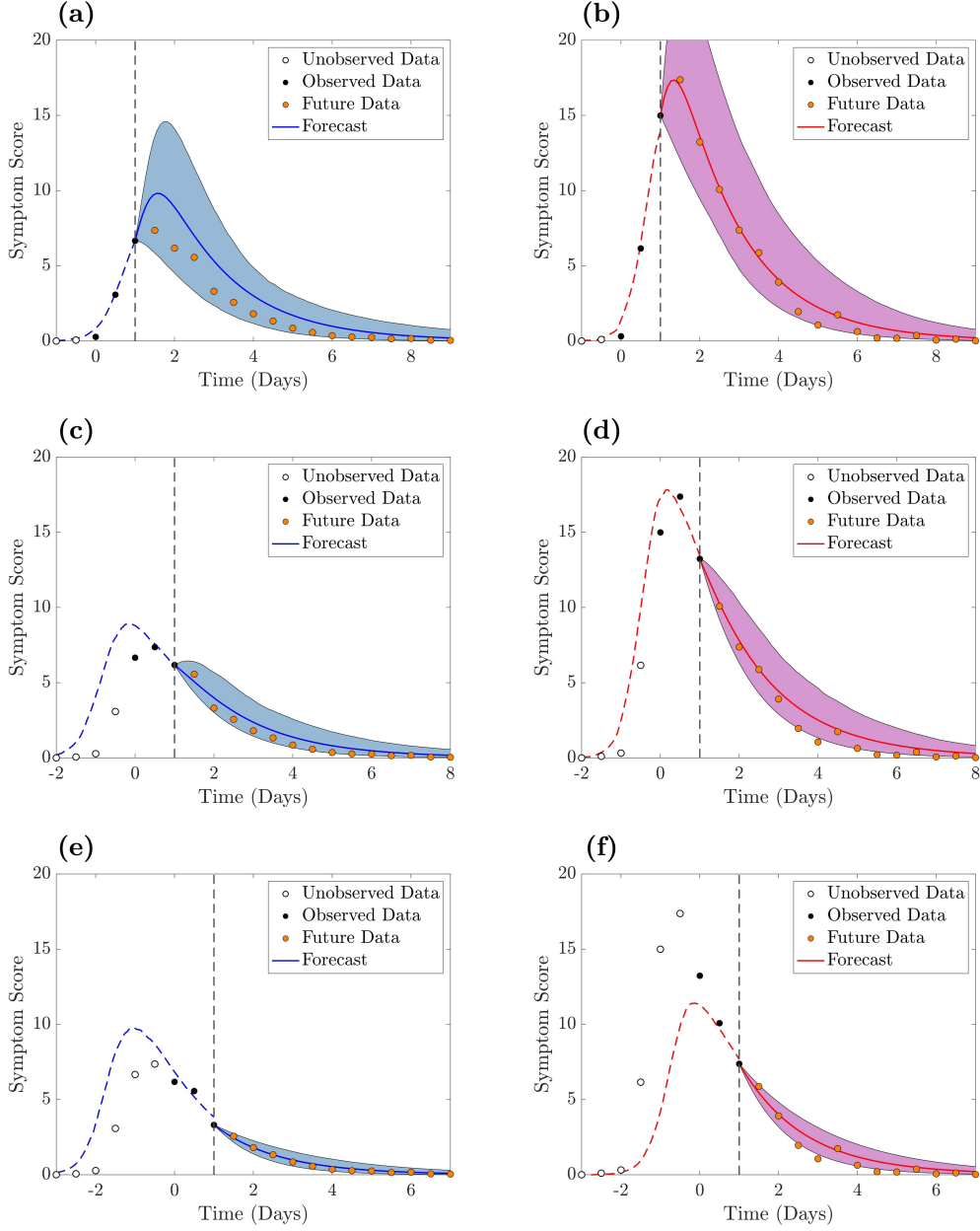


Figure 5.7: Real-time forecasts obtained from the TS model of the symptom score dynamics of a healthy and comorbid patient where the time of infection is unknown. Prior information of symptom scores from 24 previous patients is used (i.e. informative priors are set on the fitting parameters b , q , and r). A non-negative uniform prior is set on the fitting parameter t_0 . In each case, forecasts are made with three observed datapoints (black dots) to estimate model parameters, where the true values of t_0 are 1 (a)-(b), 2 (c)-(d), and 3 days (e)-(f) respectively. Dashed vertical lines separate the calibration and forecasting periods. Other dashed lines indicate the mean of the symptom score trajectories in the model calibration period. Solid lines and shaded regions indicate the mean and 95% CrIs of the forecasts, based on the posterior distributions of the parameters. The synthetic data were generated from the SD model using parameter values stated in Tables 5.1 and 5.2, with normally distributed noise added. DIC values are given in Table D.1.

5.3.4 Fitting the TS model to symptom score data to classify patients

Finally, we determine whether the estimated parameter values of the TS model can be used to classify patients as healthy or comorbid. As training data, we use the same synthetically generated data of symptom scores from 12 healthy and 12 comorbid patients over 10 days as to inform priors in Figures 5.5 and 5.7 (shown in Figure D.3). We assume that the classifications (healthy or comorbid) of these 24 patients are known. We then observe ‘test’ symptom score data for a new patient twice per day for up to 10 days. By fitting the TS model to the data for the new patient, and comparing the parameter estimates to those of the previous healthy and comorbid patients as detailed in Section 5.2.5, we predict whether the new patient is healthy or comorbid. As each new symptom score datapoint for the patient is observed, the classification can be updated by re-fitting the TS model.

In Figure 5.8 (a) we show simulations of symptom scores using the TS model for healthy and comorbid patients, using parameter values estimated by fitting the model to the 12 observed patients in each group. The large credible intervals are due to the differences in symptom scores between patients in the same group (see Figure D.3). Data for a new unclassified patient is shown (black dots), and the model fitted to the patient presented in green. Scatter plots comparing the healthy and comorbid parameter distributions are given in Figure 5.8 (b). In most cases, different regions of parameter space describe the two groups of patients, although there is some overlap due to patient variability (for instance a healthy patient may have a higher peak symptom score than a comorbid patient). Parameters relating to the new patient are shown in green, where in this case the new patient is classified as being healthy.

We consider 100 new patients (50 healthy and 50 comorbid), and make classifications every half day after having observed at least one day of data. In Figure 5.8 (c) we show the percentage of patients that were classified correctly. Early on, at day one or two, the chance of a correct classification is shown to be just only over 50% (random chance). After three days, approximately when the peak symptom score occurs on average, almost 90% of patients are classified correctly. Observing more data at later times does not significantly improve the classification accuracy, with no changes in classification given after five or more days of data. In total, 89 of 100 patients were classified correctly by using the parameters of the TS model, showing that the TS model can distinguish well between different groups of patients. We note that due to the variability in symptom score data in this synthetic study and the overlap in symptom score dynamics between groups, achieving a more accurate classification percentage would be challenging.

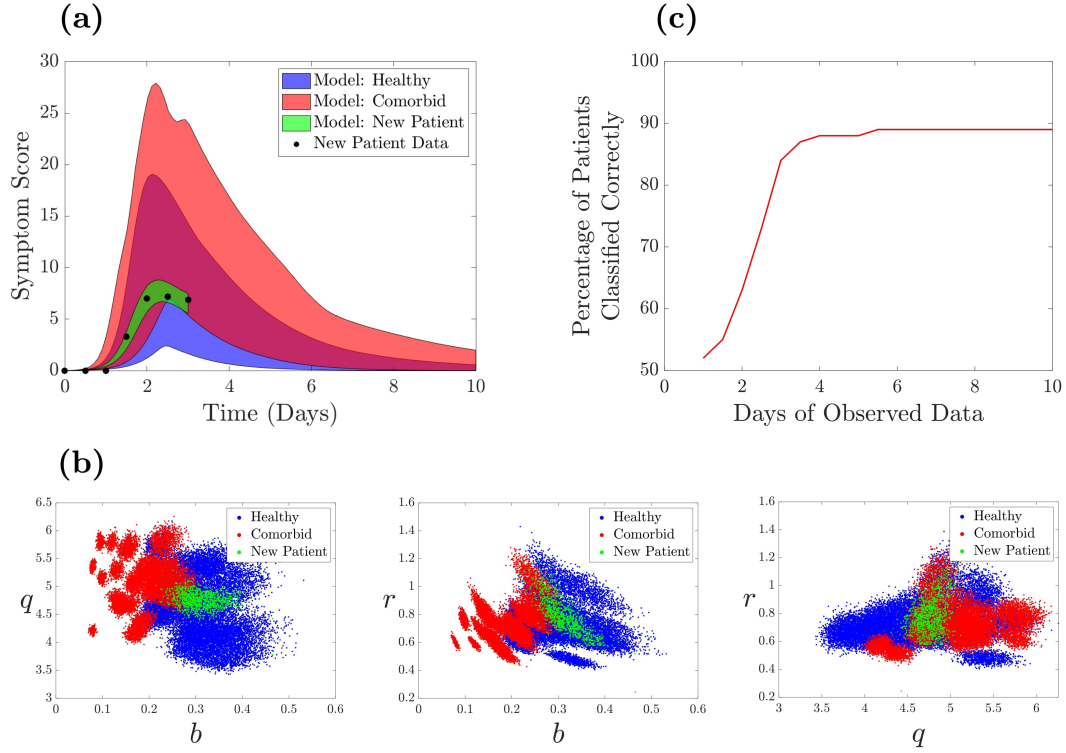


Figure 5.8: The TS model used to classify patients as healthy or comorbid. Prior symptom scores from 12 healthy and 12 comorbid patients are observed as training data, and the symptom score of a new patient is observed in real-time. The new patient is classified as healthy or comorbid by comparing the posterior distributions of the TS model parameters fitted to the data with the posterior distributions for the previously observed healthy and comorbid patients. (a): The 95% CrIs of the model trajectories of the combined posterior parameters of the previously observed healthy and comorbid patients, as well as the new patient data (black dots) and the 95% CrIs of the TS model trajectories fitted to the new patient data. (b): A scatter plot of the combined posterior distributions of the estimated parameters relating to the previously observed healthy and comorbid patients and the posterior distributions of the estimated parameters relating to the new patient. (c): The percentage of patients classified correctly using the TS model parameter distributions as the number of days of observed data are increased, for 100 different patients (50 healthy and 50 comorbid). The synthetic data relating to all patients in this figure were generated from the SD model using parameter values stated in Tables 5.1 and 5.2, with normally distributed noise added.

5.4 Discussion

Although making real-time forecasts is common when modelling the dynamics of an epidemic at the population level, less attention has focused on forecasting host level symptoms. Accurate forecasting of a patient's symptoms can improve patient outcomes by informing treatment, and may also impact control strategies during an epidemic, as symptom progression is related to transmissibility [291]. In this chapter, we have combined the forecasting techniques from Chapter 3 with the within-host models from Chapter 4 to address this gap. We have

proposed a new model, the SD model, by extending the IR model to include a symptom score variable.

By fitting this model to pooled human patient data of influenza infection viral load, the innate immune response (cytokine levels), and symptom scores, we were able to estimate values for the unknown fitting parameters β , p , ν , μ , η , and ω . We then used the SD model to generate synthetic symptom score data for two sets of patient groups, healthy individuals, and comorbid individuals who had higher peak symptom scores.

We adapted the TV model from Chapter 4 to formulate the TS model, in which symptom scores are assumed to be proportional to viral load, which, in turn, is assumed to be proportional to the number of infected cells. Thus, the TS model is analogous to the TV model, but with the number of cells nondimensionalised and the symptom score variable, S , replacing the viral load, V . We showed that the TS model can describe synthetic data from the SD model. Further, its parameters are identifiable and differ in values between healthy and comorbid patients. When using the TS model to generate forecasts in real-time without prior information, accurate forecasts can only be made after the symptom scores have reached their peak. When informed by previous patient data, forecasts are shown to be more accurate, even during the early stages of an infection.

Symptom score data are typically self-reported. In the first few days of an infection, individuals may not be aware that they are infected if their symptoms are mild, and thus, may not record them. As such, the time of infection may not be known. We showed that if the first datapoint is observed at some unknown time after the infection occurred, the original time of infection can be estimated by fitting the TS model to the observed data and simulating the model backwards in time. Further, real-time forecasts can be made, even when the time of infection is unknown. Finally, by fitting the TS model to observed patient data, we were able to correctly classify patients as healthy or comorbid based on the fitted values of their model parameters.

As for the IR model in Chapter 4, the innate and adaptive immune responses are incorporated into the SD model by reducing the rates of viral production and clearance respectively. Models in the literature have included additional terms relating to the killing of infected cells, by Natural Killer cells and macrophages which form part of the innate immune response, and T-cells which form part of the adaptive immune response [107, 172]. We have also assumed explicitly that an increase in symptom severity is due to the innate immune response, while there is evidence that symptoms may also be caused by the death of infected cells as well as virions, and thus, the innate immune response may not be the only mechanism to contribute to symptoms [327]. We have chosen to neglect these mechanisms in our model as values for the corresponding parameters are not well known, and we do not have sufficient data to determine them through model fitting. More experimental data, including studies of

immunocompetent and immunocompromised hosts [189], would enable us to formulate models which capture these additional mechanisms, identify parameter values and, thereby, assess the relative importance of each mechanism on the infection dynamics and symptoms.

We have considered a single measurable variable in this study, the symptom score. Symptom scores can be subdivided into systemic and respiratory symptoms. Following Canini et al. [57], in future work we could formulate a symptom dynamics model which outputs both symptom scores. Further model outputs such as temperature or nasal discharge weight, which have been used as measures of disease severity [120, 162], could also be incorporated into the symptom dynamics modelling framework described here. A forecasting model which extends the TS model to account for these quantities could then be formulated and used to investigate which data are most informative for making predictions, and to establish whether additional model complexity and additional data improve forecast accuracy.

One advantage of using output metrics such as self-reported symptom scores, is that collecting these data does not require medical professionals or resources that may only be available in hospitals or clinics. However, the disadvantage is that self-reported scores are inherently subjective, especially when reported outside of a clinical setting. What one individual may deem to be a severe symptom may be considered mild by another. Establishing clear distinctions for different symptom score levels where possible may help, although there will always be uncertainty in the data. Thus, having a large dataset is important. An example of a study in which time series data were collected on a large scale is the COVID Symptom Study, where over 4 million individuals in the UK and US used a mobile application to record their COVID-19 symptoms [248]. A model coupled to such data could enable individuals who record their symptoms to have updated real-time forecasts of their predicted disease evolution, and potentially be alerted if high symptom scores are predicted in future based on their initial data. However, for this to be clinically useful, the relationship between self-reported symptoms and disease severity would need be established to ensure that symptom thresholds are aligned with specific states of disease severity. A model could also be used to guide treatment. For this to be effective though, accurate forecasts may need to be attainable during early stages of the infection [205].

While non-complicated influenza infections are short-lived (on the order of one week), the methodology used in this chapter could be extended to other viral infections which occur over longer time-scales such as Hepatitis C [310]. For longer-term infections, forecasts may be more useful as they could be made many days or weeks before peak symptom severity. As well as forecasting the entire symptom score progression, the TS model could also be used to predict scalar quantities such as the duration of infectiousness and peak symptom scores. If sufficient data were available on previous patients, this could be used to train models and improve predictions for subsequent individuals and to estimate the likelihood of hospitalisations, severe disease, or mortality.

Medical data are often noisy, partially observed, and sampled infrequently or at uneven intervals. Future work could explore the accuracy of TS model forecasts where measurement noise levels, sampling intervals, and the numbers of prior patient symptom scores are varied. The findings of such a study could also be validated with real-world data to establish whether the TS model can make accurate forecasts with more irregular data. As the methods in this study may be widely applicable to a range of scenarios, this does not necessarily have to be done using influenza data or even symptom scores, but some measure of disease severity for any relevant viral disease such as COVID-19 [108].

We have used an ordinary differential equation model to make forecasts. Supervised and unsupervised machine learning progression models could also be used [257, 262, 354, 381]. It would be interesting to compare machine learning forecasts to ODE model forecasts. An advantage of an ODE model is that prior parameter distributions to inform forecasts can be estimated from the symptom scores of a single patient, while machine learning models typically require a large amount of training data. As noted in Chapter 4, a further advantage of using a mechanistic model such as the TS model is that it is based on an infection process, and hence, its parameters are interpretable. One of its disadvantages though is that it can only describe symptoms which increase to a single peak and then decay. While this describes most primary influenza infections, as discussed in Chapter 4, more complex models would be needed to describe dynamics which resurge.

We considered two groups of patients, healthy and comorbid, and accounted for intra-group variability by assuming that model parameters from each group are drawn from a distribution. This work could be extended to consider multiple patient groups, for instance groups with specific comorbidities (e.g. COPD or diabetes) where individuals in the different comorbid groups exhibit different dynamics. As we have demonstrated using synthetic data, the TS model has the potential to be used to classify patients based on the distributions of its parameters. Often, especially in developing countries, comorbidities are not diagnosed, so individuals may be unaware that they have underlying health conditions [340]. If reliable classification is possible, by discovering that a patient's symptoms align with the symptoms of those with comorbidities, they could be referred for further testing.

In this chapter, we have used an ODE model to make real-time forecasts of influenza symptom scores. We generated synthetic data from a symptom dynamics model which explicitly describes the innate and adaptive immune responses, and then showed that a simpler two-variable model can accurately generate real-time forecasts of symptom scores and classify patients into different groups. Although there is significant scope for future work, in particular validating the TS model with real-world data to determine its utility in a clinical setting, we have shown that the model has the potential to make predictions about a patient's symptom progression.

Chapter 6

Discussion

6.1 Introduction

Infectious diseases are a leading cause of mortality worldwide [352]. The COVID-19 pandemic has, thus far, claimed the lives of over three million people [390], and influenza epidemics have caused millions of deaths around the world and the disease remains a global burden [389]. Increased understanding of infectious disease dynamics is important for developing effective strategies to fight and control these diseases. Mathematical modelling is a valuable tool which can help in this quest.

In this thesis, we have considered mathematical models at the population and host scales, focusing on influenza viruses and SARS-CoV-2. In each of Chapters 2–5, we have sought to address a specific epidemiological modelling question. As such, our work is relevant to modellers and public health officials concerned with understanding and controlling epidemics at the population scale (Chapters 2 and 3), and experimentalists and medical professionals concerned with understanding and treating infection dynamics at a patient level (Chapters 4 and 5). Although each chapter represents a self-contained piece of work, they address common themes in terms of investigating which models should be used to make projections and assessing how more accurate forecasts can be made in heterogeneous populations.

6.2 Summary of results

In Chapter 2, we sought to assess the risk of an outbreak occurring in populations characterised by low case prevalence, no non-pharmaceutical interventions, and ongoing vaccination programmes. Using SARS-CoV-2 in the Isle of Man and Israel as case studies, we used four metrics to quantify the risk of an outbreak, under the assumption that infections can be described via a branching process model in the initial stages of a possible outbreak. We calculated the temporal changes in the outbreak risk as vaccinations take place and illustrated how vaccinations can lead to a significant drop in the potential for virus transmission. We showed that, if the Instantaneous Outbreak Risk metric is used (the IOR considers the reproduction number at a single point in time), then the risk of an outbreak is overestimated, as future

vaccinations are neglected. By contrast, the Numerical Outbreak Risk calculates the risk in a population with varying infection susceptibility due to ongoing vaccinations. The NOR can be computed without having to perform multiple simulations of the branching process model. In the two geographical regions we considered, we found that even when vaccine rollouts are complete, the risk of an imported case causing an outbreak will not be entirely eliminated. Therefore, to reduce the risk of local outbreaks, it may be necessary to continue monitoring inbound passengers for infection. If infections are detected, contact tracing and isolation strategies can then be used to prevent substantial local transmission.

In Chapter 3, we investigated the importance of accounting for cross-immunity when forecasting influenza epidemics. We considered two mathematical models that describe influenza transmission; a 2-group model which distinguishes immunologically naive and cross-immune individuals, and a 1-group model which assumes a homogeneous population. We parameterised each model using case data from the 2009 H1N1 influenza epidemic in Japan, and then used the models to predict the epidemic dynamics caused by an antigenically related strain occurring at a future date. If no prior knowledge about the values of the unknown model parameters is assumed, then predictions of the future dynamics of both models are uncertain. If we neglect cross-immunity (and use the 1-group model), then the epidemic dynamics are identical to those from 2009, regardless of when the future epidemic occurs.

We then used knowledge of parameters from the 2009 epidemic to inform our forecasts which were also updated using synthetic real-time case notification data. We found that the 1-group model only produces accurate forecasts once the peak of the epidemic has passed. By contrast, the 2-group model generates accurate forecasts, even early during the outbreak. We conclude that cross-immunity should be included in models used to forecast the long-term dynamics of influenza epidemics, and that data from previous related epidemics caused by antigenically related strains should also be taken into account.

In Chapter 4, we modelled the within-host dynamics of influenza. We considered three mathematical models: the IR model which explicitly describes the host immune responses, and the target-cell limited TEIV and TV models, which do not. We showed, via structural and practical identifiability analyses, that parameter identification is possible. We then fitted the IR model to experimental data from immunocompetent and immunosuppressed mice. We also generated synthetic viral load data using the IR model, and showed that the TV model can reproduce these data and, furthermore, that the TV model should be preferred over the TEIV model since it has fewer parameters and equivalent model flexibility. This study demonstrates that the TV model can be used to describe the viral load dynamics in (hypothetical) patient groups that differ in the strength of their immune response. Our analysis also reveals a relationship between the parameters of the TV and IR models, when the TV model is fitted to synthetic data generated from the IR model.

Finally in Chapter 5, we extended the IR model from Chapter 4 to include a symptom score variable to describe the severity of symptoms in patients, thus formulating the SD model. We fitted the SD model to pooled experimental time series data describing the viral load, innate immune response, and symptom scores in influenza-infected patients. We then generated synthetic symptom score data for two patient groups; healthy individuals, and comorbid individuals, the latter characterised by higher peak symptom scores. We adapted the TV model from Chapter 4, to formulate the two-variable TS model, in which it is implied that symptom severity is proportional to viral load. We showed that the TS model can reproduce the synthetically generated data from the SD model, that its parameters are identifiable, and that they differ in value between healthy and comorbid patients. We conclude that the TS model may be able to distinguish between patient groups.

When the TS model is informed by previous patient data, it accurately forecasts synthetic symptom scores during early stages of the infection. In practice, however, the initial time of infection may not be known. We showed that, in such situations, the original time of infection can be estimated by fitting the TS model to the observed data and simulating backwards in time. Further, real-time forecasts can be made when the time of infection is unknown. Finally, by fitting the TS model to observed patient data, we were able to classify patients as healthy or comorbid based on the estimated model parameter values. The work in Chapters 4 and 5 also provided insight into within-host disease forecasting for our industrial partner, and the models developed may be built upon for wider clinical use.

6.3 Model complexity and incorporating immunity

In Chapter 2, we investigated which metrics should be used when calculating the risk of an outbreak given imported cases, and in Chapters 3–5, we investigated which models should be used for forecasting epidemic and within-host dynamics respectively. Thus, a key theme in this thesis is investigating which approach should be used to make the most accurate possible projections of future dynamics. A variety of model types can be used for infectious disease forecasting, including ordinary differential equations [329], stochastic branching processes [10], agent-based models [174], and machine learning approaches [405]. The level of epidemiological detail included in a model varies greatly depending on the available data and the problems being addressed. Early in an outbreak caused by a novel pathogen, lack of data may mean that simple models which assume population homogeneity may be preferred when making projections [191], while more detailed models may be appropriate at later times, as more epidemiological parameter values are quantified [16]. Many different modelling approaches have been used to forecast the COVID-19 pandemic, and an important yet difficult question is to determine which approach is best suited for a specific problem as different models may generate different projections [324]. During the early stages of an outbreak, it may be challenging to determine which model forecasts are accurate [74]. To address this, ‘ensemble’ models have

been proposed which combine forecasts from multiple mathematical and statistical models into a single forecast to improve robustness [317, 392].

An overarching theme of this thesis was to investigate whether it is necessary to explicitly incorporate immunity into mathematical models of viral infectious diseases. In practice, population level immunity increases following vaccination. In Chapter 2, we demonstrated the importance of accounting for the future increased immunity due to vaccinations to accurately predict the likelihood of an outbreak occurring. In a similar manner, in Chapter 3, we considered a scenario where population level immunity wanes over time due to births and deaths, and showed that accounting for changes in immunity increases the accuracy of forecasts of the dynamics of a future epidemic. In both population level studies, accounting for immunity was important for generating accurate projections.

By contrast, in Chapters 4 and 5, we showed that in the context of within-host modelling, the simpler TV and TS models can reproduce viral load and symptom score data generated by the more complex IR and SD models. Hence, the immune response does not need to be included in within-host models to reproduce these quantities in these scenarios when the viral load or symptom score grow to a single peak and then decay. The parameter values of the simple models can be varied to represent hosts which have differing immune responses.

However, in the within-host modelling studies, we only considered partially observed data, with a single model output when we fitted the TV and TS models to the synthetic data. In future work, it would be interesting to consider situations in which more than one model output is observed (for example, viral load and symptom scores, or alternatively, the fraction of dead cells, viral load, and symptom scores). It might then be necessary to incorporate different aspects of the immune response. Clearly, the level of detail included in a model should relate to the available data and the problem of interest.

In Chapter 3, we assumed that individuals previously infected with a similar influenza strain only suffered from mild disease, and that naive individuals suffered from severe disease. In Chapter 2, we assumed that being vaccinated reduced susceptibility to SARS-CoV-2. Other studies have indicated that immunity, through prior infection or vaccination, may also reduce infectiousness [157]. As of June 2021, ongoing work is investigating the effects of immunity and vaccinations on the spread of SARS-CoV-2 and outcomes of COVID-19 [48, 84]. Once this is understood, it could inform future work on predicting the likelihood of outbreaks in vaccinated or previously infected populations.

With a number of circulating variants, multiple prior waves, and different vaccines, there is a heterogeneous immunological landscape within the global population. An important question is what level of granularity is needed in mathematical models to accurately forecast the future dynamics of the COVID-19 pandemic. For instance, is it necessary to account for different vaccines with different efficacies, different variants with different transmission rates, and waning immunity from different waves in a mathematical model to make an accurate

projection? Or will assuming ‘average’ values and a homogeneous population suffice in certain scenarios? By comparing forecasts generated from models of different complexities and using synthetic data, as we did in Chapter 3, it should be possible to shed light on the problem of model complexity in the context of the COVID-19 pandemic.

6.4 Areas for future research

In this thesis, we have investigated a range of problems at the population and within-host scales. However, the mathematical methods and models used at each scale are similar. For example, in Chapter 3, we made real-time forecasts, and showed that, without prior data from a previous epidemic, forecasts were uncertain. However, with prior data, we were able to make accurate projections. Similarly, in Chapter 5, we showed that making accurate real-time forecasts of symptom dynamics requires previous patient data.

Future work could combine within-host and population level forecasts using multi-scale models [68, 135, 266]. At a population scale, we track the number of cases, hospitalisations, and deaths to inform control measures and to assess the spread of the pathogen. However, these predictions require knowledge of the infection at a host level; the length of the latent period, the duration of infection, and the probability of severe disease. A better understanding of within-host dynamics could improve forecasts and enable more informed decisions to be made at the population level [158]. Further, observing more patients at a host level increases the knowledge of the distributions of population level parameters. Additionally, model structure and complexity can be informed by observing patient level data as the appropriate heterogeneities can then be included in population level models.

By collecting data at a within-host level we can build more sophisticated epidemic models, in which, for instance, the infectiousness profiles are known accurately for groups of hosts with different epidemiological characteristics. Within-host models could be parameterised using patient level data and coarse-grained into a population model to make more accurate forecasts [275]. An open question is whether full multi-scale models which explicitly describe the epidemiological mechanisms at a host level are needed, or whether simply understanding within-host dynamics to obtain parameter values for population scale models is sufficient. By varying the parameters related to the immune response in a within-host model, we could investigate how this affects symptom scores, the duration of infection, and infectiousness, and feed this information into a population scale model. Thus, being able to forecast within-host dynamics more accurately could lead to improved forecasts of population level dynamics.

Surveillance is key early in an outbreak, for making informed predictions and optimising interventions [365]. Although it is impractical to measure viral load or aspects of the immune response, such as cytokine and antibody levels in patients at a large scale (due to the resources required), the majority of patients can track their own symptoms. This has been done for COVID-19, with millions of individuals recording their symptoms using a mobile app [248].

If patients track their symptoms, then variability in disease progression between and within groups of patients sharing epidemiological characteristics can be assessed. As stated previously, this will also aid population modelling efforts, as it can inform the level of heterogeneity that should be included in models and facilitate parameter estimation.

The evolution of an epidemic in a population and the dynamics of viral disease progression within a host are intrinsically complex processes. Models represent these processes by classifying individuals and cells respectively into different states, with the transition between states determined by epidemiological and immunological parameters. Models require a multitude of parameter estimates which are each subject to uncertainty [401]. As we have done throughout this thesis, we have examined uncertainty by producing multiple realisations of model simulations under different parameter sets, showing credible intervals for forecasts. However, for tractability, we have made assumptions about certain parameter values which could be relaxed to explore uncertainty in forecasts further. For example, future work could consider uncertainty in the latent and infectious periods of infected individuals in Chapter 3, and analogously the latent and infectious periods of infected cells in Chapters 4 and 5.

We could extend the work in this thesis to model the infection dynamics of different strains of influenza [293] and SARS-CoV-2 [218]. At a host level, if sufficient data were available, through model fitting and parameter estimation, we could elucidate the epidemiological mechanisms that cause different infection kinetics for different strains, for instance, determining if the viral production rate is greater for one strain compared to another. Moreover, we could investigate whether the simple TV and TS models can reproduce the dynamics of hosts infected with different strains, as was done for different groups of hosts with different epidemiological characteristics. In Chapter 2, we considered two different strains of SARS-CoV-2 with different levels of transmissibility. If newer strains were to emerge, then we could use the same methodology to assess the risk of an outbreak caused by these novel strains.

Future work could build upon the models used in this thesis and identify whether it is necessary to incorporate age-structure, a common heterogeneity considered in infectious disease modelling [73, 364]. In Chapter 3, we modelled a future epidemic caused by a hypothetical strain of influenza antigenically related to the 2009 H1N1 variant. We could also consider epidemics caused by different subtypes and investigate whether an age-structured model is required to make accurate forecasts for each subtype. Models which distinguish between groups of individuals based on their age may be required because different subtypes have been the dominant circulating strain at different periods of time over the past century; H1N1 1918–1957, H2N2 1957–1968, H3N2 1968–2015 [139]. Hence, different generations of people have varying levels of protection against these subtypes as an individual’s first influenza infection confers lifelong protection against subtypes of the same phylogenetic group. Similarly, we could incorporate age-structure into the branching process infection model used in Chapter 2 as there is evidence that susceptibility to infection depends on age [84]. This ties in with

the work in Chapter 4, where we considered a group of elderly patients who exhibited different infection viral dynamics and assumed that their immune responses differed from those of younger patients.

The models used in this thesis are tied to real-world data. The IR and SD models in Chapters 4 and 5 explicitly modelled the kinetics of cytokines and antibodies. However, there are many other aspects of the immune system that could be considered, including cellular components such as Natural Killer cells and T-cells. Host level modelling is currently limited by the lack of data. By working together, experimentalists and modellers have the potential to develop a better understanding of how the immune system responds to viral infection, and heterogeneity in the immune response between different hosts [343]. In Chapter 3, at a population level, the data we considered were of the number of individuals who suffered from severe disease. It would be interesting to investigate whether the conclusions of this chapter still hold if we also forecast the total number of cases or only mild cases [38]. This would require additional data collected through heightened surveillance and reporting. We could also incorporate more information about the specific vaccines administered to different individuals into the methodology in Chapter 2. This may also tie in with using an age-structured model, as in some countries the type of vaccine administered is age-dependent [236].

Whilst in each of Chapters 2–5 we focused on either SARS-CoV-2 or the influenza virus, our results and methodologies may be applicable to both pathogens. In Chapter 2, we investigated the spread of SARS-CoV-2 in Israel and the Isle of Man, two regions with high vaccination rates. We showed that importation of a case may still result in an outbreak, especially for more transmissible variants. So, even if all eligible individuals in a population are vaccinated, localised outbreaks may still occur. It is likely that SARS-CoV-2 will continue to spread in many regions of the world for the foreseeable future [206], with new strains constantly emerging, and large mutations in the virus potentially resulting in strains which may cause new localised epidemics [54]. In that case, the work in Chapter 3 may be relevant to modelling SARS-CoV-2, as it could be necessary to explicitly account for cross-immunity between coronavirus strains when forecasting epidemics. Moreover, although in Chapters 4 and 5 we have considered the host level dynamics of influenza, we could adapt these models to describe the dynamics of SARS-CoV-2 infection [205]. We could investigate whether simple models can describe SARS-CoV-2 infection dynamics in patients with different characteristics and furthermore in hosts infected with different strains of the virus.

6.5 Concluding remarks

Mathematical modelling has a key role to play in understanding and combatting infectious disease epidemics. This has been highlighted in the past 18 months during the COVID-19 pandemic. Although the pandemic is starting to be suppressed in many regions of the world, it will continue to cause disease burden, and its secondary effects will be felt for years to

come. Future outbreaks may occur, even in regions where SARS-CoV-2 has been controlled, and COVID-19 may become endemic, in a similar manner to influenza [370].

Being able to predict epidemic dynamics enables public health officials to prepare interventions and reduce disease burden. Similarly, being able to predict the disease progression of patients already infected at a host level can result in more effective treatment and care. Thus, forecasting at the population and within-host levels go hand-in-hand in reducing morbidity and mortality. The work in this thesis has provided insights into the level of model complexity needed to describe infectious disease dynamics and make more accurate projections. We hope that the approaches described here will be developed further and applied during a range of future epidemics.

Appendix A

Appendix to Chapter 2

A.1 The effect of the assumed vaccine uptake on the outbreak risk

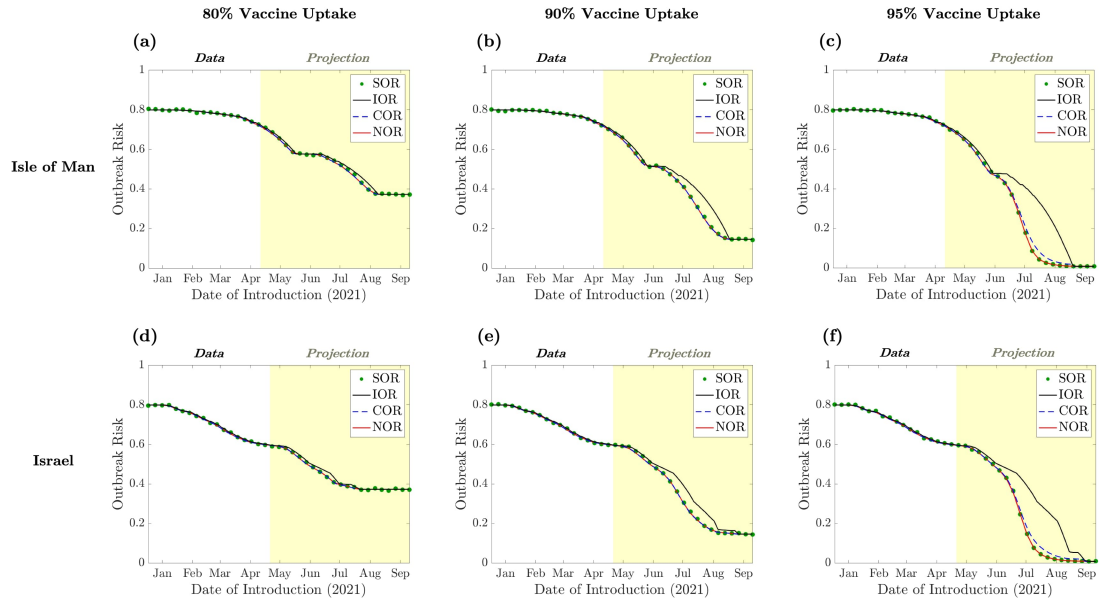


Figure A.1: The effect of the assumed vaccine uptake on the outbreak risk. (a): Isle of Man - A reproduction of Figure 2.3 (b) in the main text, with vaccine uptake $\nu = 0.8$. (b): Analogous panel to Figure 2.3 (b) in the main text, but with vaccine uptake $\nu = 0.9$. (c): Analogous panel to Figure 2.3 (b) in the main text, but with vaccine uptake $\nu = 0.95$. (d): Israel - Analogous panel to Figure 2.3 (d) in the main text, but with vaccine uptake $\nu = 0.8$. (e): Analogous panel to Figure 2.3 (d) in the main text, but with vaccine uptake $\nu = 0.9$. (f): Analogous panel to Figure 2.3 (d) in the main text, but with vaccine uptake $\nu = 0.95$. Ticks on the x-axes refer to the starts of the months labelled.

A.2 The effect of immunity due to prior infections on the outbreak risk

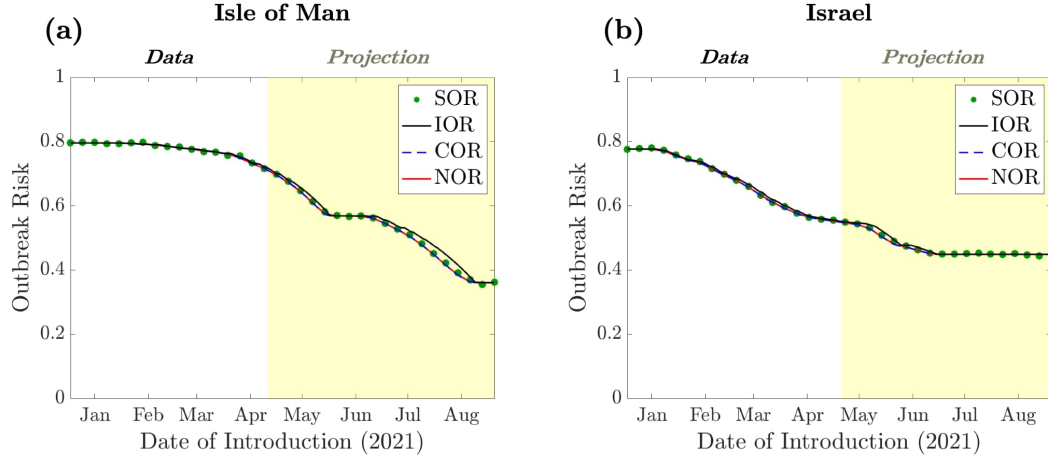


Figure A.2: The effect of immunity due to prior infections on the outbreak risk. (a): Isle of Man - Analogous panel to Figure 2.3 (b) in the main text, but with R_0 reduced by 1.4% (an approximation of the proportion of the population that had been infected before 1st May 2021, based on 1,154 confirmed cases out of a population of $N = 84,500$). (b): Israel - Analogous panel to Figure 2.3 (d) in the main text, but with R_0 reduced by 9.6% (an approximation of the proportion of the population that had been infected before 1st May 2021, based on 838,000 confirmed cases out of a population of $N = 8,772,800$). Ticks on the x-axes refer to the starts of the months labelled.

A.3 The effect of the assumed effectiveness of two vaccine doses on the outbreak risk

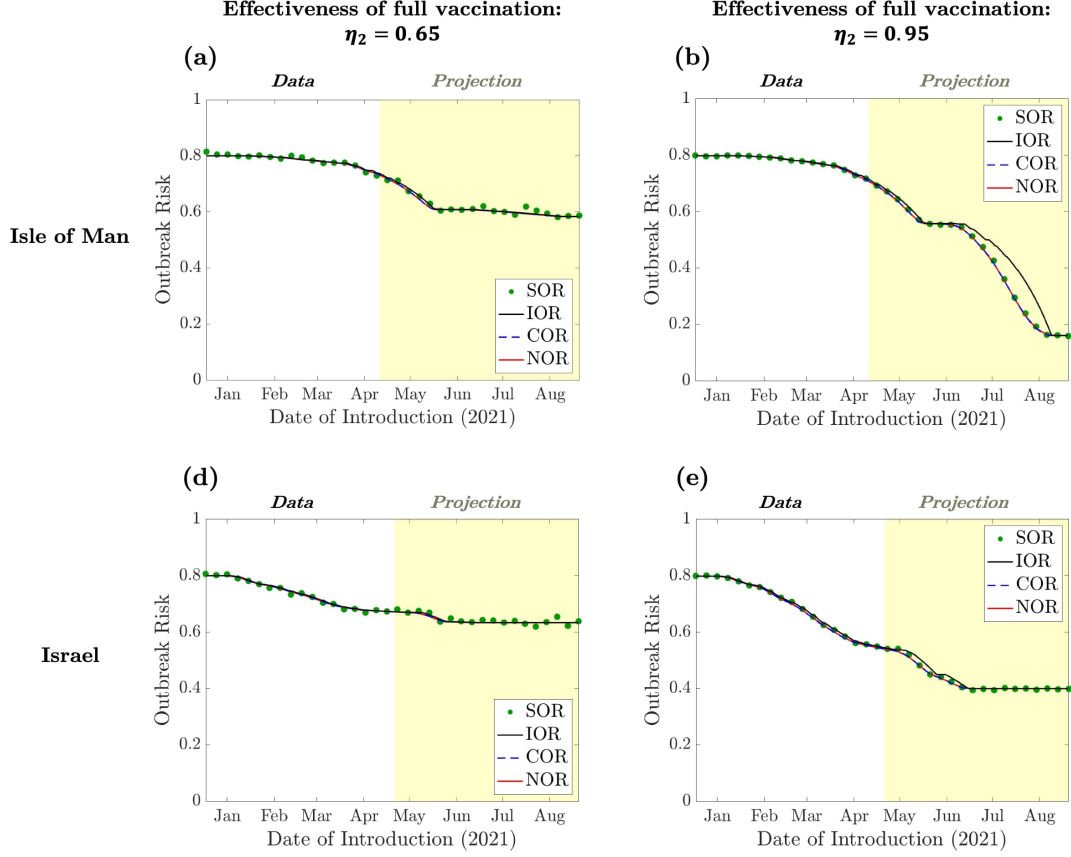


Figure A.3: The effect of the assumed effectiveness of two vaccine doses on the outbreak risk. (a): Isle of Man - Analogous panel to Figure 2.3 (b) in the main text, but with $\eta_2 = 0.65$. (b): Analogous panel to Figure 2.3 (b) in the main text, but with $\eta_2 = 0.95$. (c): Israel - Analogous panel to Figure 2.3 (d) in the main text, but with $\eta_2 = 0.65$. (d): Analogous panel to Figure 2.3 (d) in the main text, but with $\eta_2 = 0.95$. Ticks on the x-axes refer to the starts of the months labelled.

A.4 Sensitivity analysis of the outbreak risk to model parameters

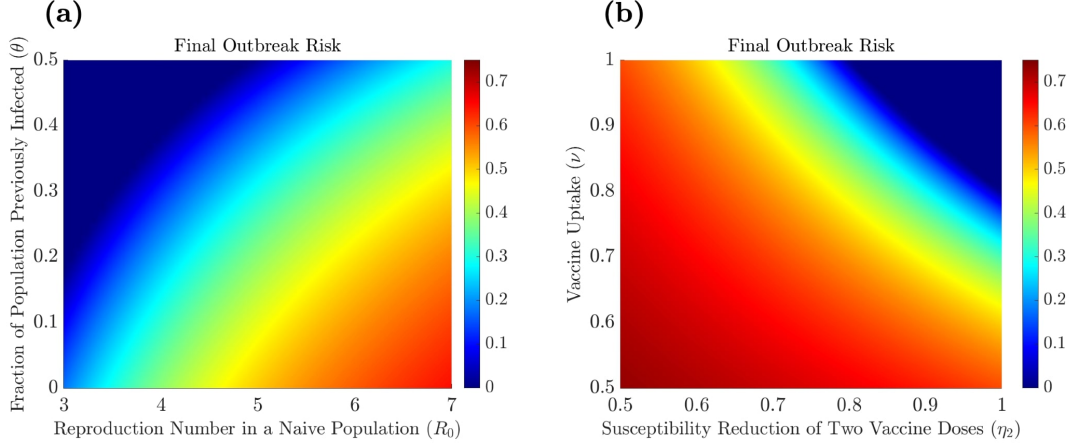


Figure A.4: Sensitivity analysis of the outbreak risk to model parameters. The final outbreak risk after the vaccination rollout has been completed in Israel as (a): the reproduction number in a fully naive population, R_0 , and the fraction of population previously infected, θ , are varied, and (b): the susceptibility reduction of two vaccine doses, η_2 , and the vaccine uptake, ν , are varied. Fixed parameter values are $\eta_2 = 0.85$ and $\nu = 0.7$ in (a), and $R_0 = 5$ and $\theta = 0$ in (b).

A.5 Individual level variation in SARS-CoV-2 transmission

Under the simple model of SARS-CoV-2 transmission considered in the main text, infected individuals are infectious for an exponentially distributed period of time (with mean $1/\mu$ days) and generate new infections at rate $\beta(1 - \Lambda(t))$ per day. Using this model, we found that the risk of outbreaks is likely to be greater than zero in low prevalence settings when NPIs are removed, even following vaccination programmes in which significant numbers of individuals are fully vaccinated. Here, we test the robustness of this conclusion to the assumed level of heterogeneity in transmission between different infected individuals.

In the analysis in the main text, when vaccination programmes are completed, the offspring distribution (the number of infections generated by each infected individual) follows a geometric distribution with mean $R_V(t) = \frac{\beta(1-\eta_2\nu)}{\mu}$ (we denote this final value of $R_V(t) = \frac{\beta(1-\eta_2\nu)}{\mu}$ by R_V for the remainder of this section). Since virus transmissibility is then constant in future, the IOR is appropriate for assessing the risk of outbreaks.

In reality, however, there is significant variation in the number of infections generated by different infected individuals. The potential that superspreading events occur leads to a longer tailed offspring distribution than that of a geometric distribution. To allow for this in epidemiological models, a negative binomial distribution can be used instead of a geometric distribution to characterise the offspring distribution.

Here, we explore the outbreak risk at the end of the vaccination programmes in the Isle of Man and Israel using a negative binomial offspring distribution. Specifically, the probability that an infected individual generates x secondary cases is given by

$$P(X = x) = \frac{\Gamma(k + x)}{x! \Gamma(k)} \left(\frac{R_V}{R_V + k} \right)^x \left(1 + \frac{R_V}{k} \right)^{-k}, \quad (\text{A.1})$$

where k is the dispersion parameter [228]. When $k = 1$, this is equivalent to the geometric distribution assumed in the main text, but when $k < 1$ then there is more individual level variation in transmission (i.e. a higher potential for superspreading).

Under this more complex transmission model, the probability that an infected individual, introduced into the population at any time following the vaccination programme, initiates an outbreak driven by sustained local transmission, is given by

$$\text{IOR} = 1 - q, \quad (\text{A.2})$$

where q is the minimal non-negative solution of

$$q = \left(1 + \frac{R_V(1 - q)}{k} \right)^{-k}, \quad (\text{A.3})$$

as described by Lloyd-Smith et al. [228] and Nishiura et al. [279]. As noted above, when the dispersion parameter $k = 1$, then the offspring distribution is geometric, in which case the IOR is given by the expression in equation (2.6) (with $R_V(t) = R_V$, since the vaccination programme is assumed to be completed).

We estimated the IOR at the ends of vaccination campaigns in the Isle of Man and Israel for different values of the dispersion parameter, k (Figure A.5). As in the main text, we considered two scenarios corresponding to different mean R_0 values (these values are $R_0 = 3$ and $R_0 = 5$). Blue dots in Figure A.5 correspond to $k = 1$ (so that the IOR matches the IOR values at the ends of the vaccination programmes in the corresponding panels of Figure 2.3 in the main text), and red dots in Figure A.5 correspond to $k = 0.1$ (which is line with published estimates of k for SARS-CoV-2 transmission [105]).

In each scenario, the IOR once vaccination programmes are completed is greater than zero for realistic values of the dispersion parameter, k . The IOR is particularly large in the scenarios in which the mean value of $R_0 = 5$. We expect these scenarios to reflect the current and future risks most accurately, due to the emergence of variants that are more transmissible than the original SARS-CoV-2 virus. The results in Figure A.5 support our main conclusion that the risk of outbreaks in low prevalence settings (in the absence of NPIs) is unlikely to be eliminated by vaccination. Surveillance strategies that aim to find infected inbound travellers in locations with low infection prevalence are of clear importance, even once vaccination programmes are completed.

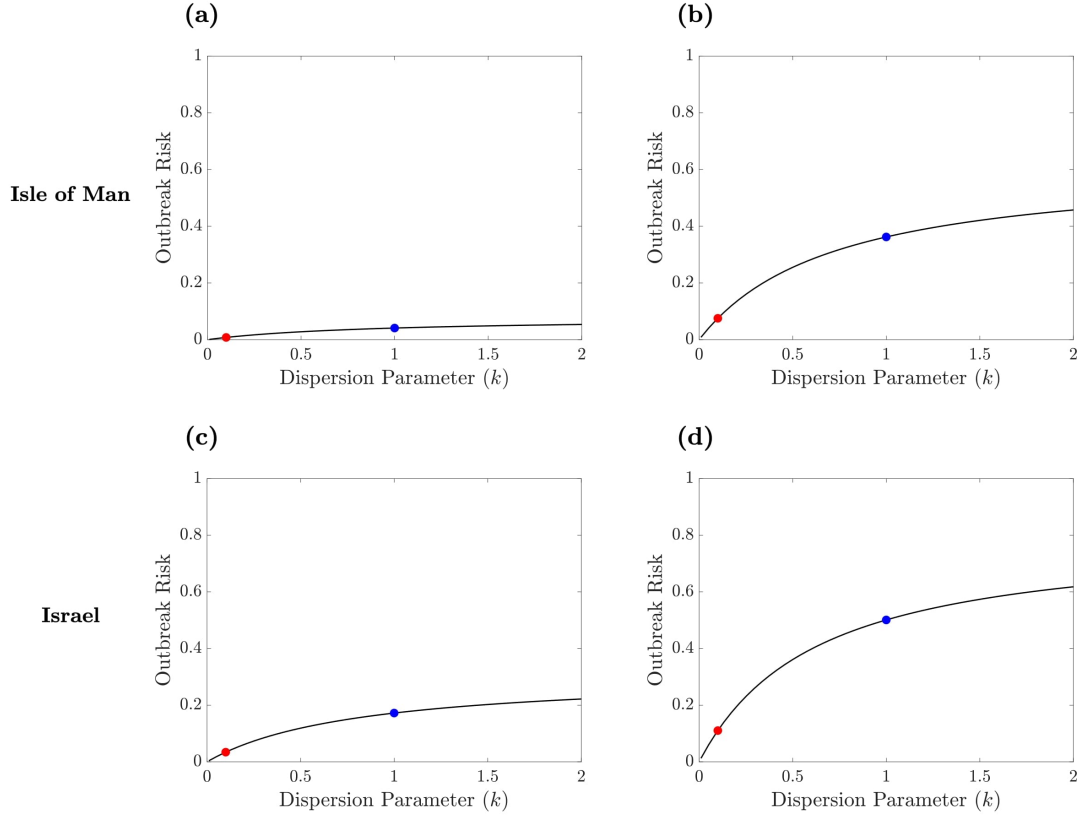


Figure A.5: The effect of individual level variation in transmission between different infected individuals (determined by the dispersion parameter, k , of the negative binomial offspring distribution) on the IOR following the completion of the vaccination programmes. (a): Isle of Man, with mean $R_0 = 3$ at the beginning of the vaccination programme (as in Figure 2.3 (a) in the main text). (b): Isle of Man, with mean $R_0 = 5$ at the beginning of the vaccination programme (as in Figure 2.3 (b) in the main text). (c): Israel, with mean $R_0 = 3$ at the beginning of the vaccination programme (as in Figure 2.3 (c) in the main text). (d): Israel, with mean $R_0 = 5$ at the beginning of the vaccination programme (as in Figure 2.3 (d) in the main text). Blue dots correspond to $k = 1$ (the underlying assumption of the simple branching process model used in the main text) and red dots correspond to $k = 0.1$ (as estimated for SARS-CoV-2 transmission in [105], although larger values of k have been estimated in some studies [3, 5]). The values of other parameters are given in Table 2.1 in the main text.

Appendix B

Appendix to Chapter 3

B.1 Sensitivity analysis: latent and infectious periods

In the 1-group and 2-group models, the parameter values relating to the latent and infectious periods are taken from the literature [112, 373]. However, these are only estimates, and therefore it is instructive to carry out a sensitivity analysis by simulating the models with different parameter values and seeing the effects on the epidemic dynamics. We show results of this sensitivity analysis in Figure B.1.

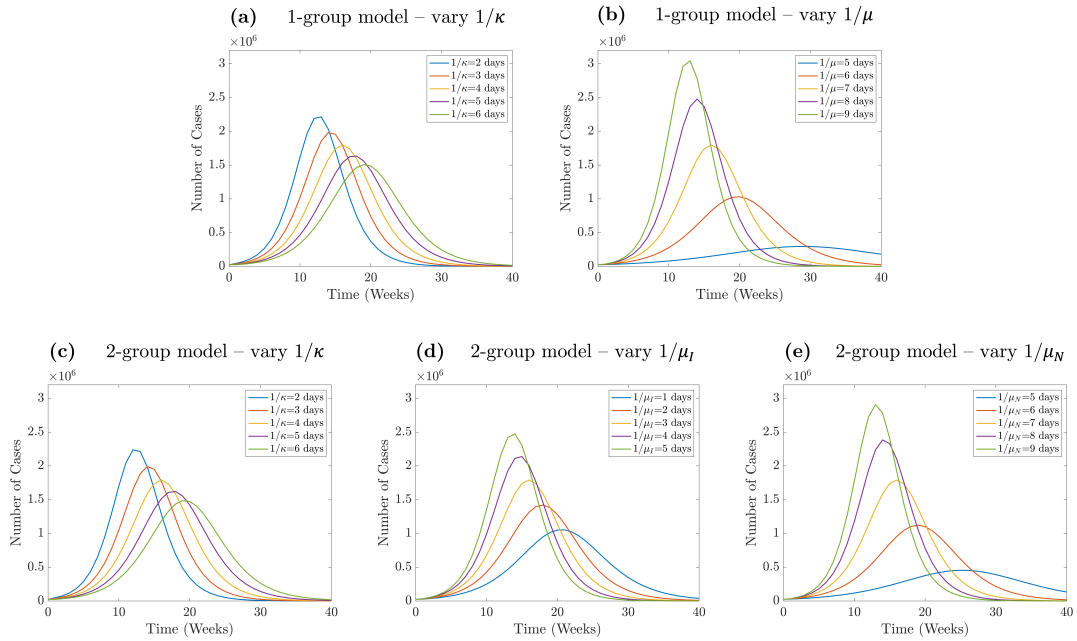


Figure B.1: Sensitivity analysis of the latent and infectious period parameters in the 1-group and 2-group models. (a)-(b): Simulations of the 1-group model with varied κ and μ . (c)-(e): Simulations of the 2-group model with varied κ , μ_I , and μ_N . In each case, all other parameters values remain fixed and are given in Table 3.1.

B.2 Sensitivity analysis: the cross-immune fraction

Based on data describing the proportion of individuals who sought medical attention, we assumed that the cross-immune fraction of the population during the 2009 epidemic was $\nu = 0.3149$. To verify that the conclusions obtained in this study are robust to the assumed value of ν , here we show the primary results of the study with $\nu = 0.5 \times 0.3149 = 0.1575$ and $\nu = 2 \times 0.3149 = 0.6298$. Using these different values of ν , we estimate different values of β and N when fitting the 2-group model to data from Japan from the 2009 epidemic, and hence the future epidemic 25 years later is predicted to vary in size. Forecasts made in week 10 of the future epidemic with no prior information, in advance at week 0, and in week 10 with prior information from the 2009 epidemic, in these two different scenarios, are presented in Figures B.2 and B.3. In these cases, accounting for cross-immunity by using the 2-group model still results in more accurate forecasts compared to when the 1-group model is used, and thus the qualitative conclusions of our study are robust to the assumed value of ν .

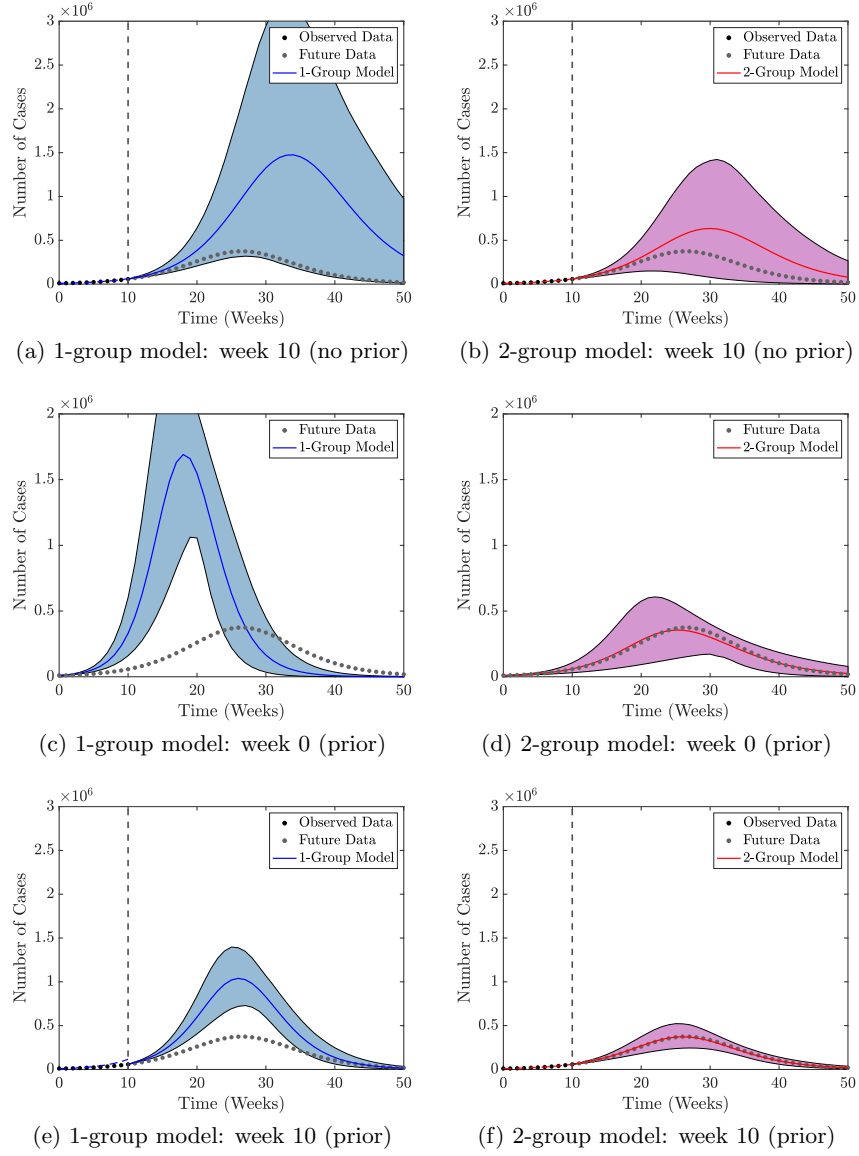


Figure B.2: The primary results of the study, repeated with half the value of the cross-immune fraction ν at the beginning of the 2009 epidemic compared to the value assumed in the main text ($\nu = 0.5 \times 0.3149 = 0.1575$). Forecasts obtained using the 1-group and 2-group models of a future epidemic occurring 25 years after the 2009 epidemic. Forecasts are generated in three scenarios. (a)-(b): Forecasts generated in week 10 of the future epidemic, with no prior information about the fitting parameters. (c)-(d): Forecasts generated at the start of the future epidemic, with parameters based on the model fits to data from the 2009 epidemic. (e)-(f): Forecasts generated in week 10 of the future epidemic, with priors of the fitting parameters based on the model fits to data from the 2009 epidemic.

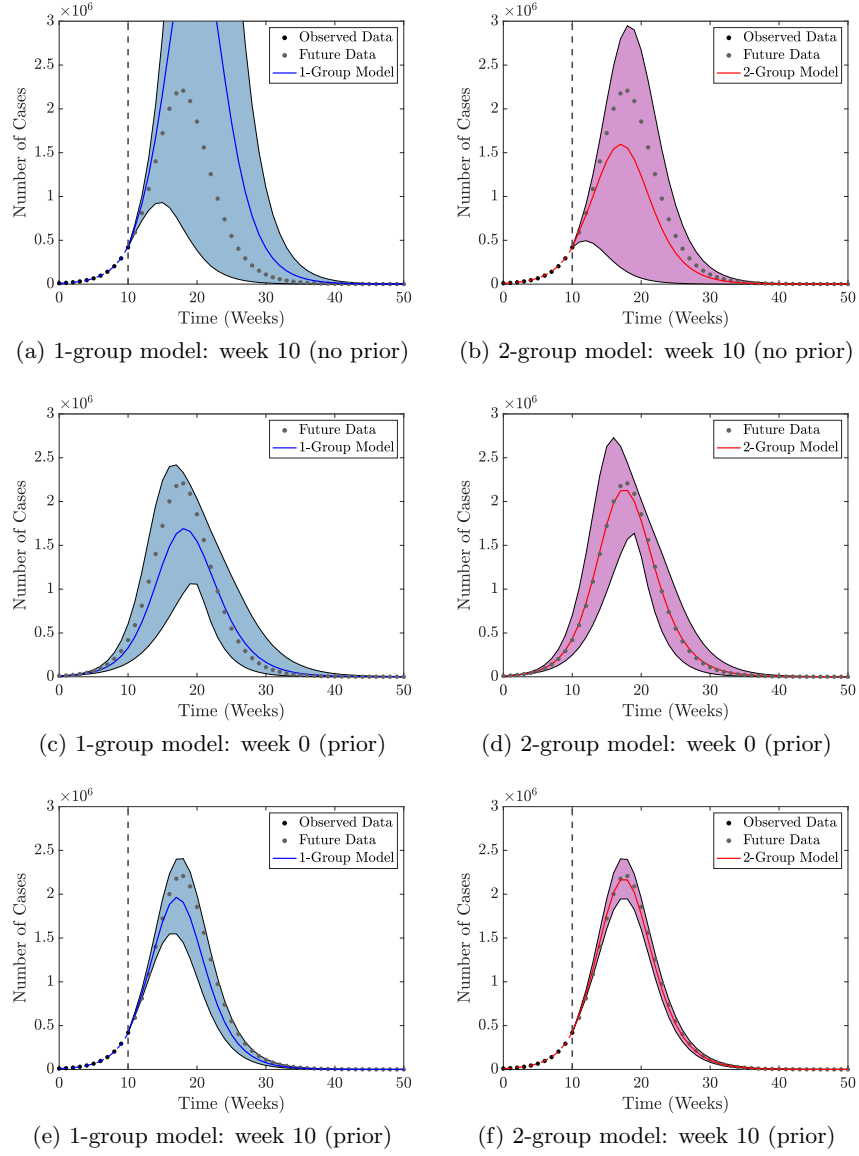


Figure B.3: The primary results of the study, repeated with double the value of the cross-immune fraction ν at the beginning of the 2009 epidemic compared to the value assumed in the main text ($\nu = 2 \times 0.3149 = 0.6298$). Forecasts obtained using the 1-group and 2-group models of a future epidemic occurring 25 years after the 2009 epidemic. Forecasts are generated in three scenarios. (a)-(b): Forecasts generated in week 10 of the future epidemic, with no prior information about the fitting parameters. (c)-(d): Forecasts generated at the start of the future epidemic, with parameters based on the model fits to data from the 2009 epidemic. (e)-(f): Forecasts generated in week 10 of the future epidemic, with priors of the fitting parameters based on the model fits to data from the 2009 epidemic.

B.3 Calculation of R_0 for the 2-group model

Following Diekmann et al. [89] and van den Driessche and Watmough [374], we use the next generation matrix approach to calculate R_0 for the 2-group model. Defining $\mathbf{x} = (E_I, I_I, E_N, I_N)^T$ and $\mathbf{y} = (S_I, S_N)^T$, then using equations (3.7)-(3.14), we can write the following subsystem

$$\dot{\mathbf{x}} = \mathcal{F}(\mathbf{x}, \mathbf{y}) - \mathcal{V}(\mathbf{x}, \mathbf{y}). \quad (\text{B.1})$$

We note that

$$\frac{\partial \mathcal{F}_i}{\partial y_j}(\mathbf{0}, \mathbf{y}_0) = \frac{\partial \mathcal{V}_i}{\partial y_j}(\mathbf{0}, \mathbf{y}_0) = 0, \quad (\text{B.2})$$

where $\mathbf{y}_0 = (\nu N, (1 - \nu)N)^T$ is the disease-free equilibrium. Thus, the linearized equations for $\mathbf{x}$ are given by

$$\dot{\mathbf{x}} = (F - V)\mathbf{x}, \quad (\text{B.3})$$

where F and V have entries

$$F_{ij} = \frac{\partial \mathcal{F}_i}{\partial x_j}(\mathbf{0}, \mathbf{y}_0), \quad V_{ij} = \frac{\partial \mathcal{V}_i}{\partial x_j}(\mathbf{0}, \mathbf{y}_0), \quad (\text{B.4})$$

which are given explicitly by

$$F = \begin{pmatrix} 0 & \beta\nu & 0 & \beta\nu \\ 0 & 0 & 0 & 0 \\ 0 & \beta(1-\nu) & 0 & \beta(1-\nu) \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad V = \begin{pmatrix} \kappa & 0 & 0 & 0 \\ -\kappa & \mu_I & 0 & 0 \\ 0 & 0 & \kappa & 0 \\ 0 & 0 & -\kappa & \mu_N \end{pmatrix}. \quad (\text{B.5})$$

Hence, the next generation matrix $K = FV^{-1}$ is given by

$$K = \begin{pmatrix} \beta\nu/\mu_I & \beta\nu/\mu_I & \beta\nu/\mu_N & \beta\nu/\mu_N \\ 0 & 0 & 0 & 0 \\ \beta(1-\nu)/\mu_I & \beta(1-\nu)/\mu_I & \beta(1-\nu)/\mu_N & \beta(1-\nu)/\mu_N \\ 0 & 0 & 0 & 0 \end{pmatrix}. \quad (\text{B.6})$$

The basic reproduction number R_0 is the spectral radius of K , which is equal to

$$R_0 = \frac{\beta\nu}{\mu_I} + \frac{\beta(1-\nu)}{\mu_N}. \quad (\text{B.7})$$

B.4 Residual analysis of models fitted to the 2009 epidemic

There may be situations in which the assumptions of a model underlying any parameter estimation framework are violated. For instance, the amplitude of the noise in the data may scale differently compared to the corresponding assumption in the model. Incorrect assumptions may not be immediately apparent when observing the fitted models and data. However, to look for any systematic deviations between the fitted models and data, we can carry out a residual analysis, plotting the residuals (the scaled differences between the model values and observed data) against time and the model values [76].

In Figure B.4 we plot the residuals, as calculated by equation (B.13) of the 1-group and 2-group models with least squares estimated parameters (see Appendix B.5), fitted to data from the 2009 H1N1 influenza epidemic in Japan, against the model values (number of cases) and time. There are no trends or any discernible patterns in any panel in Figure B.4, which suggests that it is reasonable to assume that the noise in the observed data of the number of cases scales with the square root of the size of the data.

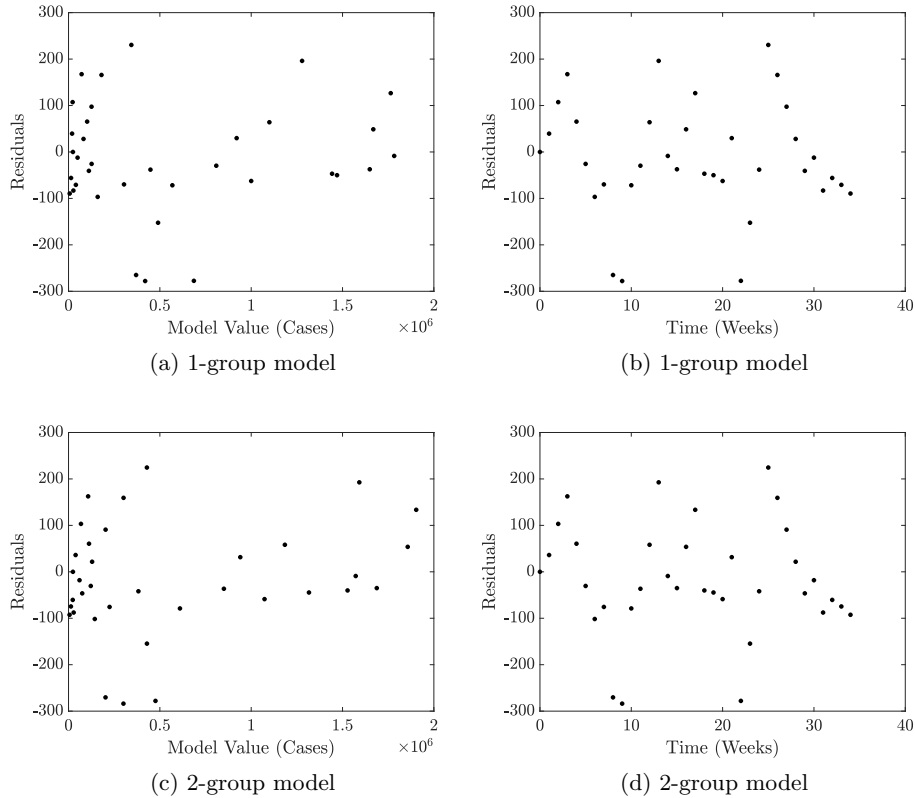


Figure B.4: Residuals of the 1-group and 2-group models fitted to data from the 2009 H1N1 influenza epidemic in Japan, with least squares estimated parameters, calculated using equation (B.13), plotted against the model values and time.

B.5 Parameter estimation

We use Markov chain Monte Carlo (MCMC) with the Metropolis-Hastings algorithm to estimate model parameters [160]. It is assumed that the noise in the observed data scales with the square root of the size of the data (i.e. the number of cases), hence we use the likelihood function

$$\mathcal{L}(\theta; y) = \prod_{j=1}^n \frac{1}{Z} \exp \left(\frac{-(z(t_j; \theta) - y_j)^2}{2z(t_j; \theta_0)\sigma^2} \right), \quad (\text{B.8})$$

where $z(t_j, \theta_0)$ are the number of cases given by our model with parameters θ , y_j are the observed data, and Z is a normalisation constant we need not consider and thus set equal to 1.

To determine σ , we first fit the 1- and 2-group models to data from the 2009 epidemic in Japan using a least squares procedure, again assuming that the observed data scales with the square root of the size of the data, and that the observed data deviates from the model forecasts due to noise which the model does not account for, such that

$$y_j = z(t_j; \theta) + \epsilon_j z(t_j; \theta)^{1/2} \quad \text{for } j = 1, \dots, n, \quad (\text{B.9})$$

where ϵ_j are assumed to be independent and identically distributed random variables with zero mean $E[\epsilon_j] = 0$ and finite variance $\text{var}[\epsilon_j] = \sigma^2$. We define the cost function

$$J(\theta; y) = \sum_{j=1}^n \frac{(y_j - z(t_j; \theta))^2}{z(t_j; \theta)}. \quad (\text{B.10})$$

The least squares parameter estimates are given by

$$\hat{\theta} = \arg \min_{\theta \in \Theta} J(\theta; y), \quad (\text{B.11})$$

where Θ is the feasible set of parameter values. We solve the optimisation problem (B.11) using MATLAB's `fminsearch` function.

Following Cintr3n-Arias et al. [76], we estimate the noise scaling parameter σ^2 by

$$\sigma^2 \approx \frac{1}{n-p} J(\hat{\theta}), \quad (\text{B.12})$$

where p is the number of parameters estimated from the data, and define the residuals by the ratio

$$\text{res}_j = \frac{y_j - z(t_j; \hat{\theta})}{z(t_j; \hat{\theta})^{1/2}}. \quad (\text{B.13})$$

When fitting the models to data from the 2009 epidemic in Japan, we estimate σ using equation (B.12), calculating $\sigma = 123.5$ for the 1-group model and $\sigma = 123.9$ for the 2-group model respectively. Hence, we use the value $\sigma = 124$ in the likelihood function (B.8) throughout this chapter. A residual analysis of both models with the least squares estimated parameters is given in Appendix B.4, justifying the assumption the noise in the observed data scales with square root size of the data.

B.6 Estimation of initial conditions

The observable data we consider is the number of cases that have been recorded each week, $C(j)$. Given that we observe $C(j)$ for $j = 0, 1, \dots, m$, we wish to estimate the number of individuals in each compartment of our models at week j as initial conditions to make forward forecasts.

When estimating initial conditions from case data from multiple previous weeks (i.e. when we are part way through an epidemic), it is necessary to take data from every week into account, as for instance, an individual who was reported as being infected many weeks ago may not have recovered yet. As we estimate initial conditions from the observed data rather than model trajectories, we make the assumption that the cases are generated uniformly in time (e.g. if there were seven cases reported in one week, we assume that there was one new case each day).

Assuming an exponentially distributed infectious period as in our models, and first considering the 1-group model, if $C(j)$ cases were reported each week for $j = 0, 1, \dots, m$, the number of infected individuals I at week m is given by

$$I(m) = \sum_{j=0}^m \int_{j-1}^j e^{-\mu(m-t)} C(j) dt \quad (\text{B.14})$$

$$= \sum_{j=0}^m \frac{C(j)}{\mu} \left[e^{-\mu(m-j)} - e^{-\mu(m+1-j)} \right]. \quad (\text{B.15})$$

Restating equation (3.6), we have

$$C(m) = \int_{m-1}^m \kappa E dt. \quad (\text{B.16})$$

Hence, by assuming a constant E over the past week, we can make the estimation

$$E(m) = \frac{C(m)}{\kappa}. \quad (\text{B.17})$$

The number of newly recovered individuals each week is given by the difference between the number of new cases and the number of individuals relating to those cases who are still infected. Hence, we estimate the number of recovered individuals R at week m by

$$R(m) = C(j) \left(1 - \frac{1}{\mu} \left[e^{-\mu(m-j)} - e^{-\mu(m+1-j)} \right] \right). \quad (\text{B.18})$$

As we have assumed a total fixed effective population size N , the number of susceptibles at week m is given by

$$S(m) = N - [E(m) + I(m) + R(m)]. \quad (\text{B.19})$$

We now consider the 2-group model. As it is assumed that only cases of naive individuals becoming infected are reported, we estimate the number of infected naive individuals at week m to be

$$I_N(m) = \sum_{j=0}^m \frac{C(j)}{\mu_N} \left[e^{-\mu_N(m-j)} - e^{-\mu_N(m+1-j)} \right]. \quad (\text{B.20})$$

As cross-immune and naive individuals are equally likely to be infected, we estimate the number of new (unrecorded) cases of infectious cross-immune individuals each week to be $\frac{\nu}{1-\nu}C(j)$. Hence, we estimate that

$$I_I(m) = \sum_{j=0}^m \frac{\nu}{1-\nu} \frac{C(j)}{\mu_I} \left[e^{-\mu_I(m-j)} - e^{-\mu_I(m+1-j)} \right]. \quad (\text{B.21})$$

The number of exposed, recovered, and susceptible individuals at week m in each group are estimated by

$$E_N(m) = \frac{C(m)}{\kappa}, \quad (\text{B.22})$$

$$E_I(m) = \frac{\nu}{1-\nu} \frac{C(m)}{\kappa}, \quad (\text{B.23})$$

$$R_N(m) = \sum_{j=0}^m C(j) \left(1 - \frac{1}{\mu_N} \left[e^{-\mu_N(m-j)} - e^{-\mu_N(m+1-j)} \right] \right), \quad (\text{B.24})$$

$$R_I(m) = \sum_{j=0}^m \frac{\nu}{1-\nu} C(j) \left(1 - \frac{1}{\mu_I} \left[e^{-\mu_I(m-j)} - e^{-\mu_I(m+1-j)} \right] \right), \quad (\text{B.25})$$

$$S_N(m) = (1-\nu)N - [E_N(m) + I_N(m) + R_N(m)], \quad (\text{B.26})$$

$$S_I(m) = \nu N - [E_I(m) + I_I(m) + R_I(m)]. \quad (\text{B.27})$$

B.7 Deviance information criterion (DIC)

To compare model forecasts throughout this chapter, we use the deviance information criterion (DIC) as stated in Spiegelhalter et al. [350]. The deviance $D(\theta) = -2 \log(\mathcal{L}(\theta; y)) + C$ where C is a constant we need not consider and therefore set equal to 0. With our likelihood function $\mathcal{L}(\theta; y)$ defined as in (B.8), the deviance is

$$D(\theta) = \sum_{j=1}^n \frac{(z(t_j; \theta) - y_j)^2}{z(t_j; \theta) \sigma^2}. \quad (\text{B.28})$$

Then the DIC is given by

$$\text{DIC} = \overline{D(\theta)} + p_D, \quad (\text{B.29})$$

where $p_D = \overline{D(\theta)} - D(\bar{\theta})$ and $\bar{\theta}$ is the mean value of the parameters θ . Models with a smaller DIC value should be preferred when describing data and making forecasts.

The DIC values for each forecast in this study are given in Table B.1.

1-group model		2-group model	
Figure	DIC	Figure	DIC
3.3 (a)	50.90	3.3 (b)	52.80
3.4 (a)	24,960	3.4 (b)	1,669
3.4 (c)	8,588	3.4 (d)	1,772
3.4 (e)	4.151	3.4 (f)	3.644
3.5 (a)	2,621	3.5 (b)	175.7
3.5 (c)	440.5	3.5 (d)	258.5
3.5 (e)	2,476	3.5 (f)	306.5
3.6 (a)	2,621	3.6 (b)	175.7
3.6 (c)	414.6	3.6 (e)	25.62
3.6 (d)	608.7		
3.6 (f)	305.5	3.6 (h)	13.00
3.6 (g)	389.5		
3.6 (i)	4.556	3.6 (k)	2.535
3.6 (j)	4.666		
B.2 (a)	7,493	B.2 (b)	152.3
B.2 (c)	22,280	B.2 (d)	2,144
B.2 (e)	1,029	B.2 (f)	33.60
B.3 (a)	440.1	B.3 (b)	194.3
B.3 (c)	7,789	B.3 (d)	539.0
B.3 (e)	84.10	B.3 (f)	13.01
B.7 (a)	4,972	B.7 (b)	871.2
B.7 (c)	5,195	B.7 (d)	2,105
B.7 (e)	39,800	B.7 (f)	2,212
B.8 (a)	2,190	B.8 (b)	33.46
B.8 (c)	120.8	B.8 (d)	48.64
B.8 (e)	1,372	B.8 (f)	58.57
B.10 (a)	4,972	B.10 (b)	871.2
B.10 (c)	647.4	B.10 (d)	76.35
B.10 (e)	342.5	B.10 (f)	40.39
B.10 (g)	4.392	B.10 (h)	3.064
B.11 (a)	2,190	B.11 (b)	33.46
B.11 (c)	630.0	B.11 (d)	10.27
B.11 (e)	299.9	B.11 (f)	4.270
B.11 (g)	44.50	B.11 (h)	1.652

Table B.1: The DIC values for each forecast in Chapter 3.

B.8 Modelling the cross-immune fraction of the population

We assume that all infected naive individuals in the 2009 H1N1 influenza epidemic in Japan acquired cross-immunity to the virus and related strains. Hence, we can compute the cross-immune fraction of the population immediately after the epidemic, ν^* , by

$$\nu^* = \nu + \frac{C_T}{N}, \quad (\text{B.30})$$

where ν is the cross-immune fraction before the epidemic, C_T is the number of total cases recorded in the data, and N is the total effective population size. Hence, from the 2-group model fitted to data from the epidemic as shown in Figure 3.3, we estimate the cross-immune fraction of the population immediately after the epidemic to be $\nu^* = 0.7443$, using the mean estimated total effective population size, $N = 4.660 \times 10^7$.

To determine the cross-immune fraction when a second epidemic in the future begins, we shall compute the fraction as a function of the time since the end of the first epidemic. We shall assume that once an individual acquires cross-immunity, they remain cross-immune for their lifetime.

Using population data from Japan in 2009 [271], which we display in Figure B.5, we define the number of individuals aged a by $b(a)$, and the number of deaths of individuals aged a by $d(a)$. Hence, the probability, $P(a, n)$, that an individual of age a survives for n more years is given by

$$P(a, n) = \prod_{i=0}^{a-1} \left(1 - \frac{d(n+i)}{b(n+i)} \right), \quad (\text{B.31})$$

where $d(a)/b(a)$ is the probability than an individual aged a dies within one year.

The fraction of individuals who have survived for n years is then given by

$$\frac{1}{M} \sum_{a=0}^{a_m} b(a) P(a, n), \quad (\text{B.32})$$

where $M = \sum_{a=0}^{a_m} b(a)$, and a_m is the assumed maximum age of an individual, which we take to be 110 years old. Hence, assuming a constant population size and that all individuals are born immunologically naive, we can calculate the immune fraction ν , n years after the first epidemic to be

$$\nu(n) = \frac{\nu^*}{N} \sum_{a=0}^{\infty} P(a, n), \quad (\text{B.33})$$

where ν^* is the immune fraction immediately after the first epidemic. The result is shown in Figure B.9 (a) (see Appendix B.12). For an epidemic occurring 25 years after 2009 as we consider in this study, the cross-immune fraction of the population at the start of the epidemic is $\nu = 0.5129$.

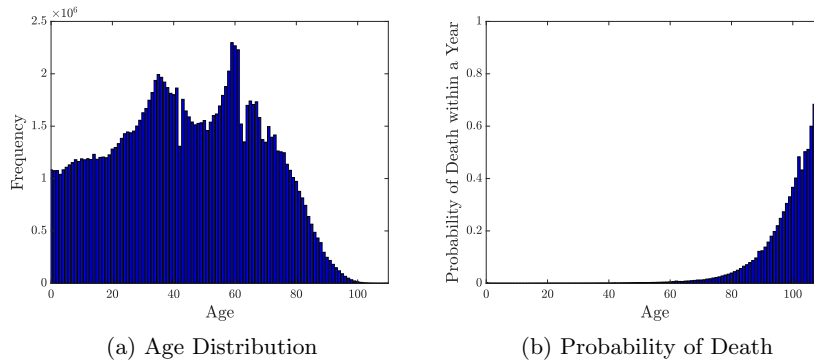


Figure B.5: (a): The age distribution of the population of Japan in 2009. (b): The age distribution of the probability that an individual in Japan dies within one year, based on mortality and population data from 2009. Data accessed from National Institute of Population and Social Security Research [271].

B.9 The 2-group model fitted to the 2009 epidemic: total recorded and unrecorded cases

We assume that all cases of immunologically naive infected individuals are recorded in the data, and that no cases relating to cross-immune individuals are recorded, due to the assumption that naive individuals suffer more severe symptoms and are therefore more likely to seek medical attention. In Figure B.6 we display the number of (recorded) infected naive individuals, the number of (unrecorded) infected cross-immune individuals, and their combined total, estimated by the 2-group model fitted to the data from the 2009 epidemic in Japan.

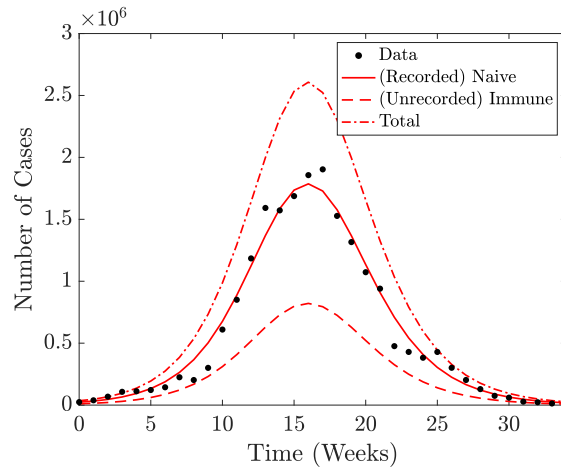


Figure B.6: The weekly number of recorded, unrecorded, and total (recorded and unrecorded) cases as estimated by the 2-group model with mean estimated parameters, fitted to data from the 2009 H1N1 influenza epidemic in Japan.

B.10 Prior fitting parameter distributions

We prescribe prior gamma parameter distributions on β and N using estimated values from the 2009 epidemic. Different variances of these distributions may be considered when making forecasts, values of which are given in Table B.2.

Figure	Parameter	Mean	Variance	Relative Variance
3.5, 3.6 (a), 3.6(c), 3.6(f), 3.6(i)	β	1.644	5×10^{-3}	1
3.5, 3.6 (a), 3.6(c), 3.6(f), 3.6(i)	N	3.063×10^7	1×10^{13}	1
3.6 (d), 3.6 (g), 3.6 (j), B.7, B.10	β	1.644	2.5×10^{-2}	5
3.6 (d), 3.6 (g), 3.6 (j), B.7, B.10	N	3.063×10^7	5×10^{13}	5
B.8, B.11	β	1.644	1×10^{-3}	0.2
B.8, B.11	N	3.063×10^7	2×10^{12}	0.2

(a): 1-group model

Figure	Parameter	Mean	Variance	Relative Variance
3.5, 3.6	β	1.947	5×10^{-3}	1
3.5, 3.6	N	4.649×10^7	1×10^{13}	1
B.7, B.10	β	1.947	2.5×10^{-2}	5
B.7, B.10	N	4.649×10^7	5×10^{13}	5
B.8, B.11	β	1.947	1×10^{-3}	0.2
B.8, B.11	N	4.649×10^7	2×10^{12}	0.2

(b): 2-group model

Table B.2: Mean and variance of prior gamma distributions of the transmission rate β and effective population size N used to make forecasts of future epidemics using the 1-group and 2-group models, set by fitting models to data from the 2009 epidemic.

B.11 Forecasting epidemics in advance using different prior parameter distributions

We show forecasts of future epidemics made in advance of future outbreaks occurring 25, 50, and 75 years after the 2009 epidemic, if variances five times larger (Figure B.7) or five times smaller (Figure B.8) of both β and N are used, compared to forecasts made in Figure 3.5.

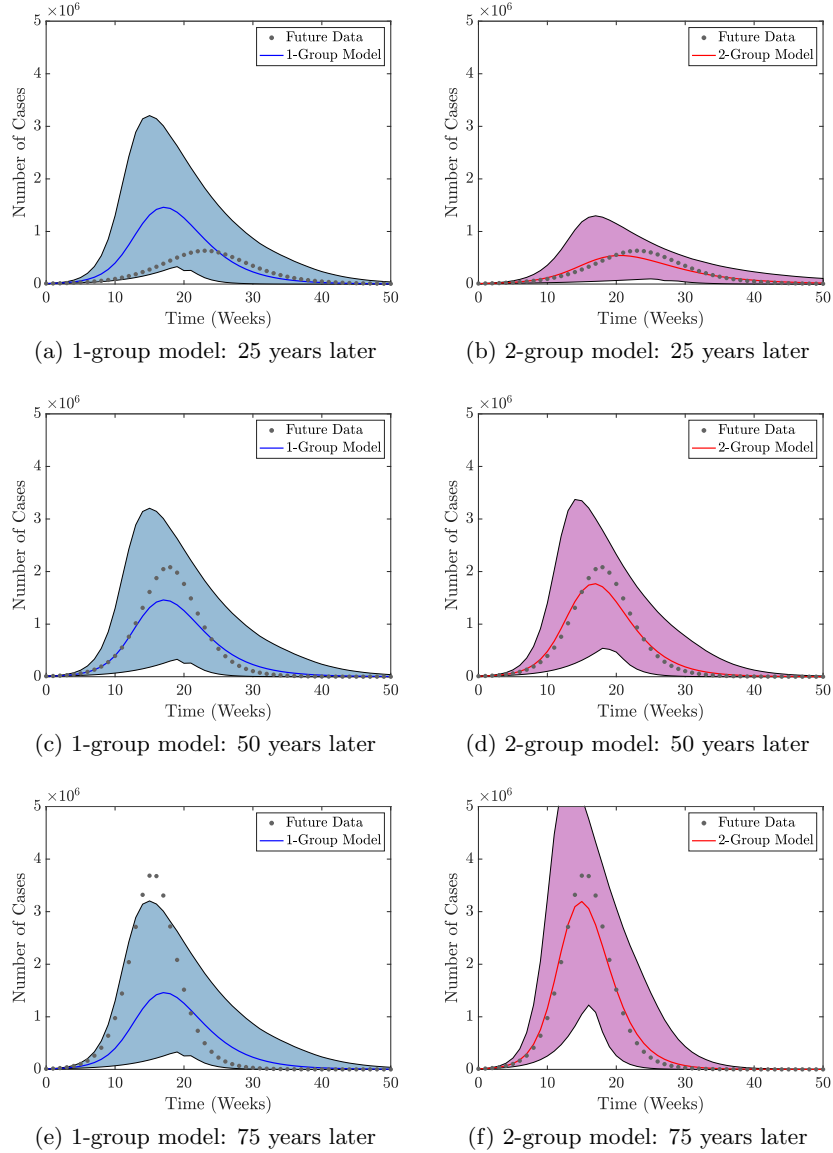


Figure B.7: Forecasts obtained from the 1-group and 2-group models of epidemics occurring 25, 50, and 75 years after the 2009 epidemic as in Figure 3.5, but with variances of the gamma distributions of β and N five times larger (details in Appendix B.10).

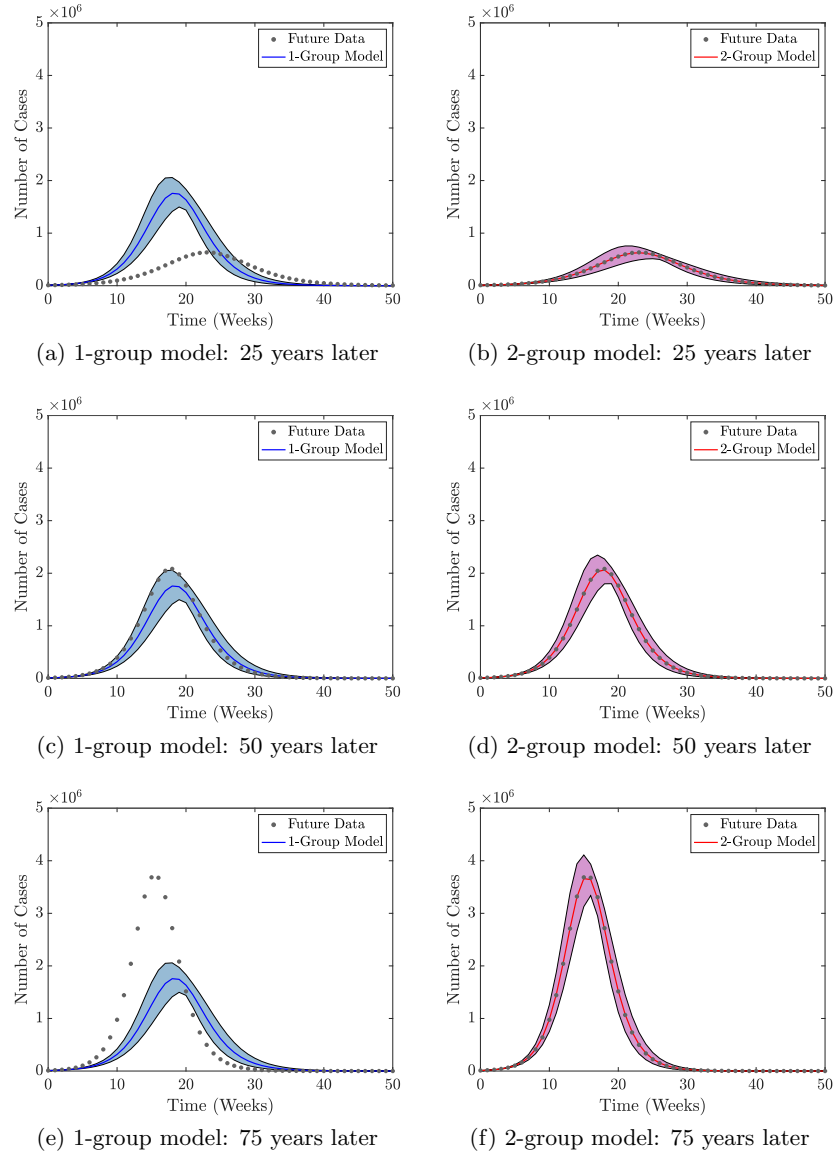


Figure B.8: Forecasts obtained from the 1-group and 2-group models of epidemics occurring 25, 50, and 75 years after the 2009 epidemic as in Figure 3.5, but with variances of the gamma distributions of β and N five times smaller (details in Appendix B.10).

B.12 Outbreak statistics of future epidemics

In Figure B.9 we show how the cross-immune fraction, basic reproduction numbers, and various quantities of interest relating to the dynamics of a future epidemic (estimated using both the 1-group and 2-group models) vary dependent on when the future epidemic occurs. The cross-immune fraction decreases approximately linearly before tapering to zero as time increases (Figure B.9 (a)). This results in an increase in the basic reproduction number, R_0 , of the 2-group model (Figure B.9 (b)).

We explore various quantities of interest which would inform policy makers on how best to control the epidemic: (i) the total number of recorded cases of individuals seeking medical attention, (ii) the duration of the epidemic, (iii): the maximum weekly number of recorded cases, and (iv): the time at which the maximum weekly number of recorded cases occurs. Under the assumptions of the 1-group model, these are all forecasted to be fixed irrespective of when the next epidemic begins, and identical to that of the 2009 epidemic. However, under the assumptions of the 2-group model, these quantities all vary, which we now describe, as well as motivations for determining them.

The total number of recorded cases of individuals seeking medical attention reflects the number of total medical resources (e.g. antivirals, hospital beds, and staff) required to treat individuals (Figure B.9 (c)). This increases in the 2-group model the further into the future the next epidemic begins, due to the decreasing cross-immune population fraction. It would be useful for public health officials and medical facilities to know the expected duration of the epidemic so they can plan and allocate resources accordingly. Here we define the duration of the epidemic to be the time until the number of recorded cases reaches fewer than 1000 per week (orders of magnitude smaller than the peak of the epidemic). The duration of the next epidemic would decrease the further into the future it occurred, and is inversely correlated to the number of cases (Figure B.9 (d)). This is because if there was a greater proportion of immunologically naive individuals in the population, the basic reproduction number would be larger and hence the virus would spread more quickly leading to a shorter epidemic duration.

The maximum weekly number of recorded cases (i.e. the peak of the epidemic) and the time at which this peak occurs are important to public health officials with regards to determining the maximum number of resources required to help treat patients, and when these resources would be needed. There is a direct correlation between the total number of recorded cases and the peak number of weekly recorded cases, as well as the duration of the epidemic and the time of the peak number of cases (Figure B.9 (e)-(f)).

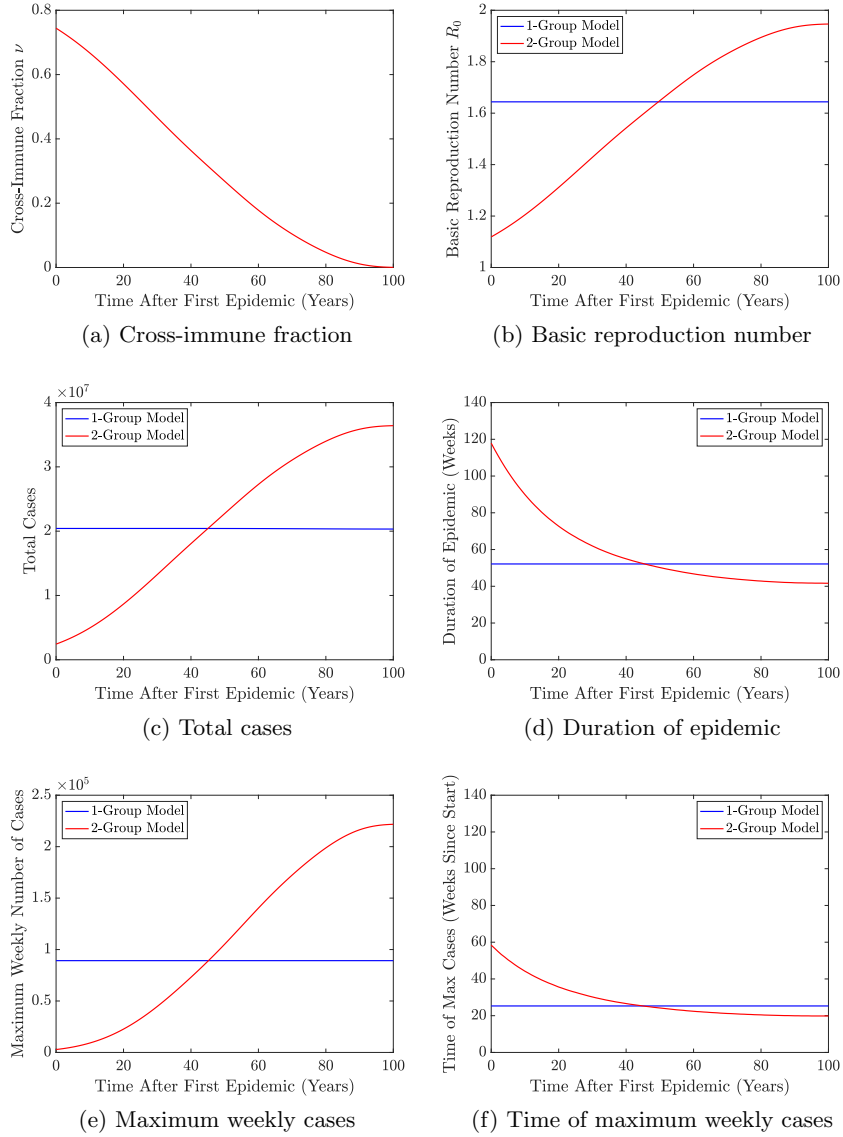


Figure B.9: (a)-(b): The cross-immune fraction of the 2-group model and the basic reproduction numbers of the 1-group and 2-group models if a second epidemic were to occur at a given time after the 2009 epidemic. (c)-(f): The total number of recorded cases, the duration of the epidemic, the maximum number of weekly recorded cases, and the time at which this maximum occurs, if a second epidemic were to occur at a given time after the 2009 epidemic, as calculated by the 1-group and 2-group models. Quantities calculated using the mean parameter values estimated by fitting models to the 2009 H1N1 epidemic in Japan.

B.13 Forecasting an epidemic in real-time with prior information using different prior parameter distributions

We show real-time forecasts of an outbreak occurring 25 years after the 2009 epidemic (using prior information from the 2009 epidemic), if variances five times larger (Figure B.10) or five times smaller (Figure B.11) of both β and N are used, compared to forecasts made in Figure 3.6. Variances remain fixed in each figure.

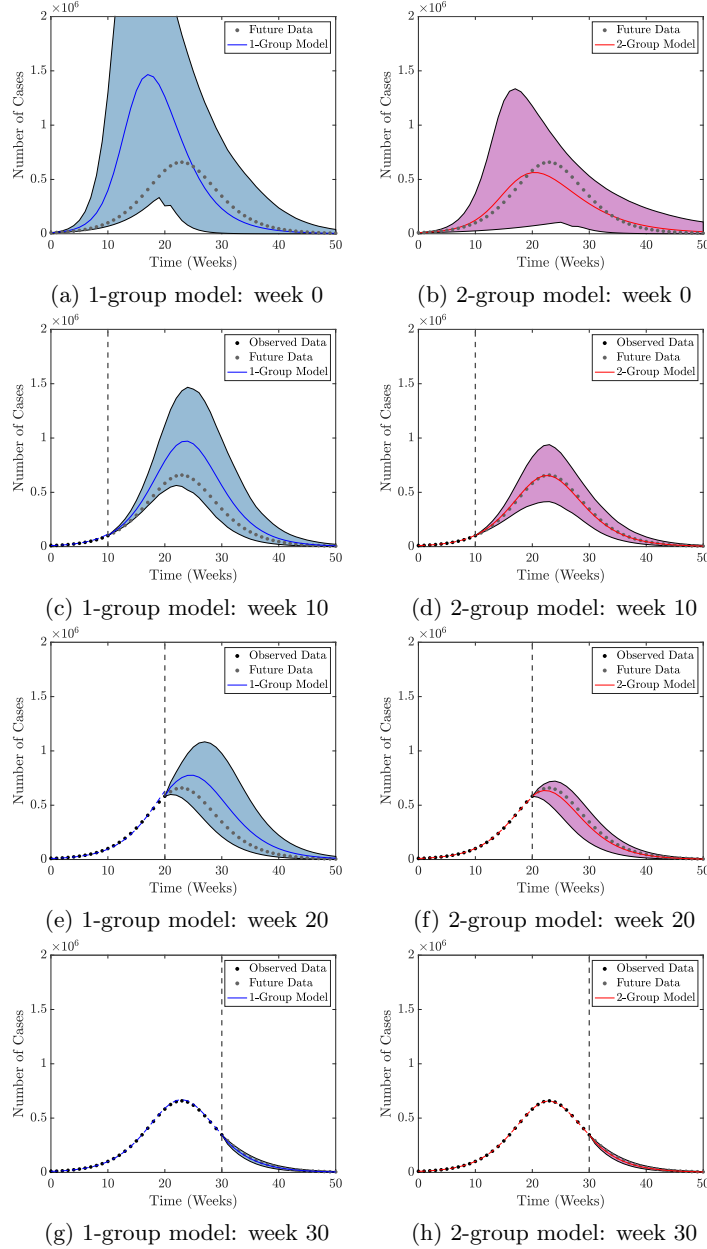


Figure B.10: Forecasts obtained from the 1-group and 2-group models of a future epidemic occurring 25 years after the 2009 epidemic in real-time using information from the 2009 epidemic as in Figure 3.6, but with variances of the gamma distributions of β and N five times larger (details in Appendix B.10).

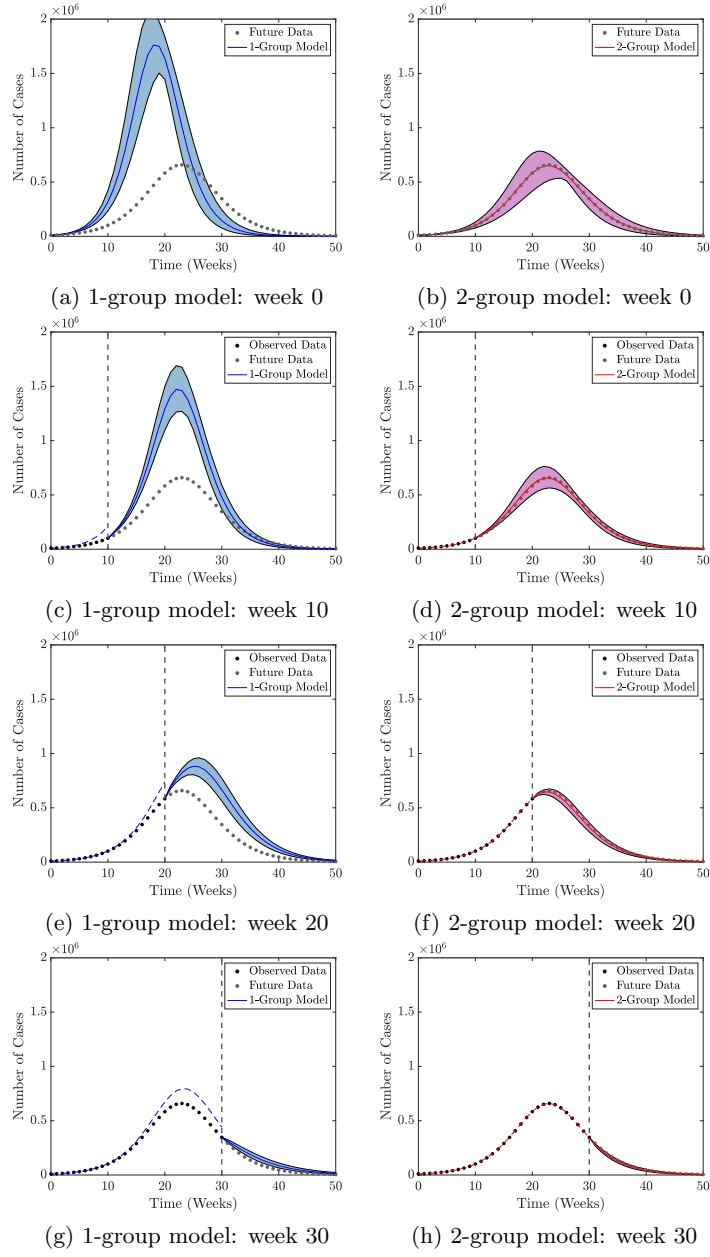


Figure B.11: Forecasts obtained from the 1-group and 2-group models of a future epidemic occurring 25 years after the 2009 epidemic in real-time using information from the 2009 epidemic as in Figure 3.6, but with variances of the gamma distributions of β and N five times smaller (details in Appendix B.10).

Appendix C

Appendix to Chapter 4

C.1 Parameter estimation

We use Markov chain Monte Carlo (MCMC) with the Metropolis-Hastings algorithm [160] and maximum likelihood estimation (MLE) to estimate model parameters. Both methods utilise the same likelihood function. We assume that the measured viral titer and adaptive immune response levels are lognormally distributed around their true values, with standard deviation $0.3 \log_{10}$ (as in Hadjichrysanthou et al. [147]) and $0.1 \log_{10}$ respectively, and that the measured dead cell fractions are normally distributed around their true values, with standard deviation 0.05 (these latter two values determined by estimating the level of noise in the data using MLE). The likelihood function for parameter estimation is given by

$$\mathcal{L}(\theta; y) = \frac{1}{Z} \prod_{i=1}^M \prod_{j=1}^{N_i} \exp \left(\frac{-[z_i(t_j; \theta) - y_{ij}]^2}{2\sigma_i^2} \right), \quad (\text{C.1})$$

where $z_i(t_j, \theta)$ represent model outputs generated with parameters θ , y_{ij} represent data, and σ_i are noise scaling parameters. The subscripts $i = 1, \dots, M$ relate to the observables (e.g. viral load, fraction of dead cells, etc.) and the subscripts $j = 1, \dots, N_i$ relate to the output time-points.

We fix the normalisation constant $Z = 1$ without loss of generality. When considering data for the viral load or adaptive immune response, as in Handel et al. [153], $z_i(t_j, \theta)$ and y_{ij} are the base 10 logarithms of the model outputs and data. If we assume normally distributed noise, then we can also define the weighted sum of squared residuals (SSR) via the cost function $J(\theta; y)$ where

$$J(\theta; y) = \sum_{i=1}^M \sum_{j=1}^{N_i} \left(\frac{-[z_i(t_j; \theta) - y_{ij}]^2}{2\sigma_i^2} \right). \quad (\text{C.2})$$

The MLE parameter values are defined in terms of the cost function as follows:

$$\hat{\theta} = \arg \min_{\theta \in \Theta} J(\theta; y), \quad (\text{C.3})$$

where Θ denotes the set of feasible parameter values. We solve the least squares optimisation problem (C.3) using MATLAB's `fminsearch` function.

C.2 Deviance information criterion (DIC) and Akaike information criterion (AIC)

To compare model fits and predictions when parameters are estimated using MCMC, we use the deviance information criterion (DIC) as stated in [128]. With the likelihood function $\mathcal{L}(\theta; y)$ defined as in (C.1), the deviance, $D(\theta) = -2 \log(\mathcal{L}(\theta; y))$ is given by

$$D(\theta) = \sum_{i=1}^M \sum_{j=1}^{N_i} \left(\frac{-[z_i(t_j; \theta) - y_{ij}]^2}{2\sigma_i^2} \right), \quad (\text{C.4})$$

and the DIC is given by

$$\text{DIC} = \overline{D(\theta)} + p_D, \quad (\text{C.5})$$

where $p_D = \frac{1}{2} \overline{\text{var}(D(\theta))}$. Models with a smaller DIC value should be preferred when describing data and making forecasts.

To compare model fits and predictions in which parameters were estimated using MLE, we use the corrected Akaike Information Criterion (AIC) [53], which is given by

$$\text{AIC} = 2m + n \log \sum_{i=1}^M \sum_{j=1}^{N_i} \left(\frac{-[z_i(t_j; \theta) - y_{ij}]^2}{2n\sigma_i^2} \right) + \frac{2m(m+1)}{n-m+1}, \quad (\text{C.6})$$

where $m - 1$ is the number of parameters, and n is the total number of datapoints.

C.3 Global sensitivity analysis

We carry out a global sensitivity analysis of the IR model, using the area under the viral load curve (AUC), duration of symptomatic infection, and fraction of dead cells at the end of the infection as quantities of interest. For robustness we use both the eFAST and Sobol techniques [333, 334, 348]. We sample from the parameter space $\theta_i \in [0.5 \times \hat{\theta}_i, 2 \times \hat{\theta}_i]$ where $\hat{\theta}_i$ are the parameter values given in Tables 4.1 and 4.6.

The first-order sensitivity index of a parameter evaluates the fractional contribution of a single parameter on the output variance, while the total-order index accounts for the interaction between the parameter of interest and other model parameters. A first-order sensitivity index of 1 indicates that all the variance in the model output is due to variation in that parameter. Similarly, a total-order sensitivity index of 1 indicates that all the variance in the model output is due to variation in that parameter and its interactions with other parameters. Parameters with a total-order sensitivity index less than or equal to that of the dummy parameter ‘du’ using the eFAST technique should be considered not significantly different from zero.

The results presented in Figure C.1 reveal that the viral infection rate β , viral production rate p , and initial number of uninfected epithelial cells T_0 are the most important parameters with regards to the three model outputs. We note further that each of these parameters relate explicitly to viral production and cell infection rather than viral clearance. Many other parameters, notably δ , c , κ , τ , r , and m have a significant effect on the duration of infection, which is to be expected as these parameters determine the immune response and viral clearance. The good agreement between the results of the eFAST and Sobol sensitivity indices demonstrates the robustness of our findings.

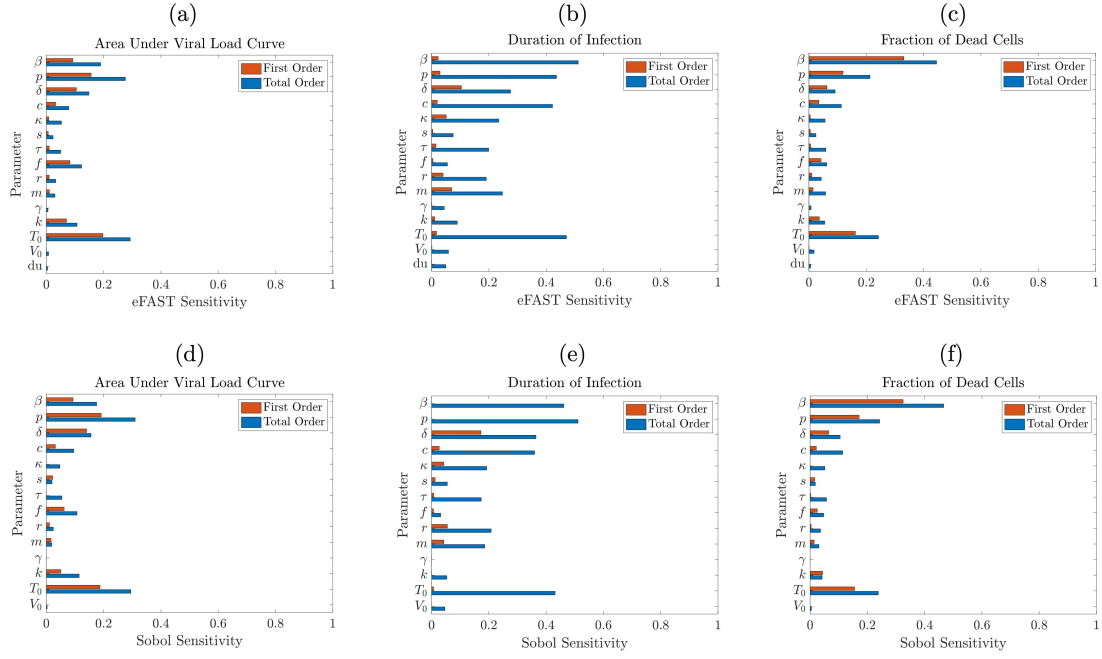


Figure C.1: eFAST (a),(b),(c) and Sobol (d),(e),(f) sensitivity analyses of the IR model, where model parameters θ_i are sampled from the range $\theta_i \in [0.5 \times \hat{\theta}_i, 2 \times \hat{\theta}_i]$ and $\hat{\theta}_i$ correspond to the parameters stated in Tables 4.1 and 4.6. The unit of the area under the viral load curve is (PFU days), the unit of the duration is (days), while the fraction of dead cells is dimensionless.

C.4 Structural identifiability analysis of the IR model

C.4.1 Nondimensionalisation

We partially nondimensionalise the IR model by defining

$$\tilde{T} = \frac{T}{T_0}, \quad \tilde{E} = \frac{E}{T_0}, \quad \tilde{I} = \frac{I}{T_0}, \quad C = \frac{D}{T_0}, \quad \tilde{X} = \frac{X}{f}. \quad (\text{C.7})$$

Under this partial nondimensionalisation, equations (4.1)–(4.7) transform to give

$$\text{fraction of uninfected cells} \quad \frac{d\tilde{T}}{dt} = -\beta\tilde{T}V, \quad (\text{C.8})$$

$$\text{fraction of latently infected cells} \quad \frac{d\tilde{E}}{dt} = \beta\tilde{T}V - \kappa\tilde{E}, \quad (\text{C.9})$$

$$\text{fraction of productively infected cells} \quad \frac{d\tilde{I}}{dt} = \kappa\tilde{E} - \delta\tilde{I}, \quad (\text{C.10})$$

$$\text{fraction of dead cells} \quad \frac{dC}{dt} = \delta\tilde{I}, \quad (\text{C.11})$$

$$\text{free virions} \quad \frac{dV}{dt} = \frac{(pT_0)\tilde{I}}{1 + sF(t - \tau)} - cV - (fk)V\tilde{X}, \quad (\text{C.12})$$

$$\text{normalised adaptive IR} \quad \frac{d\tilde{X}}{dt} = V + \frac{r\tilde{X}V}{1 + mV} - \gamma\tilde{X}, \quad (\text{C.13})$$

$$\text{innate IR} \quad F(t) = \begin{cases} 0, & \text{for } t \leq 0, \\ A_1 e^{\lambda_1 t} & \text{for } 0 < t \leq t^*, \\ A_2 e^{-\lambda_2 t} & \text{for } t > t^*. \end{cases} \quad (\text{C.14})$$

We conclude that if only the viral load V is observed and we do not observe data on the number of cells (T , E , I , or D) or the adaptive immune response X , then we can only identify the parameter groupings pT_0 and fk .

C.4.2 Taylor series calculations: $\tau = 0$ (Case 1b extension)

We first consider the case where we observe data for the viral load V , fraction of dead cells $C = D/T_0$, and adaptive immune response X . We assume that the values of the parameters T_0, δ, κ, A_1 , and λ_1 are known, and that $\tau = 0$ (i.e. no time delay). We seek to determine the identifiability of parameters $\beta, p, V_0, s, f, r, m, k, \gamma$, and c (an extension of Case 1b in Table 4.2 where we assume γ, r , and c are also known).

We expand the observables, $V(t; \theta)$, $C(t; \theta)$, and $X(t; \theta)$ in a neighbourhood of $t = 0$ as a Taylor series given by

$$V(t; \theta) = a_0 + a_1 t + a_2 \frac{t^2}{2!} + \dots, \quad (\text{C.15})$$

$$C(t; \theta) = b_0 + b_1 t + b_2 \frac{t^2}{2!} + \dots, \quad (\text{C.16})$$

$$X(t; \theta) = d_0 + d_1 t + d_2 \frac{t^2}{2!} + \dots, \quad (\text{C.17})$$

where

$$a_i = \lim_{t \rightarrow 0^+} V^{(i)}(t; \theta), \quad (\text{C.18})$$

$$b_i = \lim_{t \rightarrow 0^+} \frac{D^{(i)}(t; \theta)}{T_0}, \quad (\text{C.19})$$

$$d_i = \lim_{t \rightarrow 0^+} X^{(i)}(t; \theta). \quad (\text{C.20})$$

The non-zero coefficients needed to determine model identifiability are

$$a_0 = V_0, \quad (\text{C.21})$$

$$a_1 = -V_0 c, \quad (\text{C.22})$$

$$a_2 = -fkV_0^2 + V_0 c^2, \quad (\text{C.23})$$

$$a_3 = V_0 f(2V_0 ck + V_0 k(\gamma - (V_0 r)/(V_0 m + 1))) - V_0 c(c^2 - 2V_0 fk) \quad (\text{C.24})$$

$$+ (T_0 V_0 \beta \kappa p)/(A_1 s + 1), \quad (\text{C.25})$$

$$\begin{aligned} a_4 = & -V_0 c(f(2V_0 ck + V_0 k(\gamma - (V_0 r)/(V_0 m + 1))) - c(c^2 - 2V_0 fk) \\ & + V_0 f(2ck + k(\gamma - (V_0 r)/(V_0 m + 1))) \\ & - V_0 k(r/(V_0 m + 1) - (V_0 mr)/(V_0 m + 1)^2)) + 2V_0 cfk + (T_0 \beta \kappa p)/(A_1 s + 1) \\ & - V_0 f((\gamma - (V_0 r)/(V_0 m + 1))(2V_0 ck + V_0 k(\gamma - (V_0 r)/(V_0 m + 1))) \\ & - 2V_0^2 fk^2 + V_0 k(c^2 - 2V_0 fk) + V_0 c(2ck + k(\gamma - (V_0 r)/(V_0 m + 1))) \end{aligned} \quad (\text{C.26})$$

$$\begin{aligned} & - V_0 k(r/(V_0 m + 1) - (V_0 mr)/(V_0 m + 1)^2)) - T_0 V_0 \beta (\kappa((cp)/(A_1 s + 1) \\ & + (\delta p)/(A_1 s + 1) + (A_1 \lambda ps)/(A_1 s + 1)^2) + (\kappa^2 p)/(A_1 s + 1) \\ & + (A_1 \kappa \lambda ps)/(A_1 s + 1)^2) - (T_0 V_0^2 \beta^2 \kappa p)/(A_1 s + 1) - (A_1 T_0 V_0 \beta \kappa \lambda ps)/(A_1 s + 1)^2, \end{aligned}$$

$$b_3 = V_0 \beta \delta \kappa, \quad (\text{C.27})$$

$$d_1 = fV_0, \quad (\text{C.28})$$

$$d_2 = -V_0 cf - V_0 f(\gamma - (V_0 r)/(V_0 m + 1)), \quad (\text{C.29})$$

$$\begin{aligned} d_3 = & V_0 c(cf + f(\gamma - (V_0 r)/(V_0 m + 1)) - V_0 f(r/(V_0 m + 1) - (V_0 mr)/(V_0 m + 1)^2)) \\ & - V_0 f(V_0 fk - (\gamma - (V_0 r)/(V_0 m + 1))^2 + V_0 c(r/(V_0 m + 1) \\ & - (V_0 mr)/(V_0 m + 1)^2)), \end{aligned} \quad (\text{C.30})$$

$$\begin{aligned}
d_4 = & V_0 c(f(V_0 f k - (\gamma - (V_0 r)/(V_0 m + 1))^2 + V_0 c(r/(V_0 m + 1) \\
& - (V_0 m r)/(V_0 m + 1)^2)) - c(c f + f(\gamma - (V_0 r)/(V_0 m + 1)) - V_0 f(r/(V_0 m + 1) \\
& - (V_0 m r)/(V_0 m + 1)^2)) + V_0 f(f k + c(r/(V_0 m + 1) - (V_0 m r)/(V_0 m + 1)^2) \\
& + 2(r/(V_0 m + 1) - (V_0 m r)/(V_0 m + 1)^2)(\gamma - (V_0 r)/(V_0 m + 1)) \\
& - V_0 c((2 m r)/(V_0 m + 1)^2 - (2 V_0 m^2 r)/(V_0 m + 1)^3)) + V_0 c(2 f(r/(V_0 m + 1) \\
& - (V_0 m r)/(V_0 m + 1)^2) - V_0 f((2 m r)/(V_0 m + 1)^2 - (2 V_0 m^2 r)/(V_0 m + 1)^3))) \\
& + V_0 f((\gamma - (V_0 r)/(V_0 m + 1))(V_0 f k - (\gamma - (V_0 r)/(V_0 m + 1))^2 + V_0 c(r/(V_0 m + 1) \\
& - (V_0 m r)/(V_0 m + 1)^2)) + V_0 c(f k + c(r/(V_0 m + 1) \\
& - (V_0 m r)/(V_0 m + 1)^2) + 2(r/(V_0 m + 1) - (V_0 m r)/(V_0 m + 1)^2)(\gamma \\
& - (V_0 r)/(V_0 m + 1)) - V_0 c((2 m r)/(V_0 m + 1)^2 - (2 V_0 m^2 r)/(V_0 m + 1)^3)) \\
& + V_0 k(c f + f(\gamma - (V_0 r)/(V_0 m + 1)) - V_0 f(r/(V_0 m + 1) - (V_0 m r)/(V_0 m + 1)^2)) \\
& - 2 V_0^2 f k(r/(V_0 m + 1) - (V_0 m r)/(V_0 m + 1)^2)) + (T_0 V_0 \beta f \kappa p)/(A_1 s + 1). \tag{C.31}
\end{aligned}$$

We can solve for the unknown parameters, which are uniquely given by

$$V_0 = a_0, \tag{C.32}$$

$$c = -a_1/V_0, \tag{C.33}$$

$$f = d_1/V_0, \tag{C.34}$$

$$k = -(a_2 - V_0 c^2)/(V_0^2 f), \tag{C.35}$$

$$\beta = b_3/(V_0 \delta \kappa), \tag{C.36}$$

$$\begin{aligned}
p = & (3(a_3 + V_0 c^3 + V_0 d_2 k - 3 V_0^2 c f k)(a_3 \lambda + V_0 c^3 \lambda + V_0 d_2 k \lambda \\
& - 3 V_0^2 c f k \lambda))/(T_0 V_0 \beta \kappa(a_4 + 2 c a_3 + a_3 \delta + a_3 \kappa + 3 a_3 \lambda + V_0 c^4 + V_0^2 \beta c^3 \\
& + V_0 \beta a_3 + V_0 d_3 k - 3 V_0^3 f^2 k^2 + V_0 c^3 \delta + V_0 c^3 \kappa + 3 V_0 c^3 \lambda + V_0^2 \beta d_2 k \\
& - 2 V_0 c d_2 k + V_0 d_2 \delta k + V_0 d_2 k \kappa + 3 V_0 d_2 k \lambda - 3 V_0^3 \beta c f k - 3 V_0^2 c \delta f k \\
& - 3 V_0^2 c f k \kappa - 9 V_0^2 c f k \lambda)), \tag{C.37}
\end{aligned}$$

$$\begin{aligned}
r = & (18(c^3 d_2^6 + V_0^6 c^9 f^6 - 3 V_0^7 c^7 f^7 k + 6 V_0^2 c^5 d_2^4 f^2 - V_0^3 c^3 d_3^3 f^3 + 7 V_0^3 c^6 d_2^3 f^3 \\
& + 6 V_0^4 c^7 d_2^2 f^4 + 3 V_0^4 c^5 d_3^2 f^4 + 3 V_0^8 c^5 f^8 k^2 - V_0^9 c^3 f^9 k^3 + 3 V_0 c^4 d_2^5 f + 3 V_0^5 c^8 d_2 f^5 \\
& - 3 V_0^5 c^7 d_3 f^5 - 3 V_0 c^3 d_2^4 d_3 f + 3 V_0^2 c^3 d_2^2 d_3^2 f^2 + 3 V_0^6 c^3 d_2^2 f^6 k^2 - 6 V_0^4 c^6 d_2 d_3 f^4 \\
& - 6 V_0^6 c^6 d_2 f^6 k + 6 V_0^6 c^5 d_3 f^6 k - 6 V_0^2 c^4 d_2^3 d_3 f^2 + 3 V_0^3 c^4 d_2 d_3^2 f^3 - 9 V_0^3 c^5 d_2^2 d_3 f^3 \\
& - 3 V_0^3 c^3 d_2^4 f^3 k - 6 V_0^4 c^4 d_2^3 f^4 k - 9 V_0^5 c^5 d_2^2 f^5 k - 3 V_0^5 c^3 d_3^2 f^5 k + 3 V_0^7 c^4 d_2 f^7 k^2 \\
& - 3 V_0^7 c^3 d_3 f^7 k^2 + 6 V_0^4 c^3 d_2^2 d_3 f^4 k + 6 V_0^5 c^4 d_2 d_3 f^5 k))/(V_0(3 c d_2^3 + 3 V_0^3 c^4 f^3 \\
& - 3 V_0^5 f^5 k^2 + 3 V_0^2 d_2^2 f^2 k + 6 V_0 c^2 d_2^2 f - 2 V_0^2 c a_3 f^3 + 2 V_0^2 c d_4 f^2 - 3 V_0^3 d_3 f^3 k \\
& + 6 V_0^2 c^3 d_2 f^2 - 5 V_0^2 c^2 d_3 f^2 - 5 V_0 c d_2 d_3 f^2)), \tag{C.38}
\end{aligned}$$

$$\begin{aligned}
m = & -(3cd_2^3 - 3V_0^3c^4f^3 - 3V_0^5f^5k^2 + 6V_0^4c^2f^4k + 3V_0^2d_2^2f^2k - 2V_0^2ca_3f^3 \\
& + 2V_0^2cd_4f^2 - 3V_0^3d_3f^3k + V_0^2c^2d_3f^2 - 5V_0cd_2d_3f)/(V_0(3cd_2^3 + 3V_0^3c^4f^3 \\
& - 3V_0^5f^5k^2 + 3V_0^2d_2^2f^2k + 6V_0c^2d_2^2f - 2V_0^2ca_3f^3 + 2V_0^2cd_4f^2 - 3V_0^3d_3f^3k \\
& + 6V_0^2c^3d_2f^2 - 5V_0^2c^2d_3f^2 - 5V_0cd_2d_3f)),
\end{aligned} \tag{C.39}$$

$$\begin{aligned}
\gamma = & -(3c^2d_2^3 + cd_2^2d_3 + 6V_0^4c^3f^4k - 6V_0^5cf^5k^2 - 3V_0^4d_2f^4k^2 - 3V_0cd_3^2f \\
& + 3V_0d_3^2fk + 3V_0c^3d_2^2f - 2V_0^2c^2a_3f^3 + 3V_0^2c^4d_2f^2 + V_0^2c^3d_3f^2 + 2V_0^2c^2d_4f^2 \\
& - 9V_0^3cd_3f^3k - 3V_0^2d_2d_3f^2k + 2V_0cd_2d_4f + 9V_0^2cd_2^2f^2k + 6V_0^3c^2d_2f^3k \\
& - 2V_0cd_2a_3f^2 - 4V_0c^2d_2d_3f)/(3cd_2^3 + 3V_0^3c^4f^3 - 3V_0^5f^5k^2 + 3V_0^2d_2^2f^2k \\
& + 6V_0c^2d_2^2f - 2V_0^2ca_3f^3 + 2V_0^2cd_4f^2 - 3V_0^3d_3f^3k + 6V_0^2c^3d_2f^2 \\
& - 5V_0^2c^2d_3f^2 - 5V_0cd_2d_3f),
\end{aligned} \tag{C.40}$$

$$\begin{aligned}
s = & -(a_4 + 2ca_3 + a_3\delta + a_3\kappa + V_0c^4 + V_0^2\beta c^3 + V_0\beta a_3 + V_0d_3k - 3V_0^3f^2k^2 + V_0c^3\delta \\
& + V_0c^3\kappa + V_0^2\beta d_2k - 2V_0cd_2k + V_0d_2\delta k + V_0d_2k\kappa - 3V_0^3\beta cfk - 3V_0^2c\delta fk \\
& - 3V_0^2cfk\kappa)/(A_1a_4 + 2A_1ca_3 + A_1a_3\delta + A_1a_3\kappa + 3A_1a_3\lambda + A_1V_0c^4 + A_1V_0c^3\delta \\
& + A_1V_0c^3\kappa + 3A_1V_0c^3\lambda + A_1V_0^2\beta c^3 + A_1V_0\beta a_3 + A_1V_0d_3k - 3A_1V_0^3f^2k^2 \\
& - 2A_1V_0cd_2k + A_1V_0d_2\delta k + A_1V_0d_2k\kappa + 3A_1V_0d_2k\lambda + A_1V_0^2\beta d_2k \\
& - 3A_1V_0^3\beta cfk - 3A_1V_0^2c\delta fk - 3A_1V_0^2cfk\kappa - 9A_1V_0^2cfk\lambda).
\end{aligned} \tag{C.41}$$

We conclude that the parameters β , p , V_0 , s , f , r , m , k , γ , and c are globally identifiable.

C.4.3 Taylor series calculations: $\tau > 0$ (Case 1a extension)

We now consider the case where the time delay $\tau > 0$. As above, we assume V , C , and X are observed and seek to determine whether β , p , V_0 , s , f , r , m , k , γ , c , and now τ are identifiable. We assume that T_0 , δ , κ , A_1 , A_2 , λ_1 , λ_2 , and t^* are known (an extension of Case 1a in Table 4.2 where we assume that γ , r , c are also known). We construct Taylor expansions around three time-points t_0 , t_1 , and t_2 , where $t_0 = 0$, $\tau < t_1 < \tau + t^*$, and $\tau + t^* < t_2$.

Let us define the Taylor series coefficients

$$a_{0i} = \lim_{t \rightarrow t_0} V^{(i)}(t; \theta), \quad a_{1i} = \lim_{t \rightarrow t_1} V^{(i)}(t; \theta), \quad a_{2i} = \lim_{t \rightarrow t_2} V^{(i)}(t; \theta), \tag{C.42}$$

$$b_{0i} = \lim_{t \rightarrow t_0} C^{(i)}(t; \theta), \quad b_{1i} = \lim_{t \rightarrow t_1} C^{(i)}(t; \theta), \quad b_{2i} = \lim_{t \rightarrow t_2} C^{(i)}(t; \theta), \tag{C.43}$$

$$d_{0i} = \lim_{t \rightarrow t_0} X^{(i)}(t; \theta), \quad d_{1i} = \lim_{t \rightarrow t_1} X^{(i)}(t; \theta), \quad d_{2i} = \lim_{t \rightarrow t_2} X^{(i)}(t; \theta). \tag{C.44}$$

We first consider the expansions around $t = t_0$. As $\tau > 0$ we do not consider the sF term as

this is zero at $t = t_0$. The non-zero coefficients needed to determine model identifiability are

$$a_{00} = V_0, \quad (C.45)$$

$$a_{01} = -V_0 c, \quad (C.46)$$

$$a_{02} = -fkV_0^2 + V_0 c^2, \quad (C.47)$$

$$a_{03} = V_0 f(2V_0 ck + V_0 k(\gamma - (V_0 r)/(V_0 m + 1))) - V_0 c(c^2 - 2V_0 fk) + (T_0 V_0 \beta \kappa p), \quad (C.48)$$

$$b_{03} = V_0 \beta \delta \kappa, \quad (C.49)$$

$$d_{01} = fV_0, \quad (C.50)$$

$$d_{02} = -V_0 cf - V_0 f(\gamma - (V_0 r)/(V_0 m + 1)), \quad (C.51)$$

$$d_{03} = V_0 c(cf + f(\gamma - (V_0 r)/(V_0 m + 1)) - V_0 f(r/(V_0 m + 1) - (V_0 mr)/(V_0 m + 1)^2)) - V_0 f(V_0 fk - (\gamma - (V_0 r)/(V_0 m + 1))^2 + V_0 c(r/(V_0 m + 1) - (V_0 mr)/(V_0 m + 1)^2)), \quad (C.52)$$

$$\begin{aligned} d_{04} = & V_0 c(f(V_0 fk - (\gamma - (V_0 r)/(V_0 m + 1))^2 + V_0 c(r/(V_0 m + 1) - (V_0 mr)/(V_0 m + 1)^2)) - c(cf + f(\gamma - (V_0 r)/(V_0 m + 1)) - V_0 f(r/(V_0 m + 1) - (V_0 mr)/(V_0 m + 1)^2)) + V_0 f(fk + c(r/(V_0 m + 1) - (V_0 mr)/(V_0 m + 1)^2) \\ & + 2(r/(V_0 m + 1) - (V_0 mr)/(V_0 m + 1)^2)(\gamma - (V_0 r)/(V_0 m + 1)) - V_0 c((2mr)/(V_0 m + 1)^2 - (2V_0 m^2 r)/(V_0 m + 1)^3)) + V_0 c(2f(r/(V_0 m + 1) - (V_0 mr)/(V_0 m + 1)^2) - V_0 f((2mr)/(V_0 m + 1)^2 - (2V_0 m^2 r)/(V_0 m + 1)^3))) \\ & + V_0 f((\gamma - (V_0 r)/(V_0 m + 1))(V_0 fk - (\gamma - (V_0 r)/(V_0 m + 1))^2 + V_0 c(r/(V_0 m + 1) - (V_0 mr)/(V_0 m + 1)^2)) - (V_0 mr)/(V_0 m + 1)^2)) + V_0 c(fk + c(r/(V_0 m + 1) - (V_0 mr)/(V_0 m + 1)^2) \\ & + 2(r/(V_0 m + 1) - (V_0 mr)/(V_0 m + 1)^2)(\gamma - (V_0 r)/(V_0 m + 1)) - V_0 c((2mr)/(V_0 m + 1)^2 - (2V_0 m^2 r)/(V_0 m + 1)^3)) + V_0 k(cf + f(\gamma - (V_0 r)/(V_0 m + 1)) - V_0 f(r/(V_0 m + 1) - (V_0 mr)/(V_0 m + 1)^2)) \\ & - 2V_0^2 fk(r/(V_0 m + 1) - (V_0 mr)/(V_0 m + 1)^2)) + T_0 V_0 \beta f \kappa p. \end{aligned} \quad (C.53)$$

The unknown parameters are uniquely given by

$$V_0 = a_{00}, \quad (\text{C.54})$$

$$c = -a_{01}/V_0, \quad (\text{C.55})$$

$$f = d_{01}/V_0, \quad (\text{C.56})$$

$$k = -(a_{02} - V_0 c^2)/(V_0^2 f), \quad (\text{C.57})$$

$$\beta = b_{03}/(V_0 \delta \kappa), \quad (\text{C.58})$$

$$p = (a_{03} + V_0 c^3 + V_0 d_{02} k - 3V_0^2 c f k)/(T_0 V_0 \beta \kappa) \quad (\text{C.59})$$

$$\begin{aligned} r = & (18(c^3 d_{02}^6 + V_0^6 c^9 f^6 - 3V_0^7 c^7 f^7 k + 6V_0^2 c^5 d_{02}^4 f^2 - V_0^3 c^3 d_{03}^3 f^3 \\ & + 7V_0^3 c^6 d_{02}^3 f^3 + 6V_0^4 c^7 d_{02}^2 f^4 + 3V_0^4 c^5 d_{03}^2 f^4 + 3V_0^8 c^5 f^8 k^2 - V_0^9 c^3 f^9 k^3 \\ & + 3V_0 c^4 d_{02}^5 f + 3V_0^5 c^8 d_{02} f^5 - 3V_0^5 c^7 d_{03} f^5 - 3V_0 c^3 d_{02}^4 d_{03} f + 3V_0^2 c^3 d_{02}^2 d_{03}^2 f^2 \\ & + 3V_0^6 c^3 d_{02}^2 f^6 k^2 - 6V_0^4 c^6 d_{02} d_{03} f^4 - 6V_0^6 c^6 d_{02} f^6 k + 6V_0^6 c^5 d_{03} f^6 k \\ & - 6V_0^2 c^4 d_{02}^3 d_{03} f^2 + 3V_0^3 c^4 d_{02} d_{03}^2 f^3 - 9V_0^3 c^5 d_{02}^2 d_{03} f^3 - 3V_0^3 c^3 d_{02}^4 f^3 k \\ & - 6V_0^4 c^4 d_{02}^3 f^4 k - 9V_0^5 c^5 d_{02}^2 f^5 k - 3V_0^5 c^3 d_{03}^2 f^5 k + 3V_0^7 c^4 d_{02} f^7 k^2 - 3V_0^7 c^3 d_{03} f^7 k^2 \\ & + 6V_0^4 c^3 d_{02}^2 d_{03} f^4 k + 6V_0^5 c^4 d_{02} d_{03} f^5 k)/(V_0(3cd_{02}^3 + 3V_0^3 c^4 f^3 - 3V_0^5 f^5 k^2 \\ & + 3V_0^2 d_{02}^2 f^2 k + 6V_0 c^2 d_{02}^2 f - 2V_0^2 ca_{03} f^3 + 2V_0^2 cd_{04} f^2 - 3V_0^3 d_{03} f^3 k + 6V_0^2 c^3 d_{02} f^2 \\ & - 5V_0^2 c^2 d_{03} f^2 - 5V_0 cd_{02} d_{03} f)^2), \end{aligned} \quad (\text{C.60})$$

$$\begin{aligned} m = & -(3cd_{02}^3 - 3V_0^3 c^4 f^3 - 3V_0^5 f^5 k^2 + 6V_0^4 c^2 f^4 k + 3V_0^2 d_{02}^2 f^2 k - 2V_0^2 ca_{03} f^3 \\ & + 2V_0^2 cd_{04} f^2 - 3V_0^3 d_{03} f^3 k + V_0^2 c^2 d_{03} f^2 - 5V_0 cd_{02} d_{03} f)/(V_0(3cd_{02}^3 + 3V_0^3 c^4 f^3 \\ & - 3V_0^5 f^5 k^2 + 3V_0^2 d_{02}^2 f^2 k + 6V_0 c^2 d_{02}^2 f - 2V_0^2 ca_{03} f^3 + 2V_0^2 cd_{04} f^2 - 3V_0^3 d_{03} f^3 k \\ & + 6V_0^2 c^3 d_{02} f^2 - 5V_0^2 c^2 d_{03} f^2 - 5V_0 cd_{02} d_{03} f)), \end{aligned} \quad (\text{C.61})$$

$$\begin{aligned} \gamma = & -(3c^2 d_{02}^3 + cd_{02}^2 d_{03} + 6V_0^4 c^3 f^4 k - 6V_0^5 c f^5 k^2 - 3V_0^4 d_{02} f^4 k^2 - 3V_0 cd_{03}^2 f \\ & + 3V_0 d_{02}^3 f k + 3V_0 c^3 d_{02}^2 f - 2V_0^2 c^2 a_{03} f^3 + 3V_0^2 c^4 d_{02} f^2 + V_0^2 c^3 d_{03} f^2 \\ & + 2V_0^2 c^2 d_{04} f^2 - 9V_0^3 cd_{03} f^3 k - 3V_0^2 d_{02} d_{03} f^2 k + 2V_0 cd_{02} d_{04} f + 9V_0^2 cd_{02}^2 f^2 k \\ & + 6V_0^3 c^2 d_{02} f^3 k - 2V_0 cd_{02} a_{03} f^2 - 4V_0 c^2 d_{02} d_{03} f)/(3cd_{02}^3 + 3V_0^3 c^4 f^3 - 3V_0^5 f^5 k^2 \\ & + 3V_0^2 d_{02}^2 f^2 k + 6V_0 c^2 d_{02}^2 f - 2V_0^2 ca_{03} f^3 + 2V_0^2 cd_{04} f^2 - 3V_0^3 d_{03} f^3 k + 6V_0^2 c^3 d_{02} f^2 \\ & - 5V_0^2 c^2 d_{03} f^2 - 5V_0 cd_{02} d_{03} f). \end{aligned} \quad (\text{C.62})$$

We conclude that we can identify the parameters $V_0, c, f, k, \beta, p, r, m$, and γ . It remains to establish whether we can identify the parameters s and τ . We define $V_j = V(t_j)$ and

$I_j = I(t_j)$ for $j = 0, 1, 2$. Considering the Taylor expansions at $t = t_1$ and $t = t_2$ we have that

$$a_{11} = \frac{pI_1}{1 + sA_1e^{\lambda_1(t_1-\tau)}} - cV_1, \quad (\text{C.63})$$

$$b_{11} = \frac{\delta I_1}{T_0}, \quad (\text{C.64})$$

$$a_{21} = \frac{pI_2}{1 + sA_2e^{\lambda_2(t_2-\tau)}} - cV_2, \quad (\text{C.65})$$

$$b_{21} = \frac{\delta I_2}{T_0}. \quad (\text{C.66})$$

Manipulating (C.63)–(C.66) using the substitutions $I_1 = T_0b_{11}/\delta$ and $I_2 = T_0b_{21}/\delta$ we get that

$$se^{-\lambda_1\tau} = \frac{1}{A_1e^{\lambda_1t_1}} \left[\frac{pT_0b_{11}}{\delta(a_1 + cV_1)} - 1 \right] = Y_1, \quad (\text{C.67})$$

$$se^{-\lambda_2\tau} = \frac{1}{A_2e^{\lambda_2t_2}} \left[\frac{pT_0b_{21}}{\delta(a_2 + cV_2)} - 1 \right] = Y_2, \quad (\text{C.68})$$

where we define Y_1 and Y_2 to simplify the following expressions. From (C.67) and (C.68) we deduce that

$$\tau = \frac{1}{\lambda_2 - \lambda_1} \log \left(\frac{Y_1}{Y_2} \right), \quad (\text{C.69})$$

$$s = Y_1^{\frac{\lambda_2}{\lambda_2 - \lambda_1}} Y_2^{\frac{\lambda_1}{\lambda_1 - \lambda_2}}. \quad (\text{C.70})$$

Hence we can identify the parameters s and τ given data at times t_1 and t_2 where $\tau < t_1 < \tau + t^*$ and $\tau + t^* < t_2$. We conclude that the parameters β , p , V_0 , s , f , r , m , k , γ , c , and τ are globally identifiable if we observe data until time t where $t > \tau + t^*$. This structural identifiability analysis was performed, in part, with the GenSSI software package [70].

C.5 Practical identifiability analysis: profile likelihood curves

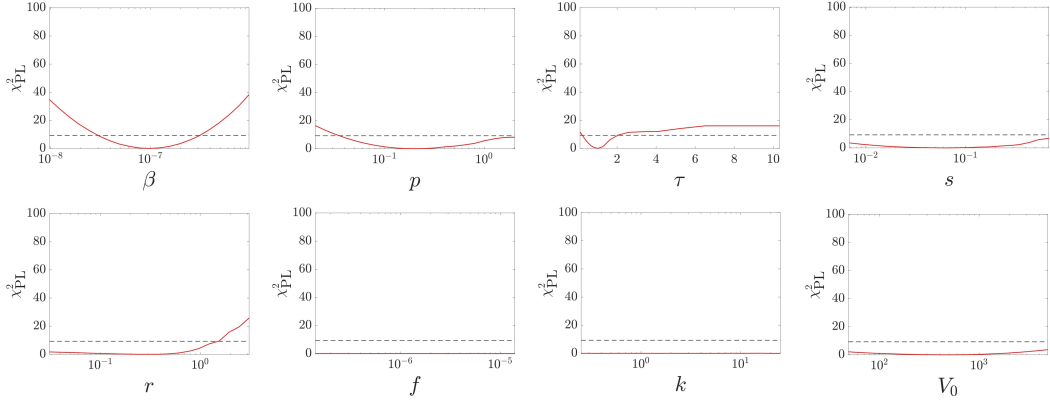


Figure C.2: Profile likelihood (χ^2_{PL}) curves for each fitting parameter of the IR model, when data for V and C are observed (Case 2). Thresholds for confidence intervals given by horizontal dashed lines.

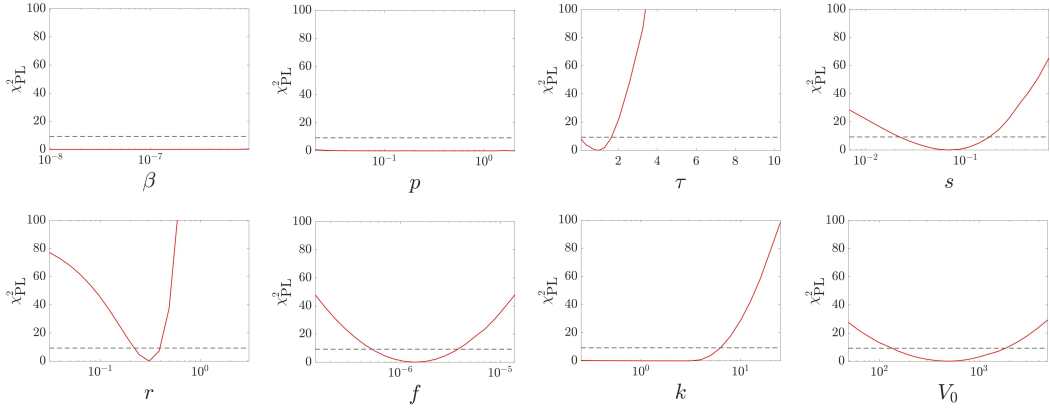


Figure C.3: Profile likelihood (χ^2_{PL}) curves for each fitting parameter of the IR model, when data for V and X are observed (Case 3). Thresholds for confidence intervals given by horizontal dashed lines.

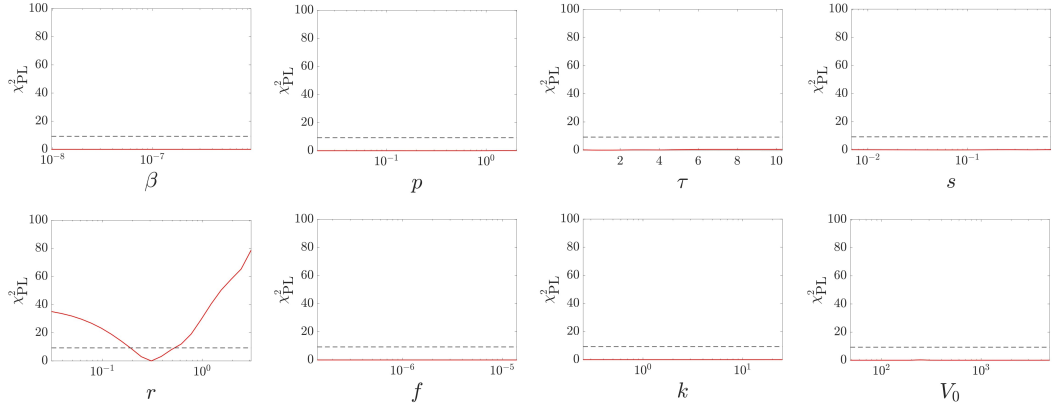


Figure C.4: Profile likelihood (χ^2_{PL}) curves for each fitting parameter of the IR model, when data for C and X are observed (Case 4). Thresholds for confidence intervals given by horizontal dashed lines.

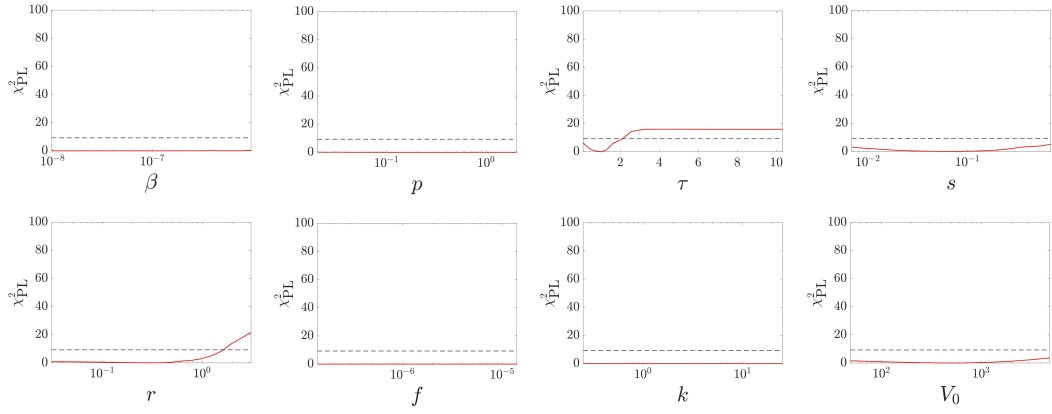


Figure C.5: Profile likelihood (χ^2_{PL}) curves for each fitting parameter of the IR model, when data for V are observed (Case 5). Thresholds for confidence intervals given by horizontal dashed lines.

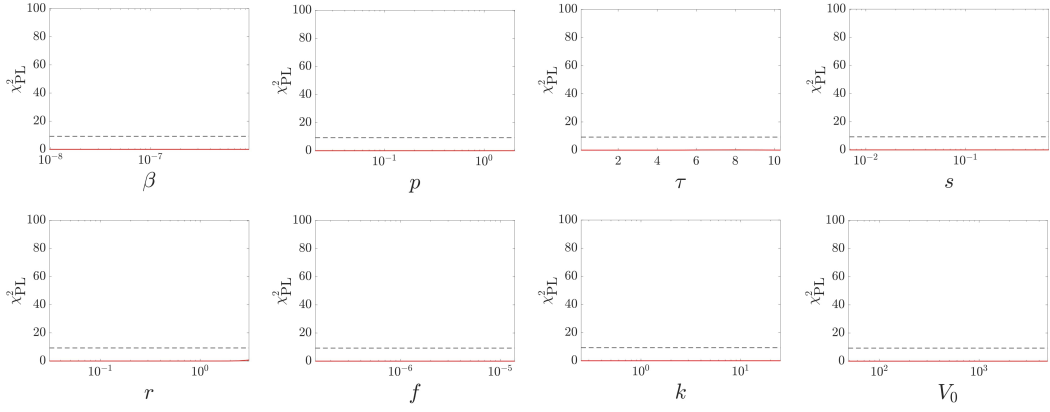


Figure C.6: Profile likelihood (χ^2_{PL}) curves for each fitting parameter of the IR model, when data for C are observed (Case 6). Thresholds for confidence intervals given by horizontal dashed lines.

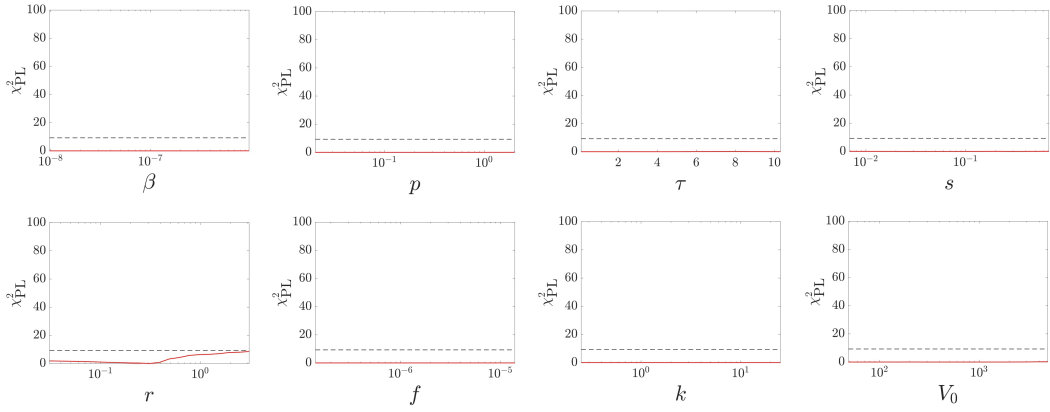


Figure C.7: Profile likelihood (χ^2_{PL}) curves for each fitting parameter of the IR model, when data for X are observed (Case 7). Thresholds for confidence intervals given by horizontal dashed lines.

Appendix D

Appendix to Chapter 5

D.1 Parameter estimation

Using the same methodology as in Chapters 3 and 4, we estimate model parameters using Markov chain Monte Carlo (MCMC) with the Metropolis-Hastings algorithm, as well as maximum likelihood estimation (MLE). We assume that the measured viral titer levels are lognormally distributed around their true values, with standard deviation $0.3 \log_{10}$ (as in Hadjichrysanthou et al. [147]) and that the innate immune response and symptom scores are normally distributed around their true values, with standard deviations of 12 and 0.7 respectively (these values determined by estimating the level of noise in the data using MLE). The likelihood function for parameter estimation is given by

$$\mathcal{L}(\theta; y) = \frac{1}{Z} \prod_{i=1}^M \prod_{j=1}^{N_i} \exp \left(\frac{-[z_i(t_j; \theta) - y_{ij}]^2}{2\sigma_i^2} \right), \quad (\text{D.1})$$

where $z_i(t_j, \theta)$ represent model outputs generated with parameters θ , y_{ij} represent data, and σ_i are noise scaling parameters. The subscripts $i = 1, \dots, M$ relate to the observables (e.g. viral load, fraction of dead cells, etc.) and the subscripts $j = 1, \dots, N_i$ relate to the output timepoints.

We fix the normalisation constant $Z = 1$ without loss of generality. When considering data for the viral load, as in Handel et al. [153], $z_i(t_j, \theta)$ and y_{ij} are the base 10 logarithms of the model outputs and data. If we assume normally distributed noise, then we can also define the weighted sum of squared residuals (SSR) via the cost function $J(\theta; y)$ where

$$J(\theta; y) = \sum_{i=1}^M \sum_{j=1}^{N_i} \left(\frac{-[z_i(t_j; \theta) - y_{ij}]^2}{2\sigma_i^2} \right). \quad (\text{D.2})$$

The MLE parameter values are defined in terms of the cost function as follows:

$$\hat{\theta} = \arg \min_{\theta \in \Theta} J(\theta; y), \quad (\text{D.3})$$

where Θ denotes the set of feasible parameter values. We solve the least squares optimisation problem (D.3) using MATLAB's `fminsearch` function.

D.2 Scatter plots of ω and η

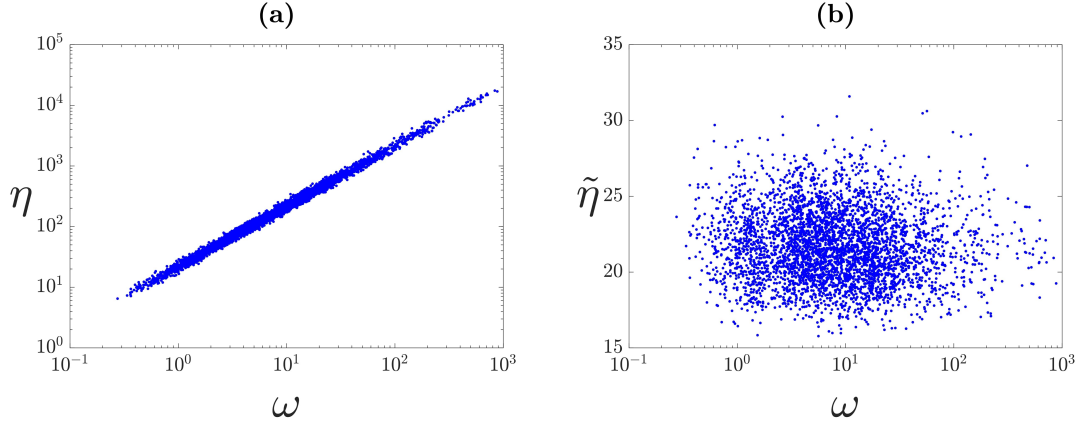


Figure D.1: Scatter plots of posterior distributions of ω and η (a), and ω and $\tilde{\eta} = \eta/\omega$ (b), when the SD model is fitted to data on the viral load, innate immune response (cytokines), and symptom score (Figure 5.2).

D.3 Modulation of ω and η

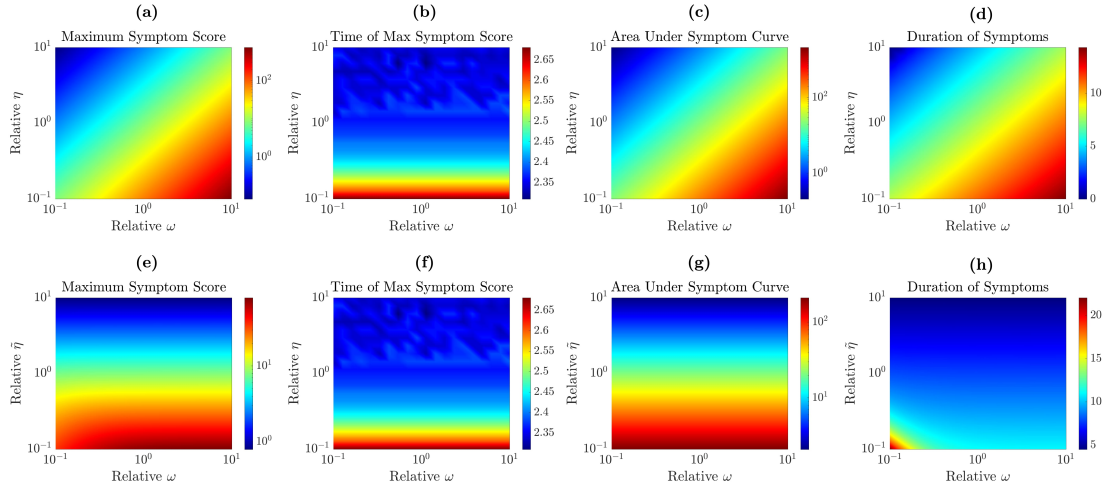


Figure D.2: Infection-related quantities of the SD model as parameters ω and η are varied (a)-(d), and ω and $\tilde{\eta} = \eta/\omega$ are varied (e)-(h). Relative parameter values refer to those normalised by those given in Table 5.2.

D.4 Synthetic prior patient data

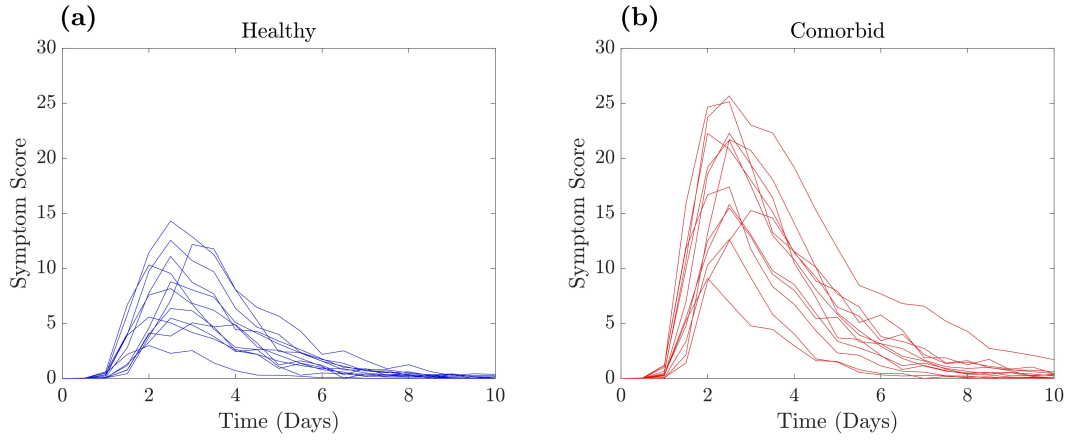


Figure D.3: The synthetic data relating to 12 previously observed healthy and comorbid patients used in Figures 5.5–5.8. Data generated from the SD model using parameters given in Tables 5.1 and 5.2, with normally distributed noise added.

D.5 Deviance information criterion (DIC)

Healthy		Comorbid	
Figure	DIC	Figure	DIC
5.4 (a)	67660	5.4 (b)	92510
5.4 (c)	9091	5.4 (d)	37230
5.4 (e)	311.4	5.4 (f)	43.95
5.5 (a)	739.5	5.5 (b)	1543
5.5 (c)	90.81	5.5 (d)	129.7
5.5 (e)	17.76	5.5 (f)	29.20
5.6 (a)	9.106	5.6 (b)	17.10
5.6 (c)	33.05	5.6 (d)	171.0
5.6 (e)	5.740	5.6 (f)	206.5
5.7 (a)	1572	5.7 (b)	2879
5.7 (c)	121.7	5.7 (d)	267.7
5.7 (e)	66.6	5.7 (f)	149.4

Table D.1: The DIC values for each forecast in Chapter 5.

Appendix E

MATLAB code

MATLAB code required to reproduce the results presented in this thesis is available at <https://doi.org/10.5287/bodleian:o0QPRkA8P>.

Appendix F

The risk of SARS-CoV-2 outbreaks in low prevalence settings following the removal of travel restrictions

Chapter 2 was published as a journal article in *Communications Medicine* (<https://doi.org/10.1101/2021.05.21.21257589>).

Each of the authors contributed to the following facets of the published work:

- Conceptualisation: RS-P, RNT, LD.
- Methodology: RS-P, RNT.
- Investigation: RS-P.
- Supervision: RNT, HMB.
- Writing – original draft: RS-P, RNT.
- Writing – review and editing: RS-P, RNT, HMB, LD.

The risk of SARS-CoV-2 outbreaks in low prevalence settings following the removal of travel restrictions

Rahil Sachak-Patwa¹, Helen M. Byrne¹, Louise Dyson^{2,3} & Robin N. Thompson^{2,3}✉

Abstract

Background Countries around the world have introduced travel restrictions to reduce SARS-CoV-2 transmission. As vaccines are gradually rolled out, attention has turned to when travel restrictions and other non-pharmaceutical interventions (NPIs) can be relaxed.

Methods Using SARS-CoV-2 as a case study, we develop a mathematical branching process model to assess the risk that, following the removal of NPIs, cases arriving in low prevalence settings initiate a local outbreak. Our model accounts for changes in background population immunity due to vaccination. We consider two locations with low prevalence in which the vaccine rollout has progressed quickly – specifically, the Isle of Man (a British crown dependency in the Irish Sea) and the country of Israel.

Results We show that the outbreak risk is unlikely to be eliminated completely when travel restrictions and other NPIs are removed. This general result is the most important finding of this study, rather than exact quantitative outbreak risk estimates in different locations. It holds even once vaccine programmes are completed. Key factors underlying this result are the potential for transmission even following vaccination, incomplete vaccine uptake, and the recent emergence of SARS-CoV-2 variants with increased transmissibility.

Conclusions Combined, the factors described above suggest that, when travel restrictions are relaxed, it may still be necessary to implement surveillance of incoming passengers to identify infected individuals quickly. This measure, as well as tracing and testing (and/or isolating) contacts of detected infected passengers, remains useful to suppress potential outbreaks while global case numbers are high.

Plain language summary

The effectiveness of public health measures against COVID-19 has varied between countries, with some experiencing many infections and others containing transmission successfully. As vaccines are deployed, an important challenge is deciding when to relax measures. Here, we consider locations with few cases, and investigate whether vaccination can ever eliminate the risk of COVID-19 outbreaks completely, allowing measures to be removed risk-free. Using a mathematical model, we demonstrate that there is still a risk that imported cases initiate outbreaks when measures are removed, even if most of the population is fully vaccinated. This highlights the need for continued vigilance in low prevalence settings to prevent imported cases leading to local transmission. Until case numbers are reduced globally, so that SARS-CoV-2 spread between countries is less likely, the risk of outbreaks in low prevalence settings will remain.

¹Mathematical Institute, University of Oxford, Oxford, UK. ²Mathematics Institute, University of Warwick, Coventry, UK. ³Zeeman Institute for Systems Biology and Infectious Disease Epidemiology Research, University of Warwick, Coventry, UK. ✉email: robin.n.thompson@warwick.ac.uk

Combinations of non-pharmaceutical interventions (NPIs) have been introduced worldwide to counter the COVID-19 pandemic^{1–3}. These measures include travel restrictions^{4,5} and a range of other NPIs intended to reduce the number of contacts between individuals⁶. The development and deployment of vaccines have also lowered transmission^{7–10} and reduced the number of individuals experiencing clinical symptoms or severe disease once infected^{11–13}.

Effective NPIs have led to low levels of transmission in some locations. Australia and New Zealand have often been cited as examples of countries that have implemented NPIs effectively, with travel restrictions and quarantine of inbound travellers combined with short-term lockdowns and contact tracing to identify infected contacts whenever cases have been discovered^{14–17}. Rigorous interventions targeted at inbound travellers in low prevalence settings reflect the fact that imported cases can contribute substantially to local incidence¹⁸, and potentially initiate outbreaks with substantial local transmission^{19,20}. At the time of writing (1st May 2021), Israel is the country that has vaccinated the largest proportion of its citizens, and the vaccination campaign there has been credited with reducing transmission⁸, prompting some NPIs to be removed.

Despite the success of both NPIs and vaccines, the current overall picture is complicated. Vaccines do not prevent transmission entirely^{7,21–24}, and vaccine uptake is incomplete, particularly in some ethnic groups and in underserved communities²⁵. Current tentative estimates suggest that first doses of the Pfizer and AstraZeneca SARS-CoV-2 vaccines are around 60% effective at preventing infection, with second doses leading to 65–85% effectiveness against infection²¹. Furthermore, the appearance of novel SARS-CoV-2 variants makes eliminating transmission more challenging. For example, the Alpha variant (B.1.1.7, or VOC 202012/01) that first appeared in the United Kingdom in late 2020 has been found to be more transmissible than the SARS-CoV-2 virus that originally emerged in China²⁶. Although public health measures have had some successes, these concerns raise a key question: Can NPIs such as travel restrictions be removed without any risk in low prevalence settings where vaccines have been distributed widely, or might outbreaks occur initiated by SARS-CoV-2 reimportations from elsewhere?

Epidemiological models are often used to assess the risk of outbreaks in scenarios in which the potential for pathogen transmission is not changing. According to the mathematical theory of branching processes, the probability that cases introduced into a new host population generate an outbreak driven by sustained local transmission is then greater than zero whenever $R > 1$, where R is the reproduction number of the pathogen. For the most common branching process model, used to represent the spread of directly transmitted pathogens, the probability that I_0 introduced infectious cases initiate an outbreak is given by

$$\text{Prob(outbreak)} = 1 - \left(\frac{1}{R}\right)^{I_0}. \quad (1)$$

This expression had been used to assess the risks of outbreaks of pathogens including the Ebola virus^{27–29}, before being applied early in the COVID-19 pandemic to assess outbreak risks outside China²⁰. Eq. (1) is based on a simple transmission model in which individuals are assumed to mix homogeneously and infected individuals are infectious for exponentially distributed time periods. Equation (1) also involves an assumption that pathogen transmissibility is fixed at its current level. In other words, the value of R is implicitly assumed to remain constant over the initial phase of the potential outbreak. With a background of rapidly changing population immunity due to vaccination acting to reduce transmission, this assumption may not be accurate. To assess the risk of outbreaks if NPIs are removed

during an ongoing vaccination campaign using branching process models, standard epidemiological modelling theory must be extended to account for temporally changing population immunity.

Here, we use branching processes to investigate whether an introduced case is likely to initiate an outbreak, accounting fully for temporal changes in population immunity due to an ongoing vaccination campaign. We use four metrics to assess the risk of an outbreak and consider two examples of vaccination campaigns from around the world, from the Isle of Man and the country of Israel. In both locations, vaccination is progressing quickly and prevalence is currently low. Given the relatively low numbers of cases in these locations during the pandemic, there is also likely to be a background of limited immunity from previous infections. We assess the risk of outbreaks in these places when travel restrictions and other NPIs are removed, considering scenarios in which NPIs are removed at different stages of the vaccination campaigns. Our goal is not to develop a detailed epidemiological model for estimating the outbreak risk precisely. Instead, we use a simple model to show that, even when all vaccines have been deployed in low prevalence settings (i.e. a maximum vaccination proportion ν of individuals have been vaccinated, where the precise value of ν varies between locations), the combination of incomplete vaccine uptake, imperfect vaccination, and variants of concern means that the risk of outbreaks due to imported cases will not be eliminated completely when NPIs are removed. This highlights the need for careful monitoring of imported cases until global prevalence is reduced to low levels. Until vaccines are rolled out worldwide, there is still a risk of local transmission arising in low prevalence settings initiated by importations from elsewhere.

Methods

Epidemiological model. We performed our main analyses using a stochastic branching process model that describes virus transmission in the initial stages of a potential outbreak. In this model, following the arrival of a case in the local population, each infected individual generates new infections at rate $\beta(1 - \Lambda(t))$ and infected individuals have a mean infectious period of $1/\mu$ days (Fig. 1a). The times between successive events in the model are assumed to follow exponential distributions. The function $\Lambda(t)$ reflects the extent to which transmission has been reduced by vaccination, where a value of $\Lambda(t) = 0$ corresponds to an entirely unvaccinated population. We consider scenarios in which travel restrictions and other NPIs are removed, and transmission is only limited by vaccine-acquired immunity.

To model an ongoing vaccination campaign, we set $\Lambda(t) = \eta_1 V_1(t - \alpha) + \eta_2 V_2(t - \alpha)$, where $V_1(t)$ is the proportion of individuals in the population who have received a single vaccine dose at time t and $V_2(t)$ is the proportion of individuals in the population who have received two vaccine doses at a time t (Fig. 1b). The parameters η_1 and η_2 reflect the effectiveness of the vaccine at preventing infection after one and two effective doses, respectively (specifically, η_1 and η_2 represent the multiplicative reduction in susceptibility following one or two effective doses, compared to an unvaccinated individual, so that $0 \leq \eta_1 \leq \eta_2 \leq 1$), and the parameter α represents the delay between a vaccine dose being administered and being effective in the recipient. In our main analyses, since we are modelling relatively low prevalence settings, we do not consider immunity due to prior infections, although we present a supplementary analysis in which we demonstrate the robustness of our results to this assumption.

Under this model, the time-dependent reproduction number, accounting for any vaccines that have been administered and are

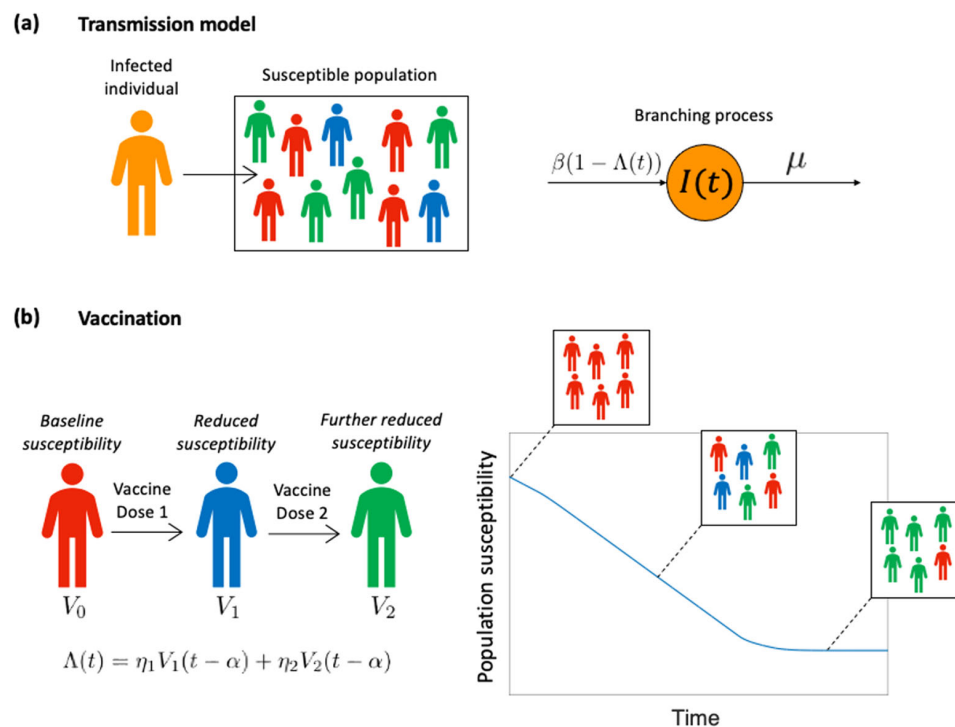


Fig. 1 The epidemiological model used in our analyses. **a** Following the introduction of an infected individual into the host population, local transmission may happen with each infected individual generating new infections at rate $\beta(1 - \Lambda(t))$, where $\Lambda(t)$ reflects the vaccination coverage in the local population. The mean infectious period of infected individuals is $1/\mu$ days, and the rates shown are per infected individual. **b** The vaccination process is modelled by setting $\Lambda(t)$ according to the proportion of individuals in the population who have been vaccinated with one (V_1) or two (V_2) doses. The first vaccine dose is assumed to have effectiveness η_1 and the second vaccine dose has effectiveness η_2 , where the values of η_1 and η_2 are the multiplicative reductions in the susceptibility of individuals who have been vaccinated with one or two vaccine doses, respectively (compared to an unvaccinated host). Vaccine doses are effective α days after they are administered. This leads to declining population susceptibility as a vaccine is rolled out across the population.

effective at time t , is given by

$$R_V(t) = \frac{\beta(1 - \Lambda(t))}{\mu}. \quad (2)$$

The value of $R_V(t)$ represents the expected number of secondary infections generated by a single infected individual in the population at time t , under the assumption that no further vaccinations take place in the future. It is therefore representative of the “instantaneous” transmissibility at time t . Indeed, $R_V(t)$ is sometimes referred to as the instantaneous reproduction number^{30–33}. In the absence of vaccination, so that $\Lambda(t) = 0$, then $R_V(t)$ is equal to the basic reproduction number, R_0 .

Vaccination data. In the model, $V_1(t)$ and $V_2(t)$ were set based on vaccination data from the location under consideration (either the Isle of Man or Israel). Data describing the proportion of the total population who had received one or two vaccine doses were available for the periods up until 11th April 2021 (for the Isle of Man³⁴) and 21st April 2021 (for Israel³⁵).

To explore how the risk of outbreaks is likely to change in the future, we projected the vaccine rollout forwards beyond these dates in the following way. We considered the total population size of the location under consideration (denoted N), as well as the numbers of individuals ($N_1(t)$ and $N_2(t)$) vaccinated with one or two doses, so that $V_1(t) = N_1(t)/N$ and $V_2(t) = N_2(t)/N$. We assumed that a constant number of vaccine doses are available each day in the future (denoted D), and that there is a target period of τ days between each vaccine dose. On any day in the future, each available dose is assigned to an individual who has been vaccinated with their first dose at least τ days ago, with

remaining doses then assigned to unvaccinated individuals. Resulting values of $N_1(t)$ and $N_2(t)$ were then converted to corresponding values of $V_1(t)$ and $V_2(t)$. To reflect the fact that vaccine uptake is imperfect, and some population groups are not vaccinated, we assumed that a maximum proportion ν of the population can ever be vaccinated. Consequently, once the vaccination programme is completed, then $\Lambda(t) = \eta_2 \nu$. The value of ν varies between locations, and depends on factors such as the age distribution of the host population. Since there is uncertainty in the vaccine uptake going forwards, we conducted supplementary analyses in which we considered a range of different values of ν . Values of the model parameters used in the analyses in the main text for the Isle of Man and Israel are shown in Table 1.

Outbreak risk metrics. We used four metrics to assess the risk that an infected individual, introduced into the population at time t , initiates an outbreak driven by sustained local transmission. The values of each of these metrics vary temporally. Each metric is used to estimate the risk that an imported case initiates an outbreak under the assumption that all travel restrictions and other NPIs are removed (in other words, pathogen transmissibility at time t is determined by $R_V(t)$, which represents an adjustment to the basic reproduction number R_0 due to vaccine-acquired immunity, as described above). An overview of the four metrics explored is provided here; additional details are available in the Supplementary Information (Supplementary Methods).

The first metric we considered is the *Instantaneous Outbreak Risk (IOR)*. This quantity represents the expression shown in Eq. (1), with $R = R_V(t)$ and $I_0 = 1$. The IOR reflects the risk of an

Table 1 Default parameter values used in our analyses.

Parameter	Description	Value	Source
β	Transmission rate	Set so that $R_0 = \frac{\beta}{\mu}$ takes a prescribed value	Two scenarios were considered, with mean R_0 values of 3 (similar to the original SARS-CoV-2 variant ⁶⁵) and 5 (similar to the Alpha variant (B.1.1.7) ²⁶)
$1/\mu$	Duration of infectiousness	5 days	⁶⁶
α	Delay between a vaccine dose being administered and being effective	14 days	⁶⁷
η_1	Multiplicative reduction in susceptibility of an individual vaccinated with a single effective dose (compared to an unvaccinated host)	0.6	²¹
η_2	Multiplicative reduction in susceptibility of an individual vaccinated with two effective doses (compared to an unvaccinated host)	0.85	²¹
ν	Vaccine uptake	0.8 (Isle of Man) and 0.7 (Israel). Other values are considered in the Supplementary Information	⁶⁸
N	Total population size	84,500 (Isle of Man) and 8,772,800 (Israel)	⁶⁹
D	Number of vaccine doses available each day (in future projections)	1000 (Isle of Man) and 56,000 (Israel)	Average number of doses administered each day in the previous 30 days (up to 11th April 2021 for Isle of Man ³⁴ and 21st April 2021 for Israel ³⁵)
τ	Target period between vaccine doses (in future projections)	84 days (Isle of Man) and 21 days (Israel)	^{70,71}

outbreak occurring starting from a single infected individual at time t , but under the assumption that the vaccine rollout does not continue after time t so that pathogen transmissibility is unchanged. While the IOR is straightforward to calculate, it does not reflect changing population immunity due to vaccination over the initial phase of the potential outbreak. This standard metric is often used to assess the risk of outbreaks in scenarios where pathogen transmissibility does not vary temporally^{20,27–29,36,37}.

The second metric is the *Case Outbreak Risk (COR)*. The COR is an extension of the IOR, accounting for changes in population susceptibility due to vaccination over the initial phase of the potential outbreak. The COR has previously been used to assess outbreak risks using branching processes for models in which pathogen transmission varies periodically^{38–40}. Its calculation involves solving the differential equation

$$\frac{dq(t)}{dt} = \beta(1 - \Lambda(t))q(t)(1 - q(t)) + \mu(q(t) - 1). \quad (3)$$

The COR at time t is then given by $1 - q(t)$, where $q(t)$ represents the probability of an outbreak failing to occur (i.e. ultimate extinction of the virus in the branching process model⁴¹) starting from a single infected individual introduced into the population at time t . Further details, including the derivation of Eq. (3), are provided in the Supplementary Information (Supplementary Methods).

The third metric we considered is the *Simulated Outbreak Risk (SOR)*. The SOR involves repeated simulation of the branching process model, using the direct version of the Gillespie stochastic simulation algorithm⁴² adapted to account for temporally varying pathogen transmissibility (due to changing population immunity as vaccines are deployed). Simulations were run starting with a single infected individual introduced into the population at time t . The SOR then corresponds to the proportion of simulations in which a local outbreak occurs; an outbreak is said to occur if the total number of individuals infected simultaneously exceeds a pre-defined threshold value, M . In our analyses, we set $M = 100$.

Finally, we considered the *Numerical Outbreak Risk (NOR)*. The NOR is analogous to the SOR, but with the advantage that it does not require large numbers of model simulations to be run. The NOR, therefore, also represents the risk that a single infected individual introduced into the population at time t initiates an outbreak in which at least $M = 100$ individuals are ever infected simultaneously.

Results

As described in the Methods, we first generated projections of the number of vaccinated individuals in the future for the Isle of Man (Fig. 2a) and Israel (Fig. 2d), based on past vaccination data in those locations. To explore the impact of vaccination on virus transmission, we calculated the time-dependent reproduction number ($R_V(t)$; Eq. (2)) throughout the vaccination campaigns. We considered two different scenarios. In the first, we set the mean value of R_0 (i.e. the reproduction number in the absence of vaccination) equal to 3, as was the situation early in the COVID-19 pandemic (Fig. 2b,e). In the second scenario, we set the mean value of R_0 equal to 5 (Fig. 2c,f) to reflect the fact that currently circulating SARS-CoV-2 variants are more transmissible than the original virus²⁶. This second scenario is therefore likely to be a more realistic reflection of the current and future risk.

We then calculated the values of the four different outbreak risk metrics throughout the period considered (18th December 2020 to 20th August 2021) based on these vaccination projections (Fig. 3). This involves a scenario in which NPIs are removed entirely, so that $R_V(t)$ is not reduced by interventions other than vaccination. These metrics then reflect the risk that a single case first entering the population at the date of introduction shown initiates an outbreak driven by sustained local transmission, given that no NPIs are in place. For each metric, and each time during the vaccination programme, we calculated the outbreak risk by integrating over the full distribution for $R_V(t)$ shown in Fig. 2. In practice, the precise value of $R_V(t)$ is not known exactly at any time: the purpose of this integration (which is over values of $R_V(t)$, rather than values of t) is to account for this uncertainty in outbreak risk estimates. The resulting outbreak risk therefore

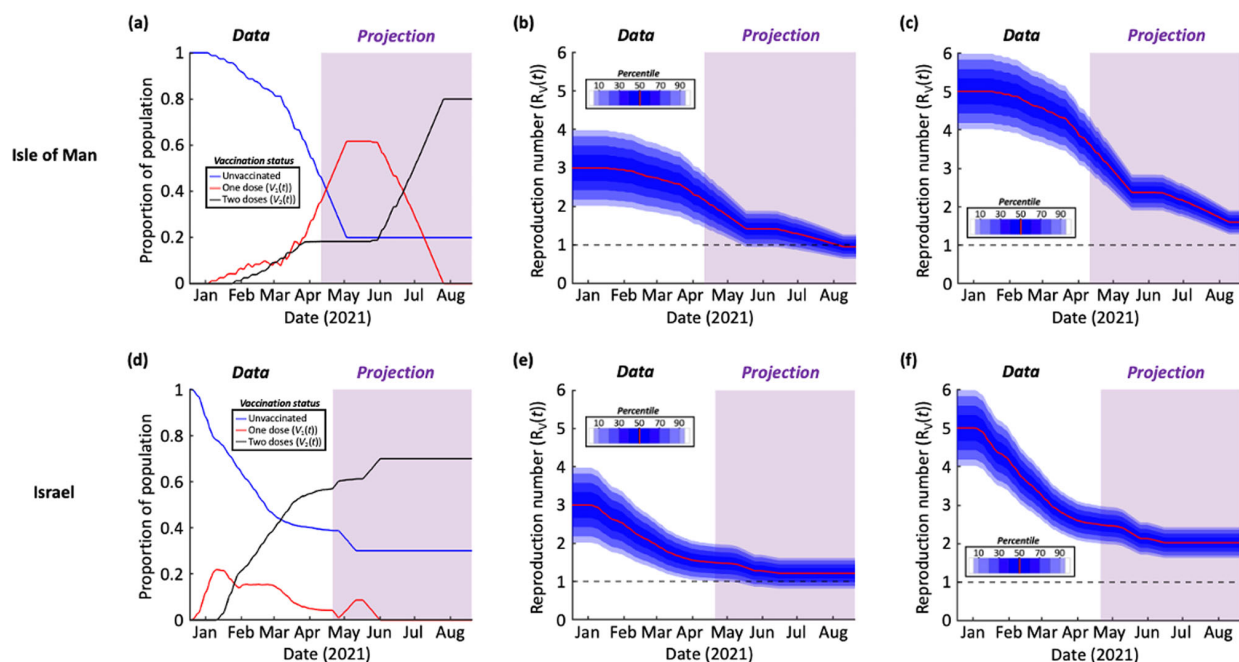


Fig. 2 Effect of vaccination on population susceptibility. **a** The proportion of the population of the Isle of Man that is unvaccinated ($1 - V_1(t) - V_2(t)$), vaccinated with a single dose ($V_1(t)$) and vaccinated with two doses ($V_2(t)$). The period in which vaccination data were available is shown in white, and the period in which vaccination data were projected is shown in purple. **b** The time-dependent reproduction number ($R_V(t)$) corresponding to the vaccination data in panel a with all travel restrictions and non-pharmaceutical interventions removed, starting from a mean initial value of $R_V(0) = R_0 = 3$. To account for uncertainty in the value of R_0 , a normal distribution was assumed about the mean value of R_0 with variance $\sigma^2 = 0.25$, which is reflective of the range of R_0 values estimated early in the COVID-19 pandemic⁶⁵. **c** Analogous to panel b but starting from a mean initial value of $R_V(0) = R_0 = 5$. **d–f** Analogous to panels a–c but using vaccination data for Israel. In all panels, $t = 0$ days corresponds to 18th December 2020. Ticks on the x-axes refer to the starts of the months labelled. Parameter values are shown in Table 1.

represents a point estimate of the risk accounting for uncertainty in pathogen transmissibility.

The IOR represents the probability of sustained local transmission based only on the value of $R_V(t)$ at the precise instant that the virus is introduced into the population. When an introduction occurs while vaccines are being deployed, there is a background of decreasing population susceptibility, and so $R_V(t)$ may in fact decrease over the initial stages of a potential outbreak. For that reason, it is unsurprising that the IOR sometimes overestimates the outbreak risk compared to the other risk metrics (see e.g. black lines in Fig. 3a, c – and see Discussion).

In contrast, we found close agreement between the COR, SOR, and NOR. Due to the high assumed vaccine uptake in the Isle of Man, the outbreak risk at the end of the vaccination programme there was calculated to be lower than when the vaccine rollout was completed in Israel (although we also considered supplementary analyses with different assumed vaccine uptake values – Fig. S1). In the first scenario that we considered (mean $R_0 = 3$), which is representative of the transmissibility of the original SARS-CoV-2 virus, the outbreak risk was projected to be low following the vaccination programme in the Isle of Man.

However, when the virus was assumed to be more transmissible (mean $R_0 = 5$), as is the case for newly emerged variants of concern such as the Alpha variant (B.1.1.7), the outbreak risk was found to be substantial even following the projected end of the vaccination campaigns in both the Isle of Man and Israel. For example, when the assumed vaccine uptake values of $\nu = 0.8$ and $\nu = 0.7$ had been achieved in the Isle of Man and Israel, respectively, the NOR took values of 0.37 (95% Equal-Tailed Credible Interval (CrI): [0.22,0.48], calculating using the 95% Equal-Tailed CrI for $R_V(t)$ at the end of the vaccination rollout) and 0.51 (95% Equal-Tailed CrI: [0.39,0.59]) in those locations.

Given the uncertainty in epidemiological modelling projections, the key point to note is that these values are greater than zero (rather than the exact outbreak risk estimates themselves), suggesting that outbreaks can still occur when NPIs are removed following the completion of vaccination programmes. This result is robust to a range of assumptions surrounding virus transmission and the effectiveness of vaccines (Figs S1–S4).

Discussion

As vaccines are administered in countries around the world, attention has turned to the possibility that transmission will soon be reduced to the extent that travel restrictions and other NPIs can be relaxed. Here, we have investigated the impact of removing NPIs on the risk of outbreaks occurring in locations with low prevalence and a substantial proportion of the population vaccinated. We used four metrics to estimate the risk that a case introduced at any stage in the vaccination rollout leads to an outbreak driven by local transmission, in a scenario in which NPIs are removed. We calculated temporal changes in the values of these metrics in the context of vaccination in the Isle of Man and in Israel, two locations with low prevalence and vaccination campaigns that have progressed quickly.

We found that vaccination is leading to a substantial drop in the potential for virus transmission in both the Isle of Man and Israel, as indicated by a decreasing value of the time-dependent reproduction number, $R_V(t)$ (Fig. 2). However, even when the vaccine rollout is completed (with 80% of individuals fully vaccinated in the Isle of Man, and 70% in Israel; other values are considered in Fig. S1), the combination of vaccines not preventing transmission entirely, incomplete vaccine uptake and the emergence of novel SARS-CoV-2 variants suggests that the risk of

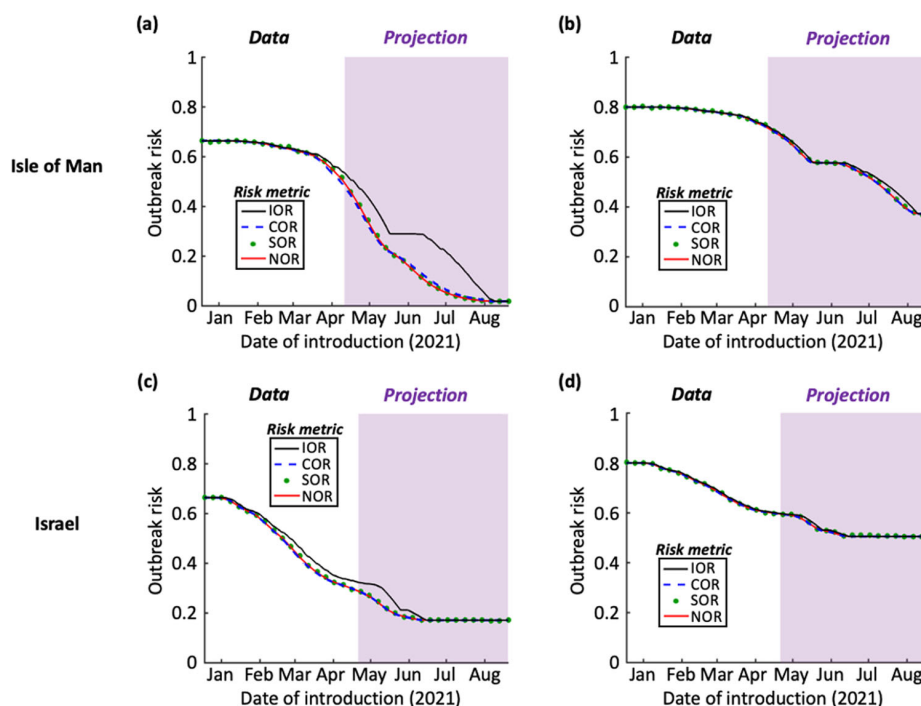


Fig. 3 The risk that an infectious case introduced at each stage of the vaccination campaign initiates an outbreak, if travel restrictions and other non-pharmaceutical interventions are removed. **a** The outbreak risk was assessed using the four metrics (Instantaneous Outbreak Risk – black; Case Outbreak Risk – blue dashed; Simulated Outbreak Risk – green dots; Numerical Outbreak Risk – red), based on vaccination data from the Isle of Man and assuming mean $R_0 = 3$, as in Fig. 2b. The period in which vaccination data were available is shown in white, and the period in which vaccination data were projected is shown in purple. **b** Analogous to panel a but with a mean $R_0 = 5$ (see Fig. 2c). **c** Analogous to panel a but using vaccination data for Israel (see Fig. 2e). **d** Analogous to panel a but using vaccination data for Israel with a mean $R_0 = 5$ (see Fig. 2f). Ticks on the x-axes refer to the starts of the months labelled. Parameter values are shown in Table 1.

outbreaks initiated by infected individuals arriving from elsewhere will not be eliminated when NPIs are removed (Fig. 3). This conclusion holds unless the vaccine uptake is very high (Fig. S1). This suggests that, when NPIs such as travel restrictions are relaxed, it will still be advisable to be aware of the potential for local transmission. Ensuring that case numbers are reduced elsewhere (i.e. in locations that imported cases might travel from) will reduce the risk of importations, and strategies should be considered to suppress outbreaks quickly if importations occur.

One potential use of our modelling framework by policy-makers is to identify the dates on which travel restrictions can be relaxed, based on a maximum acceptable outbreak risk. As an example, if the maximum acceptable value of the NOR is 0.4, then our analysis with $R_0 = 5$ suggests that travel restrictions can be lifted on the Isle of Man at the end of July 2021 (Fig. 3b; the first date on which the NOR is projected to fall below 0.4 in our illustrative analysis is 29th July 2021). However, as described above, since the risk of local transmission following introductions remains, it will be necessary to continue to monitor inbound passengers in low prevalence settings to identify infected individuals, even once vaccination programmes are completed.

Our modelling approach for assessing outbreak risks was motivated by studies in which the potential for pathogen transmission varies periodically^{38–40,43–46}, due, for example, to seasonal changes in weather conditions that affect transmission. In that context, Carmona and Gandon³⁸ describe a “winter is coming effect”, in which the risk of an outbreak is lower than current environmental conditions suggest if conditions become less favourable for transmission in the near future. In the terminology used in our manuscript, if current environmental conditions promote pathogen transmission, then the IOR is

expected to be high. This is because the IOR reflects the outbreak risk based on the conditions at the precise instance when the virus is introduced into the population. However, if environmental conditions are expected to become less favourable for transmission in the near future, then the values of the other risk metrics are lower than the IOR, since those metrics account for changes in transmissibility over the initial phase of the potential outbreak. Although in general we found a close agreement between the four metrics that we considered, a background of decreasing population susceptibility can lead to a similar effect in which the IOR is larger than the COR, SOR, or NOR (e.g. Figure 3a, c).

In this study, we used a simple branching process model to investigate the risk of outbreaks when NPIs are removed during a vaccination programme. This involved considering whether introduced cases are likely to lead to sustained local transmission or instead fade out without causing an outbreak. We made the standard branching process modelling assumption that population immunity is unaffected by infections in the earliest stages of potential outbreaks^{19,20,27–29,47}. In other words, infection-acquired immunity following the arrival of the pathogen in the host population is not considered. While this is reasonable when case numbers are low in the initial stages of potential outbreaks, a more detailed model is needed to explore other quantities, such as the eventual size of outbreaks. Following a vaccination programme, outbreaks are likely to be smaller than those that occur before vaccines are widely administered.

Another simplification of our model is that we only accounted for changes in population susceptibility due to the vaccine rollout. We did not account for prior immunity of some members of the population due to previous exposures to the virus⁴⁸. At the time of writing (1st May 2021), there have been 1,154 confirmed cases

in the Isle of Man and 838,000 confirmed cases in Israel. Since these case numbers correspond to a relatively small proportion of the host population (representing 1.4% and 9.6% of the population in the Isle of Man and Israel, respectively), we do not expect this assumption to affect our key findings. Furthermore, immunity is likely to wane over time^{49–51}, reducing the effect of previous exposures on the outbreak risk. To test the potential impact of infection-acquired immunity arising from cases occurring before May 2021, we also conducted a supplementary analysis in which the value of R_0 is reduced by 1.4% in the Isle of Man and 9.6% in Israel, and we found qualitatively similar results (Fig. S2): even in this scenario, there is still a risk of outbreaks due to imported cases once the vaccination programmes are completed. Importantly, in other countries in which higher numbers of cases have occurred, prior immunity may play a larger role in reducing the risk of outbreaks compared to the low prevalence settings considered here. Understanding the extent of this effect, based on the rate at which immunity wanes, is an important target for further study.

In this research, we assumed that vaccines reduce transmission by lowering the susceptibility of vaccinated hosts compared to unvaccinated hosts. Two effective vaccine doses were assumed to reduce the probability of infection per exposure substantially (by $\eta_2 = 0.85$), which is in line with estimates for the Pfizer vaccine used in the UK Government's COVID-19 roadmap modelling²¹. However, there is uncertainty about the precise level of protection offered by different COVID-19 vaccines. For that reason, we also conducted a supplementary analysis (Fig. S3) in which we considered a range of values of η_2 . In every scenario that we explored, the risk of outbreaks following the completion of vaccination programmes in low prevalence settings was greater than zero, with larger outbreak risks for lower values of η_2 .

In addition to reducing the risk that a host becomes infected as assumed here⁵², vaccination may also reduce the risk of onwards transmission following infection⁵³, reduce the risk of severe disease developing, or a combination of these effects⁵⁴. In principle, it would be possible to develop a more complex model that accounts for all of these effects separately, at the cost of having to estimate (or assume) values of the associated parameter values. Vaccines can be assumed to reduce the risk that vaccinated hosts become infected by lowering their susceptibility to infection compared to unvaccinated hosts (leaky vaccines), or by protecting a subset of the vaccinated hosts completely from infection (all-or-nothing vaccines)²⁴, and the implications of these different mechanisms on outbreaks while a vaccination campaign is ongoing could be considered. It would also be possible to develop a transmission model in which either vaccine- and/or infection-acquired immunity wanes. There is currently substantial uncertainty about the extent and speed of waning immunity, with the hope that antibody responses and/or cellular immune responses may provide long-lasting immune memory, at least in some individuals^{55,56}. Booster vaccination for vulnerable individuals is being considered in high-income countries, given the possibility that vaccine-acquired immunity wanes. We note that any waning of immunity will act to enhance our main qualitative conclusion that the risk of outbreaks is unlikely to be eliminated completely in low prevalence settings by vaccination programmes.

Other factors that may merit consideration in further analyses of future outbreak risks include the impact of population structure on both vaccine effectiveness⁵⁷ and transmission^{58,59}, and the possibility that individuals' behaviours may be different following the pandemic compared to beforehand, even when NPIs and travel restrictions are removed. For precise quantitative outbreak risk predictions to be made, it may be necessary to estimate R_0 in different locations. It would also be necessary to

incorporate changes in R_0 due to further new variants in our analyses, although we note that any increase in transmissibility will act to strengthen our main conclusion that vaccine-acquired immunity alone is unlikely to be sufficient to prevent outbreaks. Another important consideration is the degree of heterogeneity in transmission between different infected individuals. For a range of pathogens, including SARS-CoV-2, the majority of transmissions are generated by a relatively small proportion of infectious hosts^{60–63}. We therefore also conducted a supplementary analysis (Supplementary Notes) in which we considered the dependence of outbreak risk estimates following the completion of vaccination programmes in the Isle of Man and Israel on the assumed individual-level variation in SARS-CoV-2 transmission. While a higher degree of heterogeneity (i.e. a higher probability of “superspreading” events) acts to lower the outbreak risk, in each scenario explored outbreaks were still possible in low prevalence settings following vaccination programmes when NPIs are removed (Fig. S4).

To summarise, our analyses have demonstrated the general principle that, even following vaccination programmes in low prevalence settings, the risk of outbreaks remains when NPIs are removed. Our intention is not to argue that travel restrictions and other NPIs should not be relaxed once vaccination programmes are sufficiently advanced, but rather that measures should be taken to ensure that any clusters of cases are suppressed quickly if they arise. A local outbreak requires two steps: first, the virus must be imported from elsewhere; second, local transmission must occur. The first step emphasises the need for a global approach to minimising transmission, since higher case numbers at a potential source location translate into a higher importation risk. The second step emphasises the need to ensure that, while vaccination acts to reduce transmission substantially, continued surveillance of inbound passengers for infection, combined with isolation and/or testing of contacts of detected infected individuals, is important when travel restrictions and other NPIs are relaxed. These measures are necessary, since only once the global prevalence of SARS-CoV-2 infections is reduced substantially can the risk of outbreaks in low prevalence settings be eliminated.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

All data required to reproduce the results presented here are available at https://github.com/robin-thompson/Outbreaks_Low_Prevalence_Settings/64. Data on the proportion of the population vaccinated was obtained from ³⁴ and ³⁵ (see Methods). Source data for the main figures in the manuscript can be accessed at https://github.com/robin-thompson/Outbreaks_Low_Prevalence_Settings/64. No restrictions exist on data availability.

Code availability

All code required to reproduce the results presented here are available at https://github.com/robin-thompson/Outbreaks_Low_Prevalence_Settings/, along with relevant documentation⁶⁴.

Received: 25 May 2021; Accepted: 3 September 2021;

Published online: 15 October 2021

References

1. Liu, Y., Morgenstern, C., Kelly, J., Lowe, R. & Jit, M. The impact of non-pharmaceutical interventions on SARS-CoV-2 transmission across 130 countries and territories. *BMC Med.* **19**, 40 (2021).
2. Leung, K., Wu, J. T., Liu, D. & Leung, G. M. First-wave COVID-19 transmissibility and severity in China outside Hubei after control measures,

- and second-wave scenario planning: a modelling impact assessment. *Lancet*. **395**, 1382–1393 (2020).
3. Cowling, B. J. et al. Impact assessment of non-pharmaceutical interventions against coronavirus disease 2019 and influenza in Hong Kong: an observational study. *Lancet Public Heal*. **4**, 397–404 (2020).
 4. Chinazzi, M. et al. The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. *Science*. **368**, 395–400 (2020).
 5. Daon, Y., Thompson, R. N. & Obolski, U. Estimating COVID-19 outbreak risk through air travel. *J Travel Med*. **27**, taaa093 (2020).
 6. Prem, K. et al. The effect of control strategies to reduce social mixing on outcomes of the COVID-19 epidemic in Wuhan, China: a modelling study. *Lancet Public Heal*. **5**, e261–e270 (2020).
 7. Harris R. J. et al. Impact of vaccination on household transmission of SARS-CoV-2 in England. *medRxiv* (2021).
 8. Rossman, H. et al. COVID-19 dynamics after a national immunization program in Israel. *Nat Med*. **27**, 1055–1061 (2021).
 9. Shah A. S. et al. Effect of vaccination on transmission of COVID-19: an observational study in healthcare workers and their households. *medRxiv* (2021).
 10. Richterman A., Meyerowitz E. A. & Cevik M. Indirect protection by reducing transmission: Ending the pandemic with SARS-CoV-2 vaccination. *Open Forum Infect Dis*. ofab259 (2021).
 11. Polack, F. P. et al. Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine. *N Engl J Med*. **383**, 2603–2615 (2020).
 12. Voysey, M. et al. Safety and efficacy of the ChAdOx1 nCoV-19 vaccine (AZD1222) against SARS-CoV-2: an interim analysis of four randomised controlled trials in Brazil, South Africa, and the UK. *Lancet*. **397**, 99–111 (2021).
 13. Baden, L. R. et al. Efficacy and Safety of the mRNA-1273 SARS-CoV-2 Vaccine. *N Engl J Med*. **384**, 403–416 (2021).
 14. Baker, M. G., Wilson, N. & Blakely, T. Elimination could be the optimal response strategy for covid-19 and other emerging pandemic diseases. *BMJ*. **371**, m4907 (2020).
 15. Huang Q. S. et al. Impact of the COVID-19 nonpharmaceutical interventions on influenza and other respiratory viral infections in New Zealand. *Nat Commun*. **12** (2021).
 16. Baker, M. G., Wilson, N. & Anglemeyer, A. Successful elimination of Covid-19 transmission in New Zealand. *N Engl J Med*. **383**, e56 (2020).
 17. Hart, W. S., Maini, P. K. & Thompson, R. N. High infectiousness immediately before COVID-19 symptom onset highlights the importance of continued contact tracing. *Elife*. **10**, e65534 (2021).
 18. Russell, T. W. et al. Effect of internationally imported cases on internal spread of COVID-19: a mathematical modelling study. *Lancet Public Heal*. **6**, e12–e20 (2021).
 19. Hellewell, J. et al. Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts. *Lancet Glob Heal*. **8**, e488–e496 (2020).
 20. Thompson, R. N. Novel coronavirus outbreak in Wuhan, China, 2020: Intense surveillance is vital for preventing sustained transmission in new locations. *J Clin Med*. **9**, 498 (2020).
 21. Keeling M. J., Moore S., Dyson L., Tildesley M. J. & Hill E. M. Road map scenarios and sensitivity (2021). Available: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/975911/S1184_SPI-M_University_of_Warwick_Road_Map_Scenarios_and_Sensitivity.pdf.
 22. Moore, S., Hill, E. M., Tildesley, M. J., Dyson, L. & Keeling, M. J. Vaccination and non-pharmaceutical interventions for COVID-19: a mathematical modelling study. *Lancet Infect Dis*. **S1473-3099**, 00143–2 (2021).
 23. Moore, S., Hill, E. M., Dyson, L., Tildesley, M. J. & Keeling, M. J. Modelling optimal vaccination strategy for SARS-CoV-2 in the UK. *PLOS Comput Biol*. **17**, e1008849 (2020).
 24. Halloran, M. E., Haber, M. & Longini, I. M. Jr. Interpretation and estimation of vaccine efficacy under heterogeneity. *Am J Epidemiol*. **136**, 328–343 (1992).
 25. Razai, M. S., Osama, T., McKechnie, D. G. J. & Majeed, A. Covid-19 vaccine hesitancy among ethnic minority groups. *BMJ*. **372**, n513 (2021).
 26. Davies, N. et al. Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England. *Science*. **372**, eabg3055 (2021).
 27. Althaus, C. L., Low, N., Musa, E. O. & Shuaib, F. Gsteiger. Ebola virus disease outbreak in Nigeria: Transmission dynamics and rapid control. *Epidemics*. **11**, 80–84 (2015).
 28. Thompson, R. N., Gilligan, C. A. & Cunliffe, N. J. Detecting presymptomatic infection is necessary to forecast major epidemics in the earliest stages of infectious disease outbreaks. *PLOS Comput Biol*. **12**, e1004836 (2016).
 29. Merler, S. et al. Containing Ebola at the source with ring vaccination. *PLOS Negl Trop Dis*. **10**, e0005093 (2016).
 30. Fraser, C. Estimating individual and household reproduction numbers in an emerging epidemic. *PLOS One*. **2**, e758 (2007).
 31. White, L. F., Moser, C. B., Thompson, R. N. & Pagano, M. Statistical estimation of the reproductive number from case notification data. *Am J Epidemiol*. **190**, 611–620 (2020).
 32. Thompson, R. N. et al. Improved inference of time-varying reproduction numbers during infectious disease outbreaks. *Epidemics*. **29**, 100356 (2019).
 33. Gostic, K. M. et al. Practical considerations for measuring the effective reproductive number, Rt. *PLOS Comput Biol*. **16**, e1008409 (2020).
 34. Isle of Man Government. COVID-19 Vaccination Programme. (2021) [cited 15 Apr 2021]. Available: <https://covid19.gov.im/about-coronavirus/open-data-downloads/>.
 35. Our World in Data. Coronavirus (COVID-19) vaccinations. (2021) [cited 23 Apr 2021]. Available: <https://ourworldindata.org/covid-vaccinations>.
 36. Keeling, M. J. & Ross, J. V. On methods for studying stochastic disease dynamics. *J R Soc Interface*. **5**, 171–181 (2008).
 37. Thompson, R. N., Gilligan, C. A. & Cunliffe, N. J. 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**, 20200690 (2020).
 38. Carmona, P. & Gandon, S. Winter is coming: Pathogen emergence in seasonal environments. *PLOS Comput Biol*. **16**, e1007954 (2020).
 39. Ball, F. G. The threshold behaviour of epidemic models. *J Appl Probab*. **20**, 227–241 (1983).
 40. Bacaër, N. & Ait Dads, E. H. On the probability of extinction in a periodic environment. *J Math Biol*. **68**, 533–548 (2014).
 41. Kendall, D. G. On the generalized “birth-and-death” process. *Ann Math Stat*. **19**, 1–15 (1948).
 42. Gillespie, D. T. Exact stochastic simulation of coupled chemical reactions. *J Phys Chem*. **8**, 2340–2361 (1977).
 43. Klein, B. & MacDonald, P. D. M. The multitype continuous-time Markov branching process in a periodic environment. *Adv Appl Probab*. **12**, 81–93 (1980).
 44. Jagers, P. & Nerman, O. Branching processes in periodically varying environment. *Ann Probab*. **13**, 254–268 (1985).
 45. Nipa, K. F. & Allen, L. J. S. Disease emergence in multi-patch stochastic epidemic models with demographic and seasonal variability. *Bull Math Biol*. **82**, 152 (2020).
 46. Bacaër, N. & Ed-Darraz, A. On linear birth-and-death processes in a random environment. *J Math Biol*. **69**, 73–90 (2014).
 47. Lovell-Read, F. A., Funk, S., Obolski, U., Donnelly, C. A. & Thompson, R. N. Interventions targeting non-symptomatic cases can be important to prevent local outbreaks: SARS-CoV-2 as a case study. *J R Soc Interface*. **18**, 20201014 (2021).
 48. Sachak-Patwa, R., Byrne, H. M. & Thompson, R. N. Accounting for cross-immunity can improve forecast accuracy during influenza epidemics. *Epidemics*. **34**, 100432 (2020).
 49. Anderson, R. M., Vegvari, C., Truscott, J. & Collyer, B. S. Challenges in creating herd immunity to SARS-CoV-2 infection by mass vaccination. *Lancet*. **396**, 1614–1616 (2020).
 50. Kissler, S. M., Tedijanto, C., Goldstein, E., Grad, Y. H. & Lipsitch, M. Projecting the transmission dynamics of SARS-CoV-2 through the postpandemic period. *Science*. **368**, 860–868 (2020).
 51. Crellen, T. et al. Dynamics of SARS-CoV-2 with waning immunity in the UK population. *Philos Trans R Soc B Biol Sci*. **376**, 20200274 (2021).
 52. Pritchard, E. et al. Impact of vaccination on new SARS-CoV-2 infections in the United Kingdom. *Nat Med*. **27**, 1370–1378 (2021).
 53. Harris, R. J. et al. Effect of vaccination on household transmission of SARS-CoV-2 in England. *N Engl J Med*. **385**, 759–760 (2021).
 54. Gog, J. R., Hill, E. M., Danon, L. & Thompson, R. N. Vaccine escape in heterogeneous populations: Insights for SARS-CoV-2 from a simple model. *R Soc Open Sci*. **8**, 210530 (2021).
 55. Zuo, J. et al. Robust SARS-CoV-2-specific T cell immunity is maintained at 6 months following primary infection. *Nat Immunol*. **22**, 620–626 (2021).
 56. Tea, F. et al. SARS-CoV-2 neutralizing antibodies: longevity, breadth, and evasion by emerging viral variants. *PLOS Med*. **18**, e1003656 (2021).
 57. Bubar, K. M. et al. Model-informed COVID-19 vaccine prioritization strategies by age and serostatus. *Science*. **371**, 6532 (2021).
 58. Britton, T., Ball, F. & Trapman, P. A mathematical model reveals the influence of population heterogeneity on herd immunity to SARS-CoV-2. *Science*. **369**, 846–849 (2020).
 59. Davies, N. G. et al. Age-dependent effects in the transmission and control of COVID-19 epidemics. *Nat Med*. **26**, 1205–1211 (2020).
 60. Thompson, R. N. et al. Key questions for modelling COVID-19 exit strategies. *Proc R Soc B Biol Sci*. **287**, 20201405 (2020).
 61. Endo, A., CMMID COVID-19 working group, Abbott, S., Kucharski, A. J. & Funk, S. Estimating the overdispersion in COVID-19 transmission using outbreak sizes outside China. *Wellcome Open Res*. **5**, 67 (2020).

62. Lloyd-Smith, J. O., Schreiber, S. J., Kopp, P. E. & Getz, W. M. Superspreading and the effect of individual variation on disease emergence. *Nature*. **438**, 355–359 (2005).
63. Akhmetzhanov, A. R., Jung, S.-M., Cheng, H.-Y. & Thompson, R. N. A hospital-related outbreak of SARS-CoV-2 associated with variant Epsilon (B.1.429) in Taiwan: Transmission potential and outbreak containment under intensified contact tracing, January–February 2021. *Int J Infect Dis*. **110**, 15–20 (2021).
64. Sachak-Patwa R., Byrne H. M., Dyson L. & Thompson R. N. Code and data to accompany the article “The risk of SARS-CoV-2 outbreaks in low prevalence settings following the removal of travel restrictions.” In: Zenodo [Internet]. (2021). Available: <https://doi.org/10.5281/zenodo.5267293>.
65. Billah, M. A., Miah, M. M. & Khan, M. N. Reproductive number of coronavirus: a systematic review and meta-analysis based on global level evidence. *PLOS One*. **15**, e0242128 (2020).
66. Kucharski, A. J. et al. Effectiveness of isolation, testing, contact tracing, and physical distancing on reducing transmission of SARS-CoV-2 in different settings: a mathematical modelling study. *Lancet Infect Dis*. **20**, 1151–1160 (2020).
67. Centers for Disease Control and Prevention. Interim public health recommendations for fully vaccinated people. (2021).
68. Sallam, M. Covid-19 vaccine hesitancy worldwide: A concise systematic review of vaccine acceptance rates. *Vaccines*. **9**, 160 (2021).
69. Worldometer. Current world population. (2021) [cited 21 Apr 2021]. Available: <https://www.worldometers.info/world-population/>.
70. Isle of Man Government. Frequently asked questions about the vaccine. (2021).
71. Wise, J. Covid-19: Pfizer BioNTech vaccine reduced cases by 94% in Israel, shows peer reviewed study. *BMJ*. **372**, n567 (2021).

Acknowledgements

We thank members of the Wolfson Centre for Mathematical Biology (University of Oxford) and the Zeeman Institute for Systems Biology and Infectious Disease Epidemiology Research (University of Warwick) for helpful discussions about this work. We also thank the two reviewers (Pieter Trapman and one anonymous reviewer), as well as Ed Hill and Nick Lai, for helpful feedback that enabled us to improve this manuscript. RS-P was supported by the EPSRC through the Centre for Doctoral Training in Industrially Focussed Mathematical Modelling at the University of Oxford (grant EP/L015803/1), in

collaboration with Biosensors Beyond Borders Ltd. LD was supported by the MRC through the COVID-19 Rapid Response Rolling Call (grant MR/V009761/1) and by UKRI through the JUNIPER modelling consortium (grant MR/V038613/1). This work was

funded in part by the UKRI (grant EP/V053507/1). The funders had no role in study design, data collection and analysis, preparation of the paper, or the decision to publish.

Author contributions

Conceptualisation: RNT, RS-P, LD; Methodology: RNT, RS-P; Investigation: RS-P; Supervision: RNT, HMB; Writing – original draft: RNT, RS-P; Writing – review and editing: All authors.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s43856-021-00038-8>.

Correspondence and requests for materials should be addressed to Robin N. Thompson.

Peer review information *Communications Medicine* thanks Pieter Trapman and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permission information is available at <http://www.nature.com/reprints>

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



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021

Appendix G

Accounting for cross-immunity can improve forecast accuracy during influenza epidemics

Chapter 3 was published as a journal article in *Epidemics* (<https://doi.org/10.1016/j.epidem.2020.100432>).

Each of the authors contributed to the following facets of the published work:

- Conceptualisation: RS-P, RNT, HMB.
- Methodology: RS-P, RNT.
- Investigation: RS-P.
- Supervision: RNT, HMB.
- Writing – original draft: RS-P, RNT.
- Writing – review and editing: RS-P, RNT, HMB.



Accounting for cross-immunity can improve forecast accuracy during influenza epidemics

Rahil Sachak-Patwa^{a,*}, Helen M. Byrne^a, Robin N. Thompson^{a,b,c}

^a Mathematical Institute, University of Oxford, Andrew Wiles Building, Radcliffe Observatory Quarter, Woodstock Road, Oxford, OX2 6GG, UK

^b Christ Church, University of Oxford, St Aldates, Oxford, OX1 1DP, UK

^c Present address: Mathematics Institute, University of Warwick, Zeeman Building, Coventry, CV4 7AL, UK

ARTICLE INFO

Keywords:

Mathematical modelling
Influenza forecasting
Real-time forecast
Cross-immunity
2009 H1N1 pandemic

ABSTRACT

Previous exposure to influenza viruses confers cross-immunity against future infections with related strains. However, this is not always accounted for explicitly in mathematical models used for forecasting during influenza outbreaks. We show that, if an influenza outbreak is due to a strain that is similar to one that has emerged previously, then accounting for cross-immunity explicitly can improve the accuracy of real-time forecasts. To do this, we consider two infectious disease outbreak forecasting models. In the first (the “1-group model”), all individuals are assumed to be identical and cross-immunity is not accounted for. In the second (the “2-group model”), individuals who have previously been infected by a related strain are assumed to be less likely to experience severe disease, and therefore recover more quickly, than immunologically naive individuals. We fit both models to estimated case notification data (including symptomatic individuals as well as laboratory-confirmed cases) from Japan from the 2009 H1N1 influenza pandemic, and then generate synthetic data for a future outbreak by assuming that the 2-group model represents the epidemiology of influenza infections more accurately. We use the 1-group model (as well as the 2-group model for comparison) to generate forecasts that would be obtained in real-time as the future outbreak is ongoing, using parameter values estimated from the 2009 epidemic as informative priors, motivated by the fact that without using prior information from 2009, the forecasts are highly uncertain. In the scenario that we consider, the 1-group model only produces accurate outbreak forecasts once the peak of the epidemic has passed, even when the values of important epidemiological parameters such as the lengths of the mean incubation and infectious periods are known exactly. As a result, it is necessary to use the more epidemiologically realistic 2-group model to generate accurate forecasts. Accounting for cross-immunity driven by exposures in previous outbreaks explicitly is expected to improve the accuracy of epidemiological modelling forecasts during influenza outbreaks.

1. Introduction

Three major influenza pandemics occurred in the 20th century, in 1918, 1957, and 1968 (Kilbourne, 2006). Each pandemic resulted in over one million deaths, with the death toll of the 1918 Spanish Flu pandemic estimated to be 50 million people (Johnson and Mueller, 2002). In 2009, reassortment of North American and Eurasian swine viruses generated a new strain of H1N1, triggering the first influenza pandemic of the 21st century (Trifonov et al., 2009; Christman et al., 2011). The virus is believed to have originated in Mexico in April 2009, and then spread rapidly across the globe, reaching 43 countries by May that year (Fraser et al., 2009; Trifonov et al., 2009). The case fatality rate due to the virus was lower than that of previous pandemics in the 20th century (Kamigaki and Oshitani, 2009). However, the scale

of the pandemic, with estimates that 11%–21% of the global population contracted the virus, led to a significant burden on healthcare systems (Kelly et al., 2011).

Influenza A viruses mutate over time, either due to antigenic drift, where a small mutation in the genes of the virus leads to changes in the surface proteins producing a closely related strain, or, antigenic shift, an abrupt, major change in the virus due to recombination and reassortment, resulting in new surface proteins (Bouvier and Palese, 2008; Kim et al., 2018). A new form of the virus that has emerged due to drift is likely to have similar antigenic properties to the original virus, and thus individuals who were infected with the original virus may have acquired some immunity to the altered virus. However, when antigenic shift occurs, many individuals are likely to have little or no immunity

* Corresponding author.

E-mail address: rahil.sachak-patwa@maths.ox.ac.uk (R. Sachak-Patwa).

<https://doi.org/10.1016/j.epidem.2020.100432>

Received 5 July 2020; Received in revised form 11 December 2020; Accepted 15 December 2020

Available online 17 December 2020

1755-4365/© 2021 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

to the new virus (depending on their lifetime history of infection). Due to the random nature of the emergence of new influenza strains, it is not currently possible to predict when future pandemics will occur, and which strains will cause these pandemics (Neumann and Kawaoka, 2019). However, mathematical models have been used extensively for forecasting and informing public health measures when outbreaks are ongoing (Ferguson et al., 2006; Hall et al., 2007; Nishiura, 2011; Ohkusa et al., 2011; Tizzoni et al., 2012; Biggerstaff et al., 2016; Thompson and Brooks-Pollock, 2019).

The most basic infectious disease outbreak models assume that individuals have similar characteristics (Chowell et al., 2006; Bettencourt and Ribeiro, 2008). More complex models account for differences between individuals. For example, in many studies that aim to determine optimal vaccination strategies, populations are split into low-risk and high-risk groups (Gani et al., 2005; Dushoff et al., 2007), and spatial heterogeneity can be incorporated by partitioning individuals according to their location (Longini Jr. et al., 2004; Ohkusa et al., 2009). Commonly, due to different contact rates between individuals of different ages, as well as varying case fatality rates between age groups, age-structured models are used (Chowell et al., 2009; Medlock and Galvani, 2009; Glasser et al., 2010; Klepac et al., 2018; Davies et al., 2020; Thompson, 2020b). Models of influenza transmission that incorporate evolutionary and multi-strain dynamics (He et al., 2015; Yang et al., 2015; Du et al., 2017), as well as phylodynamics (accounting for temporal changes in pathogen phylogenies), have also been considered (Grenfell, 2004; Koelle et al., 2006; Frost et al., 2015).

Other types of heterogeneity also play an important role in the dynamics of influenza outbreaks. Previous exposure to an influenza virus confers homosubtypic cross-immunity to antigenically related strains of the original virus due to cross-reactive neutralising antibodies (Grebe et al., 2008; Kreijtz et al., 2011; Krammer, 2019; Padilla-Quirarte et al., 2019). These neutralising antibodies, predominantly against haemagglutinin (HA) proteins on the surface of the virus, block viral entry to the host cell (Sym et al., 2009; Huang et al., 2013). It has also been suggested that cross-protective non-HA antibody immunity may lead to reduced disease severity (Fang et al., 2011). Neuraminidase (NA) specific antibodies have also been shown to reduce influenza viral titres, morbidity, and virus shedding (Schulman et al., 1968; Murphy et al., 1972; Couch et al., 1974; Webster et al., 1988; Epstein et al., 1993; Mozdzanowska et al., 1999). As well as homosubtypic immunity, there is evidence of heterosubtypic immunity, namely cross-protection generated by prior exposure to a different but related influenza subtype (Nguyen et al., 2007; Grebe et al., 2008; Fox et al., 2017). Cross-immunity – homosubtypic and/or heterosubtypic – may explain why there has not been an influenza pandemic as severe as the 1918 pandemic in the last century (Thompson et al., 2019).

A significant proportion of elderly individuals carried pre-existing immunity (due to cross-reactive anti-HA antibodies) to the influenza strain that was responsible for the 2009 pandemic (Hancock et al., 2009; Xing and Cardona, 2009; Bandaranayake et al., 2010; Hardelid et al., 2010; Gostic et al., 2019). This may be due to the similarities between the 2009 H1N1 virus and the 1918 Spanish Flu virus, as viral descendants of the 1918 Spanish Flu virus continued to circulate until the 1957 pandemic (Skountzou et al., 2010; Xu et al., 2010). The suggestion that immunity was due to previous infections that occurred decades earlier implies that this form of cross-immunity is likely to be lifelong (Yu et al., 2008). The consequences of pre-existing immunity can be seen in the age distribution of infected individuals in Japan in the 2009 pandemic, where only a small proportion of the individuals who sought medical attention were elderly (Mizumoto et al., 2013). A number of studies in animals have also shown that prior infection with seasonal H1N1 strains have been able to provide substantial protection against infection with the 2009 H1N1 pandemic virus (Ellebedy et al., 2010; Kash et al., 2010; Laurie et al., 2010; Ellebedy et al., 2011). As well as the heterogeneity between hosts in infection risk and age

mentioned previously, models in which populations are structured according to whether or not individuals carry pre-existing immunity can also be formulated (Andreasen et al., 1997; Martcheva and Pilyugin, 2006; Reluga et al., 2008; Penman et al., 2016; Thompson et al., 2019).

In this paper, our attention is directed towards how cross-immunity (either homosubtypic or heterosubtypic) affects the predictability of epidemics. We use mathematical models to investigate whether or not it is necessary to account for cross-immune individuals in the population when forecasting the dynamics of future influenza epidemics. We consider two epidemiological models. In the first, cross-immunity is ignored (the “1-group model”). In the second, more epidemiologically realistic model (the “2-group model”), individuals with and without cross-immunity are accounted for explicitly.

First, we estimate the values of the parameters of each model (specifically, the transmission rate and the effective population size) using data from the 2009 H1N1 influenza epidemic in Japan. We then consider a synthetic future influenza epidemic of an antigenically similar strain, simulated using the more epidemiologically realistic 2-group model. We explore whether or not accurate forecasts of this epidemic can be obtained in real-time. If uninformative priors are used and parameters are estimated in real-time, even the more realistic 2-group model is unable to generate accurate forecasts of the remainder of the epidemic before the peak occurs. This motivates us to incorporate information from the 2009 epidemic to set informative priors. We show that forecasts made using the 1-group model in advance or at the start of a future epidemic are inaccurate, even if informative priors are used, because that model does not account for differing levels of cross-immunity between the starts of the 2009 epidemic and the future epidemic. We then use both information from the 2009 epidemic and data obtained as the future outbreak is ongoing to generate real-time forecasts using the 1-group and 2-group models. Early in the outbreak, only the 2-group model can provide accurate forecasts of the remainder of the epidemic. For that reason, cross-immunity should be included in epidemiological forecasting models whenever a large influenza outbreak is related to an antigenically similar strain that has previously caused a major epidemic.

2. Methods

2.1. Data

Our analysis involves case notification data from the 2009 H1N1 influenza epidemic in Japan comprising the estimated numbers of weekly cases seeking medical attention in that country. These nationwide data were based on sentinel data from 4800 hospitals, extrapolated to the total number of medical facilities in Japan (Nishiura, 2011; Omori and Nishiura, 2011). The data were acquired from Fig. 1 of the analysis by Nishiura (2011) using the data extraction tool <https://automeris.io/WebPlotDigitizer/> and the extracted data are available in Supplementary Data S1. The data represent the numbers of patients per week who sought medical attention and met one or more of the following criteria: (i) acute course of illness, (ii) fever higher than 38°C, (iii) cough, sputum or breathlessness (symptoms of upper respiratory infection), (iv) general fatigue, and (v) positive laboratory diagnosis.

It was estimated that 23.5% of the Japanese population were infected during the epidemic, and that 16.1% were infected and sought medical attention (Mizumoto et al., 2013). Therefore $(23.5 - 16.1)/23.5 = 31.5\%$ of infected individuals did not seek medical attention. We assume that those infected individuals who did not seek medical attention suffered mild symptoms of influenza because they were cross-immune to the virus (see Discussion). Hence, extrapolating to the rest of the population and assuming that the susceptibility of hosts is unaffected by cross-immunity, we assume when fitting the 2-group model that 31.5% of the population were cross-immune to the virus and that 68.5% were immunologically naive (i.e. had not previously acquired cross-immunity to the 2009 H1N1 pandemic strain, as they had not

been infected by a previous antigenically similar strain of the virus). However, we also test the robustness of our results to this assumption by conducting analyses for different proportions of the population that are cross-immune (Supplementary Information Section S.1).

2.2. Models

We consider two models characterising influenza outbreaks. In the first (the 1-group model), which is the commonly used SEIR model (Anderson and May, 1991; Mills et al., 2004; Chowell et al., 2006; Chen and Liao, 2008; Thompson et al., 2016), cross-immunity is neglected. In the second (the 2-group model), individuals who have been infected previously by a related strain are assumed to recover from infection more quickly than individuals who are immunologically naive. Schematics illustrating the compartmental structures of both models are shown in Fig. 1.

2.2.1. 1-group model

The 1-group SEIR model is described by the following differential equations, in which individuals are either (*S*)usceptible and available for infection, (*E*)xposed (i.e. infected but not yet infectious or symptomatic), (*I*)nfectious or (*R*)emoved:

$$\frac{dS}{dt} = -\frac{\beta}{N} SI, \quad (1)$$

$$\frac{dE}{dt} = \frac{\beta}{N} SI - \kappa E, \quad (2)$$

$$\frac{dI}{dt} = \kappa E - \mu I, \quad (3)$$

$$\frac{dR}{dt} = \mu I. \quad (4)$$

In this model, the infection rate is governed by the parameter β , the mean latent period is $1/\kappa$ weeks and the mean infectious period is $1/\mu$ weeks. The basic reproduction number under the 1-group model is given by

$$R_0 = \frac{\beta}{\mu}. \quad (5)$$

Following Cintrón-Arias et al. (2009), the number of recorded cases in week j (recorded at the end of that week), where j is the integer number of weeks since the epidemic began, is given by

$$C(j) = \int_{j-1}^j \kappa E dt. \quad (6)$$

The constant value $S + E + I + R = N$ represents the effective population size. Since pathogens are most likely to be transmitted locally, individuals in distant locations are not available for infection and so N is expected to be smaller than the true population size (Gart, 1968; Pouillot et al., 2008). In Table 1(a), we list the parameters that appear in Eqs. (1)–(4) and estimates of their values for the Japanese 2009 H1N1 epidemic (see also Sections 2.3 and 3.1).

2.2.2. 2-group model

The 2-group model is an extension of the standard SEIR model in which immunologically naive and cross-immune individuals are distinguished between. The 2-group model is given by the following system of differential equations:

$$\frac{dS_I}{dt} = -\frac{\beta}{N} S_I(I_I + I_N), \quad (7)$$

$$\frac{dE_I}{dt} = \frac{\beta}{N} S_I(I_I + I_N) - \kappa E_I, \quad (8)$$

$$\frac{dI_I}{dt} = \kappa E_I - \mu_I I_I, \quad (9)$$

$$\frac{dR_I}{dt} = \mu_I I_I, \quad (10)$$

$$\frac{dS_N}{dt} = -\frac{\beta}{N} S_N(I_I + I_N), \quad (11)$$

$$\frac{dE_N}{dt} = \frac{\beta}{N} S_N(I_I + I_N) - \kappa E_N, \quad (12)$$

$$\frac{dI_N}{dt} = \kappa E_N - \mu_N I_N, \quad (13)$$

$$\frac{dR_N}{dt} = \mu_N I_N. \quad (14)$$

The basic reproduction number under the 2-group model is given by

$$R_0 = \frac{\beta v}{\mu_I} + \frac{\beta(1-v)}{\mu_N}, \quad (15)$$

where the two terms represent the relative contributions from the cross-immune and immunologically naive groups respectively. The first term represents the expected number of infections generated by a single infected individual with previously acquired cross-immunity (assuming that no other individuals are infected at the time of introduction), scaled by the proportion of individuals who previously acquired cross-immunity. The second term represents the expected number of infections generated by a single infected individual who was previously immunologically naive, scaled by the proportion of individuals who were previously immunologically naive. See Supplementary Information Section S.2 for details of the calculation.

There is evidence that cross-immunity in the elderly population reduced the disease severity of the 2009 H1N1 pandemic virus (Miller et al., 2010; Chen, 2010). We assume that cross-immune individuals experience less severe disease and therefore typically recover more quickly than immunologically naive individuals (i.e. $1/\mu_I < 1/\mu_N$). Cross-immunity is therefore partial: it does not reduce disease completely. To isolate this effect alone on the predictability of epidemics, in this model it is assumed that cross-immune and naive individuals are otherwise identical. In particular, we assume that susceptible cross-immune individuals and susceptible naive individuals are equally likely to be infected.

We assume that only cases of severe disease (i.e. infected individuals who were previously immunologically naive) report infection, so that the number of recorded cases in week j is given by

$$C(j) = \int_{j-1}^j \kappa E_N dt. \quad (16)$$

Since we assume that only immunologically naive individuals report infection, the infectious period of immunologically naive individuals in the 2-group model is assumed to be identical to the infectious period in the 1-group model (i.e. $1/\mu_N = 1/\mu$).

Denoting the proportion of individuals in the population who are cross-immune by v , we have that $S_I + E_I + I_I + R_I = vN$ and $S_N + E_N + I_N + R_N = (1-v)N$, where $S_I + E_I + I_I + R_I + S_N + E_N + I_N + R_N = N$ is the total effective population size. In Table 1(b), we list the parameters that appear in Eqs. (7)–(14) and estimates of their values for the Japanese 2009 H1N1 epidemic (see also Sections 2.3 and 3.1).

2.3. Parameter estimation and forecasting

When fitting the models to each dataset in this study (either the data from the 2009 H1N1 epidemic in Japan or simulated future outbreak data), the transmission rate parameter, β , and effective population size, N , are estimated using Markov chain Monte Carlo (MCMC) with the Metropolis–Hastings algorithm (Hastings, 1970). All other parameters are assumed to be known. A likelihood function is used in which it is assumed that the differences between the data and model forecasts (where the differences are due to noise not accounted for in the models) are normally distributed, and that this noise scales with the square root of the size of the data (i.e. the number of cases). We estimate the noise scaling parameter, σ , in the likelihood function, by fitting each model to data from the 2009 Japan epidemic using a least squares approach (Cintrón-Arias et al., 2009). The value of σ is then fixed throughout this study. An analysis of the residuals (the scaled differences between each data point and the corresponding model values) is given in Supplementary Information Section S.3, justifying the square root noise scaling assumption. In each MCMC simulation, we perform 2×10^5 sampling iterations, discard the first 10^4 iterations

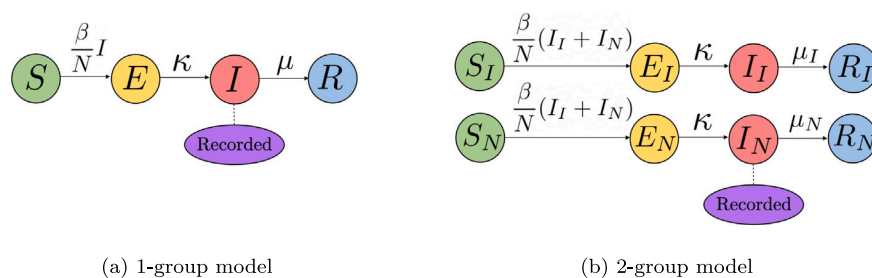


Fig. 1. Schematics illustrating the structures of the 1-group and 2-group models. In both cases, the population is compartmentalised into susceptible S , exposed (infected but not yet infectious) E , infectious I , and recovered R classes. In the 2-group model, cross-immune and immunologically naive individuals are distinguished between, and it is assumed that only infected naive individuals are recorded in case notification data (with perfect reporting). In the 1-group model, there is no distinction between cross-immune and immunologically naive individuals, and it is assumed that all infected individuals are recorded in case notification data (again with perfect reporting).

Table 1

Descriptions of parameters of the 1-group and 2-group models and estimates of their values for the Japanese 2009 H1N1 epidemic.

Parameter	Description	Values	Source
β	Transmission rate	1.644 Weeks ⁻¹	Estimated from data (Fig. 3)
N	Effective population size	3.072×10^7	Estimated from data (Fig. 3)
$1/\kappa$	Latent period	4/7 Weeks	Tuite et al. (2009)
$1/\mu$	Infectious period	1 Week	Tuite et al. (2009)
$C(0)$	Initial number of recorded cases	24073	Nishiura (2011)
(a): 1-group model			
Parameter	Description	Values	Source
β	Transmission rate	1.947 Weeks ⁻¹	Estimated from data (Fig. 3)
N	Total effective population size	4.660×10^7	Estimated from data (Fig. 3)
ν	Cross-immune fraction	0.3149	Estimate based on Mizumoto et al. (2013)
$1/\kappa$	Latent period	4/7 Weeks	Tuite et al. (2009)
$1/\mu_I$	Infectious period (cross-immune individuals)	3/7 Weeks	Estimate based on Fielding et al. (2013)
$1/\mu_N$	Infectious period (naive individuals)	1 Week	Tuite et al. (2009)
$C(0)$	Initial number of recorded cases	24073	Nishiura (2011)
(b): 2-group model			

as the ‘burn-in’ period and record every 100 iterations thereafter to reduce autocorrelation. Further details are given in Supplementary Information Section S.4.

When making forecasts in real-time after $t = m$ weeks of the epidemic, we calibrate our model forecasts with the observed data for weeks $0, 1, \dots, m$ of the epidemic, estimating model parameters using the method described above. To generate forecasts, we use these estimated model parameters and project the models forwards, starting in week m with initial conditions estimated based on the number of cases observed in weeks $0, 1, \dots, m$ (for details, see Supplementary Information Section S.5).

To compare different model forecasts throughout this study, we use the deviance information criterion (DIC) (Spiegelhalter et al., 2002). Details and DIC values are given in Supplementary Information Section S.6.

2.4. Modelling the size of the cross-immune population between epidemics

For the 2-group model, we assume that all immunologically naive individuals infected during an earlier epidemic acquired cross-immunity to antigenically related strains of the virus. Individuals who were cross-immune at the beginning of the 2009 epidemic, as well as individuals who were infected during the 2009 epidemic, are therefore assumed to be cross-immune. Furthermore, we assume that cross-immunity to related strains is lifelong (Yu et al., 2008). Using these assumptions, we can model the cross-immune fraction of the population over the years between the end of a first epidemic and the start of a future epidemic of a related strain of influenza accounting for population turnover (births and natural deaths). This is important, since the period between major influenza epidemics (or pandemics) is typically many years (Kilbourne, 2006). Further details are given in Supplementary Information Section

S.7. We seed the future epidemics by assuming that there are 10,000 recorded cases in the first week from which we generate forecasts (three orders of magnitude smaller than the estimated effective population sizes). When using the 2-group model to forecast future epidemics, we assume that the cross-immune fraction of the population is known exactly.

2.5. Approaches to forecasting a future epidemic

We use the 1-group and 2-group models (Fig. 1) to explore the accuracies of three different potential forecasting approaches (see Fig. 2 for a schematic illustrating these approaches). Specifically, we consider: generating a forecast using case notification data collected in real-time but without incorporating prior information from an earlier epidemic (Strategy 1); generating a forecast in advance of the epidemic, but using the previous epidemic as prior information (Strategy 2), and; generating a forecast using case notification data collected in real-time and also using the previous epidemic as prior information (Strategy 3).

3. Results

3.1. Fitting models to the 2009 H1N1 influenza epidemic in Japan

We fit the 1-group and 2-group models to data from the 2009 H1N1 influenza epidemic (Fig. 3(a), (b)). The shaded regions represent the 95% prediction credible intervals (CrIs) of epidemic trajectories based on the posterior distributions of fitting parameters β and N . Since the 1-group model does not account for infected but unrecorded individuals, the estimated parameters for the model represent a lower effective population size N than the for the 2-group model, while the estimated basic reproduction numbers R_0 are similar (Fig. 3(c),

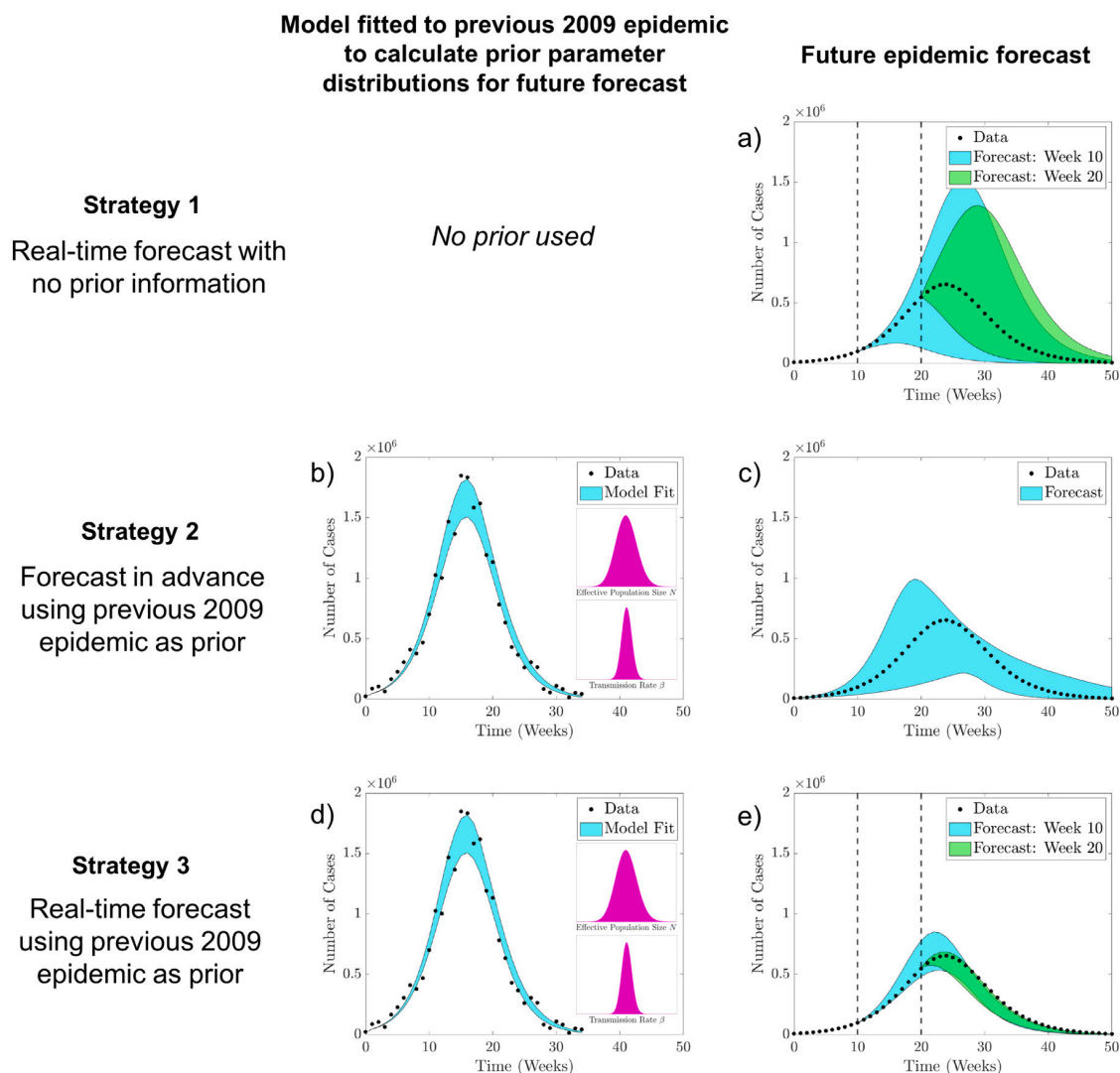


Fig. 2. Schematic illustrating the modelling approaches considered for forecasting the dynamics of a future epidemic. The shaded regions indicate the 95% CrIs of epidemic trajectories and forecasts. (a) Strategy 1: Forecast in real-time, using data from the ongoing epidemic and assuming no prior information about the fitting parameters. Cyan and green forecasts calibrated using data up to week 10 and 20 of the epidemic, respectively. (b)–(c) Strategy 2: Forecast in advance of a future epidemic using prior parameter distributions estimated by fitting the model to data from a previous epidemic (priors are shown as subfigures in (b)). (d)–(e) Strategy 3: Forecast in real-time, using data from the ongoing epidemic as well as prior parameter distributions estimated by fitting the model to data from a previous epidemic (priors are shown as subfigures in (d)). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

(d)). The mean estimated values of R_0 for the 1-group and 2-group models are 1.644 (95% CrI [1.627, 1.662]) and 1.597 (95% CrI [1.580, 1.612]) respectively, comparable with the interquartile range of 1.30 to 1.70 based on fifty-seven studies of the 2009 H1N1 pandemic strain (Biggerstaff et al., 2014). The values of the estimated parameters and their corresponding 95% CrIs are stated in Table 2. The numbers of recorded, unrecorded, and combined total weekly cases estimated using the 2-group model are shown in Supplementary Information Section S.8.

We used the mean estimated transmission rate and effective population size from the 2-group model to generate synthetic data for future epidemics. We assumed that, if the simulated epidemic takes place further into the future, then the cross-immune fraction of the population is lower due to deaths of cross-immune hosts and births of immunologically naive hosts (see Supplementary Information Section S.7).

Table 2

Mean values and CrIs of the fitting parameter posterior distributions, for the 1-group and 2-group models fitted to data from the 2009 H1N1 influenza epidemic in Japan.

Parameter	Mean	95% CrI
β	1.644	[1.627, 1.662]
N	3.072×10^7	$[2.905, 3.246] \times 10^7$

(a): 1-group model		
Parameter	Mean	95% CrI
β	1.947	[1.927, 1.966]
N	4.660×10^7	$[4.388, 4.938] \times 10^7$

(b): 2-group model		
--------------------	--	--

3.2. Forecasting an epidemic in real-time without prior information

If a major influenza epidemic were to occur, real-time forecasts could be made using live data describing the numbers of cases each week to update predictions. We considered a future epidemic due to a

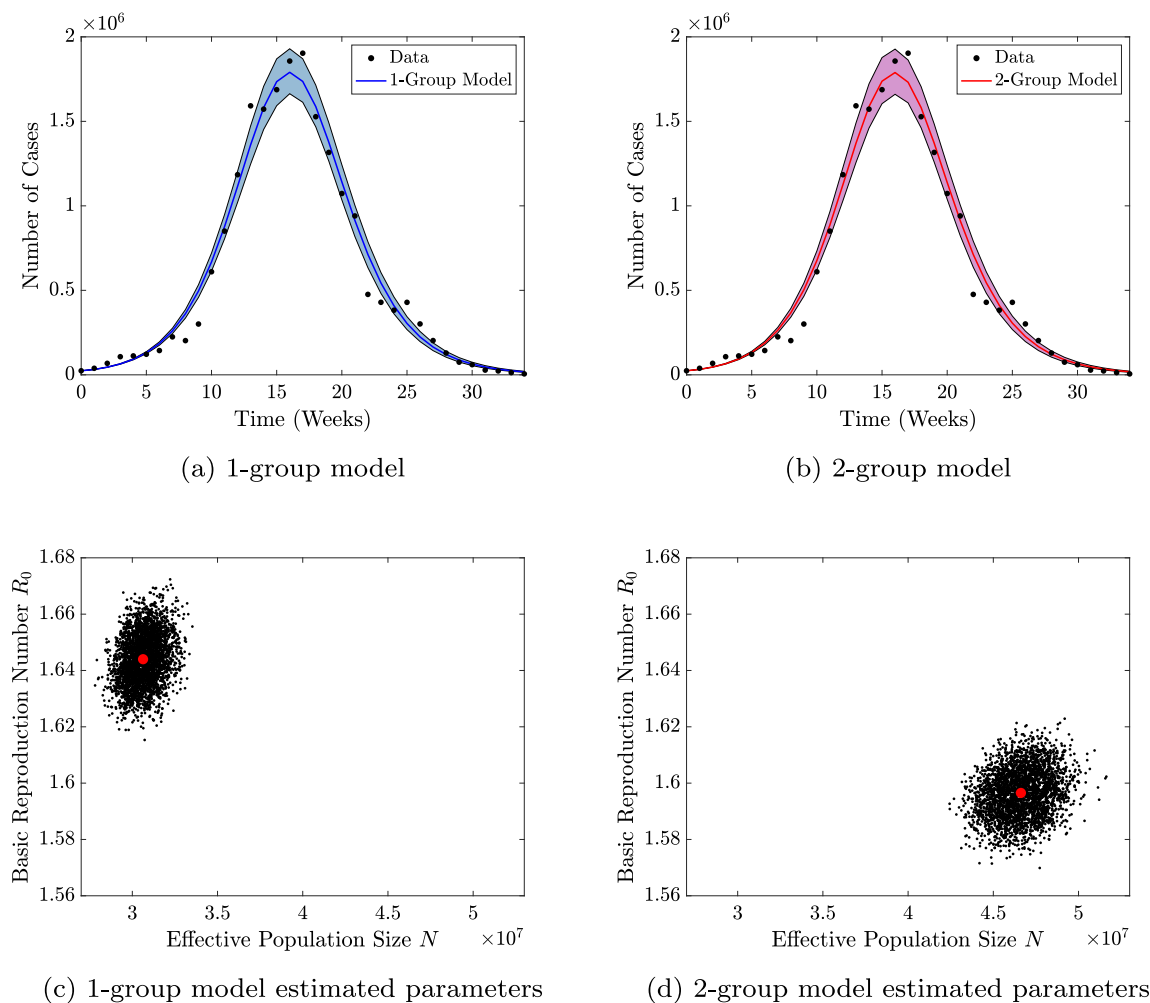


Fig. 3. (a)–(b): The 1-group and 2-group models fitted to data describing the numbers of new recorded cases each week from the 2009 H1N1 influenza epidemic in Japan, using the transmission rate β and effective population size N as fitting parameters. Solid coloured lines and shaded regions indicate the mean and 95% CrIs of epidemic trajectories based on the posterior distributions of the fitted parameters. DIC values are given in Supplementary Information Section S.6. (c)–(d): Scatter plots of the posterior distributions of R_0 (which is directly proportional to β) and N . Red dots represent the mean parameter estimates. Estimated parameters along with their CrIs are given in Table 2.

strain that is antigenically similar to the 2009 H1N1 virus occurring 25 years after the 2009 outbreak. We generated forecasts using the 1-group and 2-group models, at 10, 20, and 30 weeks after the first recorded cases. The results presented in Fig. 4 show how the model forecasts change as case notification data become available during the epidemic. The uncertainty in the forecasts of both models is large when predictions are generated early in the epidemic. By contrast, if forecasts are instead generated later, the remainder of the epidemic is predicted accurately. However, for the more accurate predictions obtained in week 30, the peak of the epidemic has already passed, and so accurate forecasting may be less practically useful. We conclude that it is challenging to generate practically useful forecasts of the remainder of an epidemic in real-time without prior information about the parameters governing pathogen transmission. This result motivates us to use information from previous epidemics when forecasting the dynamics of a future one; we investigate this further below.

3.3. Forecasting epidemics in advance

We considered informing forecasts of a future epidemic using gamma distributed estimates of the fitting parameters (β and N). The mean values of these gamma distributions, which were determined

from the fit to the 2009 epidemic data, are given in Supplementary Information Section S.9.

We used our models to generate forecasts of the dynamics of future influenza epidemics in Japan due to a related strain and occurring 25, 50, or 75 years after the 2009 epidemic (with no major epidemics occurring in each intervening period). The forecasts in Fig. 5 show that, if the 1-group model is used, the dynamics of a future epidemic are predicted to be identical to those of the 2009 epidemic, regardless of when it occurs. By contrast, for the 2-group model, a large proportion of the population would be cross-immune after the 2009 epidemic, resulting in a lower basic reproduction number. Consequently, if an antigenically related strain were to emerge 25 years later, the epidemic would be smaller than in 2009 (Fig. 5(a)–(b)). If the next major epidemic instead occurred 75 years later, a large proportion of the population would be immunologically naive to a related strain of the 2009 virus (because of population turnover due to births and deaths), resulting in a greater basic reproduction number. Hence, if all other factors were similar to those in 2009, the future epidemic would be larger than in 2009 (Fig. 5(e)–(f)).

Assuming that the 2-group model reflects the underlying epidemiology more accurately than the 1-group model, we conclude that forecasts generated using the 1-group model may not predict the dynamics of a future epidemic in which cross-immunity is present closely.

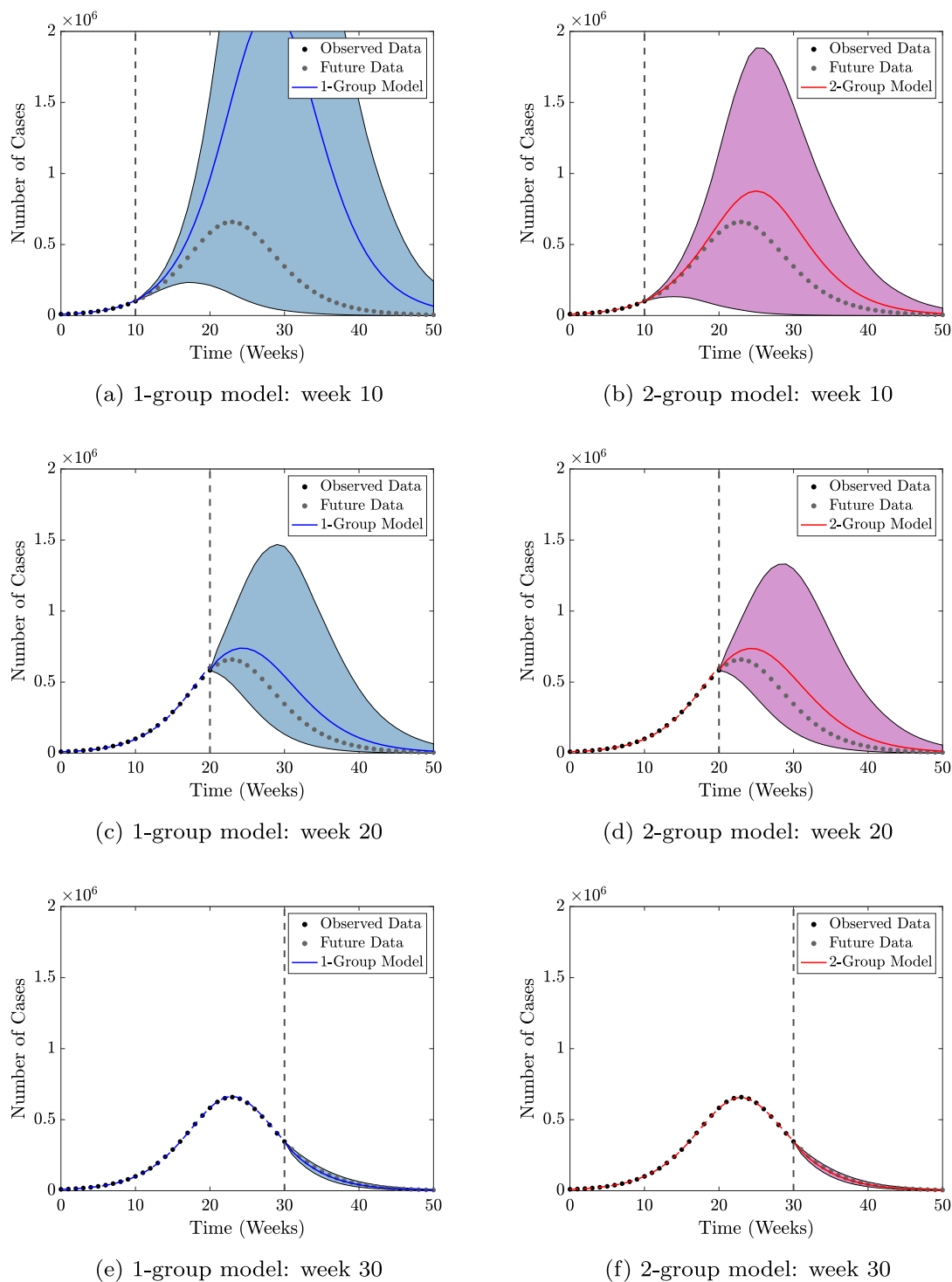


Fig. 4. Forecast Strategy 1: Real-time forecasts obtained from the 1-group and 2-group models for a future epidemic occurring 25 years after the 2009 epidemic, calibrated by fitting the model parameters β and N to data describing the weekly numbers of new cases, using uninformative uniform priors. Forecasts were made in weeks 10, 20, and 30 of the epidemic, using the observed data up until those times to estimate model parameters. Dashed vertical lines separate the calibration and forecasting periods. Other dashed lines indicate the mean of the epidemic trajectories in the model calibration period. Solid coloured lines and shaded regions indicate the mean and 95% CrIs of the forecasts, based on the posterior distributions of the parameters. The synthetic data were generated using the mean parameters of the 2-group model fitted to data from the 2009 epidemic. Uniform priors $N \in [10 \times 10^6, 128 \times 10^6]$ and $\beta \in [1, 3]$ were used in the estimation. DIC values are given in Table S1.

Forecasts using larger and smaller variances of the distributions of β and N are presented in Supplementary Information Section S.10. The dependence of a range of quantities obtained using the 2-group model (the cross-immune fraction at the start of the future epidemic, R_0 , total

number of recorded cases, duration, and size and timing of the epidemic peak) on the time period between the 2009 epidemic and the future epidemic is shown in Supplementary Information Section S.11. As seen in Fig. 5, the further into the future the next epidemic of a related

strain occurs then, all else being equal, the greater the total and peak number of cases and the shorter the duration of the epidemic. This is due to the smaller proportion of cross-immune individuals leading to a higher value of R_0 . We note that the impact of R_0 on the peak, final size and duration of an epidemic has been explored previously for simple epidemiological models (Weiss, 2013; Thompson et al., 2020a).

3.4. Forecasting an epidemic in real-time with prior information

As shown above, real-time forecasts generated during a future influenza epidemic, without informative priors for the fitting parameters, can be very uncertain (e.g. Fig. 4(a)–(b)). We therefore also considered using priors to inform the fitted parameter values, so that real-time forecasts are based on both historical data (from the 2009 epidemic) and live data from the ongoing outbreak. Gamma distributed priors were set for β and N , with mean values based on the parameter estimates obtained using the data from the 2009 epidemic.

As before, we considered a scenario in which an epidemic occurs 25 years after the 2009 epidemic. In turn, predictions were made 0, 10, 20, and 30 weeks after the start of the future epidemic (Fig. 6(a)–(b), (c)–(e), (f)–(h) and (i)–(k), respectively).

We considered using priors of different widths to inform the epidemic forecasts (Fig. 6 and Supplementary Information Section S.12). Under the baseline variance considered, when the 1-group model was used, there was sometimes a discrepancy between the calibrated model trajectory and the observed epidemic data (e.g. left part of Fig. 6(f)). For that reason, we also show epidemic forecasts obtained using a wider prior (Fig. 6(d), (g) and (j)), so that the fitted model reflects the observed data more accurately. When the 1-group model was used, forecasts were either inaccurate (when a prior with low variance was used; e.g. Fig. 6(c)) or imprecise (when a prior with higher variance was used; e.g. Fig. 6(d)). In either case, as shown quantitatively by the DIC values in Table S1, our main conclusion is unchanged: predictions improved when forecasts were generated using the more epidemiologically realistic 2-group model.

4. Discussion

Influenza epidemics and pandemics place a significant burden on healthcare systems throughout the world (Monto, 2004). Exposure to an influenza virus confers cross-immunity to antigenically related strains (Kreijtz et al., 2011), however cross-immunity is not frequently included explicitly in influenza forecasting models (Baguelin et al., 2013; Rajaram et al., 2017). In this study, we have investigated whether or not it is necessary to account for cross-immunity when generating forecasts during influenza epidemics and, moreover, whether data from previous epidemics can be used to improve forecast accuracy.

We have considered two different mathematical models that describe pathogen transmission: a 2-group model in which immunologically naive and cross-immune individuals are differentiated between and a 1-group model in which all individuals are assumed to have similar characteristics. We parameterised each model using data describing the estimated numbers of infected individuals seeking medical attention per week in Japan during the 2009 H1N1 influenza pandemic (Nishiura, 2011; Omori and Nishiura, 2011). We then considered a scenario in which we predict the dynamics of an epidemic of an antigenically related strain occurring 25 years after the 2009 pandemic. When no prior knowledge about the values of the transmission rate and effective population size is assumed, neither model can predict future epidemiological dynamics in real-time early in the epidemic with any certainty. Previous research has also indicated that significant uncertainty exists when ordinary differential equation models are used to generate real-time forecasts early in influenza epidemics (Hall et al., 2007).

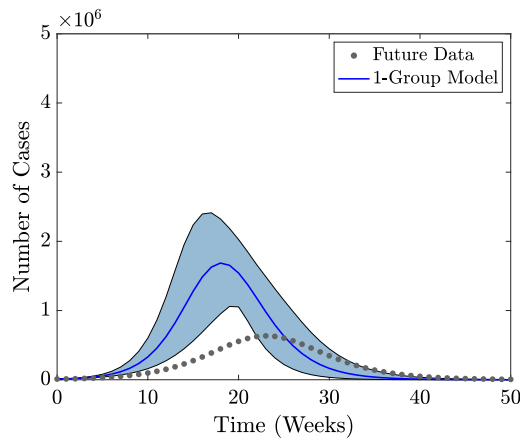
We then considered the effect of the time period between epidemics on the accuracy of epidemic forecasts made in advance of the future

epidemic. Immediately after the 2009 epidemic, a large fraction of the population will have been cross-immune to strains that are related to the 2009 H1N1 virus. This fraction reduces as the time since the 2009 epidemic increases, as individuals (a large proportion of whom may have acquired cross-immunity) die and immunologically naive individuals are born. Consequently, the expected size of a future epidemic of a related strain increases the further into the future at which it occurs. If we neglect cross-immunity by using the 1-group model, we predict the same epidemic dynamics as in 2009, regardless of when the future epidemic occurs. This is because changing numbers of cross-immune individuals are not accounted for in that model. This could result in overestimation of the size of a future epidemic, if the cross-immune fraction was greater than the baseline cross-immunity in 2009 (i.e. if the future epidemic occurred soon after 2009), or underestimation of the epidemic size if the cross-immune fraction was lower than in 2009 (i.e. if the future epidemic occurred long after 2009).

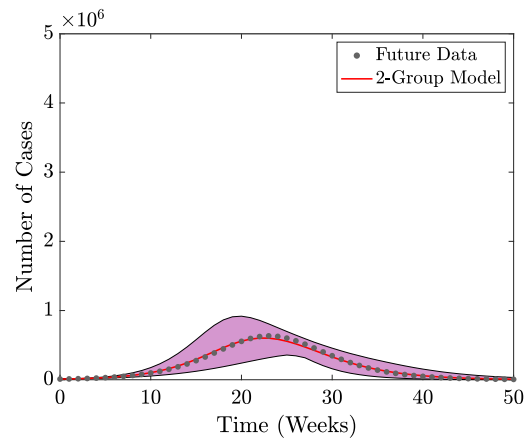
Finally, we considered incorporating knowledge of parameters from the 2009 epidemic, in combination with case notification data obtained in real-time during the future epidemic, to update forecasts as the future epidemic is ongoing. When the 1-group model was used to make forecasts, we found that priors with significant uncertainty had to be used to calibrate the model trajectories to the data. Although these priors, set based on the model fits to the 2009 epidemic data, led to improved forecasts compared to when no prior information was used, the forecasts were still inaccurate or uncertain when made before the peak of the epidemic. By contrast, the 2-group model was able to generate more accurate forecasts. This is because cross-immunity is accounted for in the 2-group model, and the priors reflect the values of the true underlying parameters more accurately. We conclude that, to forecast the long-term dynamics of major influenza epidemics caused by strains that are antigenically related to those responsible for previous epidemics, a model that accounts for cross-immunity should be used, and data obtained during previous epidemics should be taken into account. Additionally, we note that short-term forecasts on the order of one to four weeks – a time horizon used by the US Centers for Disease Control when constructing seasonal influenza forecasts (Biggerstaff et al., 2016, 2018; Centers for Disease Control and Prevention (CDC), 2020) – are also more accurate when generated using the 2-group model (see Fig. 6).

We note that there may be instances in which the underlying epidemiology of an outbreak is not well-known, or where a complex model cannot be parameterised due to ineffective surveillance or limited data availability (Gibbons et al., 2014; Thompson et al., 2020d). In such cases, using a simple model akin to the 1-group model used in this study may be the only possible option for making predictions. Although forecasts made using a simple model may lack precision, they could still be useful to policy makers during an epidemic. For example, the 1-group model was able to predict the duration of the epidemic accurately when using data from the first 10 weeks of the outbreak, even if it could not predict the trajectory of the entire epidemic perfectly. Simple models may be particularly useful for short-term forecasts (Funk et al., 2019). By constantly updating model predictions using newly acquired data, and refining the modelling framework when it is unable to replicate observed data, simple models that do not describe the underlying epidemiology of an epidemic fully may still provide useful insights.

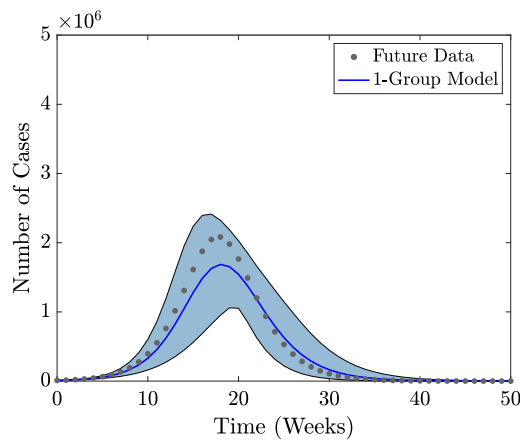
We investigated whether or not it is necessary to include a particular source of heterogeneity (i.e. cross-immunity) when constructing influenza epidemic forecasts. Our analysis used case notification data from Japan collected during the 2009 influenza pandemic. Although the exact dataset underlying our analysis was not central to our main conclusion, it would be possible to replicate our study using any number of other influenza datasets (Donaldson et al., 2009; Pourbohloul et al., 2009; Earn, 2012; National Institute of Infectious Diseases, 2020). Going forwards, we could also consider how accounting for cross-immunity would affect forecasts of seasonal epidemics, in which case



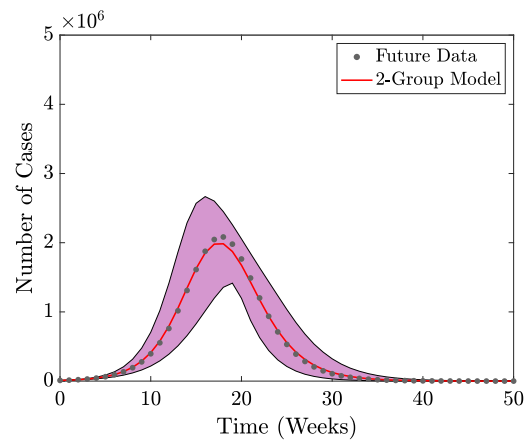
(a) 1-group model: 25 years later



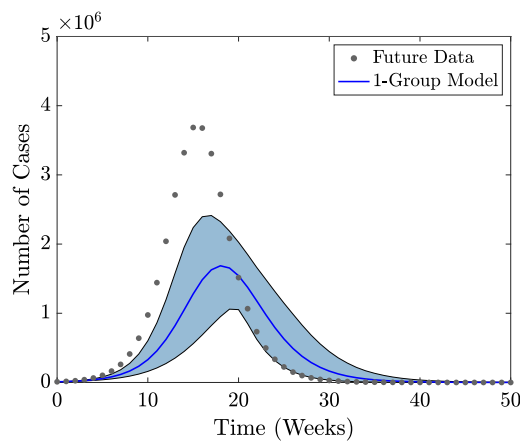
(b) 2-group model: 25 years later



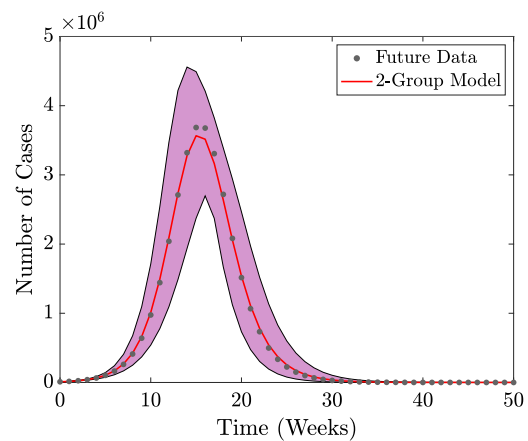
(c) 1-group model: 50 years later



(d) 2-group model: 50 years later



(e) 1-group model: 75 years later



(f) 2-group model: 75 years later

Fig. 5. Forecast Strategy 2: Forecasts obtained at the start of future epidemics using the 1-group and 2-group models, for epidemics occurring 25, 50, and 75 years after the 2009 epidemic. Forecasts were generated by assuming that the parameters β and N follow a gamma distribution, the parameters of which were determined by fitting the models to data from the previous 2009 epidemic (Supplementary Information Section S.9). Solid coloured lines and shaded regions indicate the mean and 95% CrIs of the forecasts. The synthetic data were generated using the mean parameters of the 2-group model fitted to data from the 2009 epidemic. DIC values are given in Table S1.

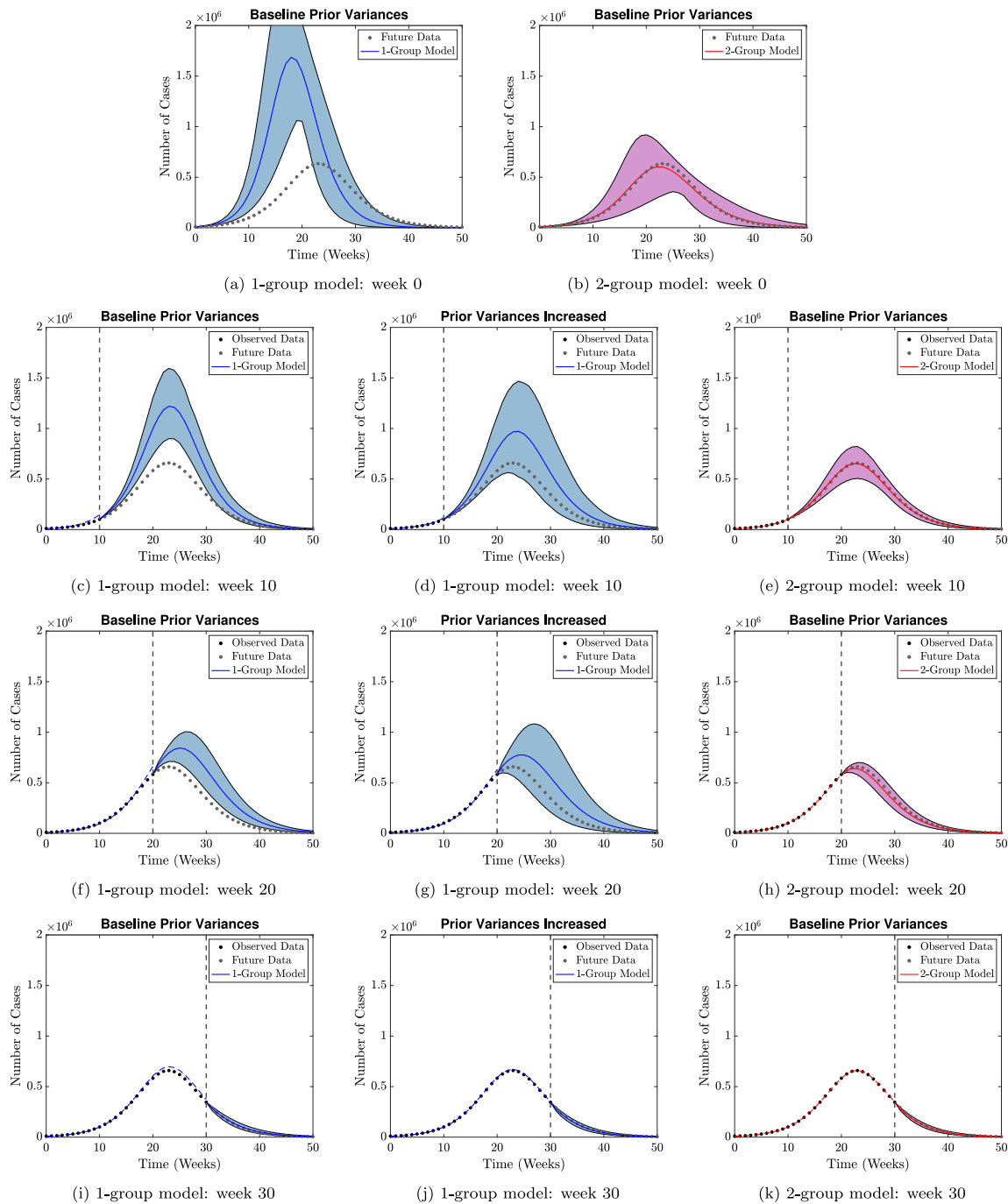


Fig. 6. Forecast Strategy 3: Forecasts obtained using the 1-group and 2-group models during a future epidemic occurring 25 years after the 2009 epidemic, at the start of the epidemic ((a)–(b)), and using case notification data obtained in real-time during the epidemic ((c)–(k)). Gamma distributed priors are prescribed for the fitting parameters β and N , the parameters of which were determined by fitting the models to data from the previous 2009 epidemic (Supplementary Information Section S.9). Forecasts are generated in weeks 10, 20, and 30 of the epidemic, using the observed data up until those times to estimate model parameters. Dashed vertical lines separate the calibration and forecasting periods. Other dashed lines indicate the mean of the epidemic trajectories in the model calibration period. Solid coloured lines and shaded regions indicate the mean and 95% CrIs of the forecasts, based on the posterior distributions of the parameters. The synthetic data were generated using the mean parameters of the 2-group model fitted to data from the 2009 epidemic. DIC values are given in Table S1.

the particular strains driving dynamics in different seasons would need to be considered (Petrova and Russell, 2017). Consideration of whether and how forecasts of epidemics of pandemic influenza may be influenced by circulating seasonal strains (Fox et al., 2017) is a target for future work. We could extend our models to investigate whether or not other sources of heterogeneity should be accounted for. For instance, age-structure could be incorporated, taking into account the effects

of immune imprinting (Gostic et al., 2016, 2019). Additionally, we could include spatial heterogeneity by partitioning the population into distinct geographical regions or introducing a contact network structure (Meyers et al., 2006; Volz and Meyers, 2007; Ohkusa et al., 2009; Volz et al., 2011; Miller and Kiss, 2014). Another potential avenue for further investigation may be to incorporate transmission models that account for cross-immunity into a phylodynamic framework to

predict which strains of influenza may emerge and the dynamics of their respective epidemics (Grenfell, 2004).

In this study, we have considered the simple example of binary immunity (i.e. a class of cross-immune individuals and a class of immunologically naive individuals were included). In reality, the landscape of immunity within a population is much more complex, with individuals exhibiting varying levels of immunity based on their life-time history of exposure to influenza viruses (Laurie et al., 2010; Gostic et al., 2016, 2019) and their age (Lambert et al., 2012; Haq and McElhaney, 2014). Underlying health conditions such as obesity and heart disease may also impact individuals' immune responses (Hui et al., 2006; Honce and Schultz-Cherry, 2019). Nonetheless, even the simple framework that we considered allowed us to show that cross-immunity should be accounted for when making influenza epidemic forecasts. Future work could involve models containing numerous classes to account for population heterogeneities or detailed infection dynamics (Thompson and Hart, 2018; Hart et al., 2020), or even a continuum level of immunity within the population based on the above-mentioned factors (Andreassen et al., 1997; Martcheva and Pilyugin, 2006).

In the 2-group model considered here, we assumed, as in Reichert et al. (2012), that immunoprotection does not prevent infection, but instead alters its consequences. Different types of cross-immunity to related strains of a virus can be considered. For example, cross-immunity could instead be assumed to reduce an individual's susceptibility to infection (Hill et al., 2019; Thompson et al., 2019). Future work could consider whether or not the results of this study hold when different assumptions are made about the precise effects of cross-immunity. In the 2-group model, in order to study the effects of cross-immunity in as simple a setting as possible, we assumed that infected individuals who did not seek medical attention had cross-immune protection against the virus. However, a range of factors (including age and behaviour) affect whether or not, and how quickly, an individual seeks medical attention (Thompson, 2020c). This could be built into the underlying modelling framework considered here, although additional data would be needed to parameterise the resulting model. Other additions to the 1- and 2-group models could also be considered. For example, the wide range of interventions that are implemented during outbreaks could be included in the models explicitly (Gani et al., 2005; Longini Jr. et al., 2005; Backer et al., 2019).

Despite its simplicity, our approach has enabled us to demonstrate that including cross-immunity in influenza epidemic forecasting models can lead to more accurate forecasts. Cross-immunity due to previous infections has been shown to play a major role in the dynamics of influenza epidemics, with clear evidence emerging from analyses of data from the 1918 Spanish Flu pandemic (Taubenberger and Morens, 2006) and the 2009 H1N1 influenza pandemic (Hancock et al., 2009). Cross-immunity will contribute to shaping the dynamics of future influenza epidemics. Consideration of cross-immunity by epidemiological modellers is therefore of obvious public health importance.

Funding

This publication is based on work supported by the EPSRC, UK Centre for Doctoral Training in Industrially Focused Mathematical Modelling (EP/L015803/1) in collaboration with Biosensors Beyond Borders Ltd. (RS-P). It was also funded by Christ Church (Oxford), UK via a Junior Research Fellowship (RNT).

CRediT authorship contribution statement

Rahil Sachak-Patwa: Conceived the study, Conducted the analysis, Writing - review & editing. **Helen M. Byrne:** Conceived the study, Supervised the research, Writing - review & editing. **Robin N. Thompson:** Conceived the study, Supervised the research, Writing - review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Thanks to Erin I Lafferty and Claude J Schmit (Biosensors Beyond Borders Ltd.), Craig P Thompson (University of Oxford), and members of the Wolfson Centre for Mathematical Biology (University of Oxford) for helpful discussions about this research.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at <https://doi.org/10.1016/j.epidem.2020.100432>.

References

- Anderson, R.M., May, R.M., 1991. *Infectious Diseases of Humans: Dynamics and Control*. Oxford University Press.
- Andreassen, V., Lin, J., Levin, S.A., 1997. The dynamics of cocirculating influenza strains conferring partial cross-immunity. *J. Math. Biol.* 35 (7), 825–842. <http://dx.doi.org/10.1007/s002850050079>.
- Backer, J., Wallinga, J., Meijer, A., Donker, G., Van der Hoek, W., Van Boven, M., 2019. The impact of influenza vaccination on infection, hospitalisation and mortality in the Netherlands between 2003 and 2015. *Epidemics* 26, 77–85. <http://dx.doi.org/10.1016/j.epidem.2018.10.001>.
- Baguelin, M., Flasche, S., Camacho, A., Demiris, N., Miller, E., Edmunds, W.J., 2013. Assessing optimal target populations for influenza vaccination programmes: an evidence synthesis and modelling study. *PLoS Med.* 10 (10), e1001527. <http://dx.doi.org/10.1371/journal.pmed.1001527>.
- Bandaranayake, D., Huang, Q.S., Bissielo, A., Wood, T., Mackereth, G., Baker, M.G., et al., 2010. Risk factors and immunity in a nationally representative population following the 2009 influenza A(H1N1) pandemic. *PLoS One* 5 (10), e13211. <http://dx.doi.org/10.1371/journal.pone.0013211>.
- Bettencourt, L.M., Ribeiro, R.M., 2008. Real time bayesian estimation of the epidemic potential of emerging infectious diseases. *PLoS One* 3 (5), e2185. <http://dx.doi.org/10.1371/journal.pone.0002185>.
- Biggerstaff, M., Alper, D., Dredze, M., Fox, S., Fung, I.C.-H., Hickmann, K.S., et al., 2016. Results from the centers for disease control and prevention's predict the 2013–2014 Influenza Season Challenge. *BMC Infect. Dis.* 16 (1), <http://dx.doi.org/10.1186/s12879-016-1669-x>.
- Biggerstaff, M., Cauchemez, S., Reed, C., Gambhir, M., Finelli, L., 2014. Estimates of the reproduction number for seasonal, pandemic, and zoonotic influenza: a systematic review of the literature. *BMC Infect. Dis.* 14 (1), <http://dx.doi.org/10.1186/1471-2334-14-480>.
- Biggerstaff, M., Johansson, M., Alper, D., Brooks, L.C., Chakraborty, P., Farrow, D.C., et al., 2018. Results from the second year of a collaborative effort to forecast influenza seasons in the United States. *Epidemics* 24, 26–33. <http://dx.doi.org/10.1016/j.epidem.2018.02.003>.
- Bouvier, N.M., Palese, P., 2008. The biology of influenza viruses. *Vaccine* 26, D49–D53. <http://dx.doi.org/10.1016/j.vaccine.2008.07.039>.
- Centers for Disease Control and Prevention (CDC), 2020. How CDC uses flu forecasting. <https://www.cdc.gov/flu/weekly/flu-sight/how-flu-forecasting.htm>. (Accessed 27 October 2020).
- Chen, M.I.C., 2010. 2009 influenza A(H1N1) seroconversion rates and risk factors among distinct adult cohorts in Singapore. *JAMA* 303 (14), 1383. <http://dx.doi.org/10.1001/jama.2010.404>.
- Chen, S.-C., Liao, C.-M., 2008. Modelling control measures to reduce the impact of pandemic influenza among schoolchildren. *Epidemiol. Infect.* 136 (8), 1035–1045. <http://dx.doi.org/10.1017/S0950268807009284>.
- Chowell, G., Nishiura, H., Bettencourt, L.M., 2006. Comparative estimation of the reproduction number for pandemic influenza from daily case notification data. *J. R. Soc. Interface* 4 (12), 155–166. <http://dx.doi.org/10.1098/rsif.2006.0161>.
- Chowell, G., Viboud, C., Wang, X., Bertozzi, S.M., Miller, M.A., 2009. Adaptive vaccination strategies to mitigate pandemic influenza: Mexico as a case study. *PLoS One* 4 (12), e8164. <http://dx.doi.org/10.1371/journal.pone.0008164>.
- Christman, M.C., Kedwaii, A., Xu, J., Donis, R.O., Lu, G., 2011. Pandemic (H1N1) 2009 virus revisited: an evolutionary retrospective. *Infect. Genet. Evol.* 11 (5), 803–811. <http://dx.doi.org/10.1016/j.meegid.2011.02.021>.
- Cintrón-Arias, A., Castillo-Chávez, C., Bettencourt, L.M., Lloyd, A.L., Banks, H., 2009. The estimation of the effective reproductive number from disease outbreak data. *Math. Biosci. Eng.* 6 (2), 261–282. <http://dx.doi.org/10.3934/mbe.2009.6.261>.

- Couch, R.B., Kasel, J.A., Gerin, J.L., Schulman, J.L., Kilbourne, E.D., 1974. Induction of partial immunity to influenza by a neuraminidase-specific vaccine. *J. Infect. Dis.* 129 (4), 411–420. <http://dx.doi.org/10.1093/infdis/129.4.411>.
- Davies, N.G., Kucharski, A.J., Eggo, R.M., Gimma, A., Edmunds, W.J., Jombart, T., et al., 2020. 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. [http://dx.doi.org/10.1016/s2468-2667\(20\)30133-x](http://dx.doi.org/10.1016/s2468-2667(20)30133-x).
- Donaldson, L.J., Rutter, P.D., Ellis, B.M., Greaves, F.E.C., Mytton, O.T., Pebody, R.G., Yardley, I.E., 2009. Mortality from pandemic A/H1N1 2009 influenza in England: public health surveillance study. *BMJ* 339 (dec10 1), b5213. <http://dx.doi.org/10.1136/bmj.b5213>.
- Du, X., King, A.A., Woods, R.J., Pascual, M., 2017. Evolution-informed forecasting of seasonal influenza A (H3N2). *Sci. Transl. Med.* 9 (413), eaan5325. <http://dx.doi.org/10.1126/scitranslmed.aan5325>.
- Dushoff, J., Plotkin, J.B., Viboud, C., Simonsen, L., Miller, M., Loeb, M., David, J., 2007. Vaccinating to protect a vulnerable subpopulation. *PLoS Med.* 4 (5), e174. <http://dx.doi.org/10.1371/journal.pmed.0040174>.
- Earn, D.J., 2012. Effects of school closure on incidence of pandemic influenza in Alberta, Canada. *Ann. Intern. Med.* 156 (3), 173. <http://dx.doi.org/10.7326/0003-4819-156-3-201207070-00005>.
- Ellebedy, A., Ducatez, M., Duan, S., Stigger-Rosser, E., Rubrum, A., Govorkova, E., et al., 2011. Impact of prior seasonal influenza vaccination and infection on pandemic A (H1N1) influenza virus replication in ferrets. *Vaccine* 29 (17), 3335–3339. <http://dx.doi.org/10.1016/j.vaccine.2010.08.067>.
- Ellebedy, A.H., Fabrizio, T.P., Kayali, G., Oguin, T.H., Brown, S.A., Rehg, J., et al., 2010. Contemporary seasonal influenza A (H1N1) virus infection primes for a more robust response to split inactivated pandemic influenza A (H1N1) virus vaccination in ferrets. *Clin. Vaccine Immunol.* 17 (12), 1998–2006. <http://dx.doi.org/10.1128/cvi.00247-10>.
- Epstein, S., Mispelon, J., Lawson, C., Subbarao, E., Connors, M., Murphy, B., 1993. Beta 2-microglobulin-deficient mice can be protected against influenza A infection by vaccination with vaccinia-influenza recombinants expressing hemagglutinin and neuraminidase. *J. Immunol.* 150 (12), 5484–5493.
- Fang, Y., Banner, D., Kelvin, A.A., Huang, S.S.H., Paige, C.J., Corfe, S.A., et al., 2011. Seasonal H1N1 influenza virus infection induces cross-protective pandemic H1N1 virus immunity through a CD8-independent, B cell-dependent mechanism. *J. Virol.* 86 (4), 2229–2238. <http://dx.doi.org/10.1128/jvi.05540-11>.
- Ferguson, N.M., Cummings, D.A., Fraser, C., Cajka, J.C., Cooley, P.C., Burke, D.S., 2006. Strategies for mitigating an influenza pandemic. *Nature* 442 (7101), 448. <http://dx.doi.org/10.1038/nature04795>.
- Fielding, J.E., Kelly, H.A., Mercer, G.N., Glass, K., 2013. Systematic review of influenza A(H1N1)pdm09 virus shedding: duration is affected by severity, but not age. *Influenza Other Respir. Viruses* 8 (2), 142–150. <http://dx.doi.org/10.1111/irv.12216>.
- Fox, S.J., Miller, J.C., Meyers, L.A., 2017. Seasonality in risk of pandemic influenza emergence. *PLoS Comput. Biol.* 13 (10), e1005749. <http://dx.doi.org/10.1371/journal.pcbi.1005749>.
- Fraser, C., Donnelly, C.A., Cauchemez, S., Hanage, W.P., Van Kerkhove, M.D., Hollingsworth, T.D., et al., 2009. Pandemic potential of a strain of influenza A (H1N1): early findings. *Science* 324 (5934), 1557–1561. <http://dx.doi.org/10.1126/science.1176062>.
- Frost, S.D., Pybus, O.G., Gog, J.R., Viboud, C., Bonhoeffer, S., Bedford, T., 2015. Eight challenges in phylogenetic inference. *Epidemics* 10, 88–92. <http://dx.doi.org/10.1016/j.epidem.2014.09.001>.
- Funk, S., Camacho, A., Kucharski, A.J., Lowe, R., Eggo, R.M., Edmunds, W.J., 2019. Assessing the performance of real-time epidemic forecasts: A case study of Ebola in the Western Area region of Sierra Leone, 2014–15. *PLoS Comput. Biol.* 15 (2), e1006785. <http://dx.doi.org/10.1371/journal.pcbi.1006785>.
- Gani, R., Hughes, H., Fleming, D., Griffin, T., Medlock, J., Leach, S., 2005. Potential impact of antiviral drug use during influenza pandemic. *Emerg. Infect. Diseases* 11 (9), 1355. <http://dx.doi.org/10.3201/eid1109.041344>.
- Gart, J.J., 1968. The mathematical analysis of an epidemic with two kinds of susceptibles. *Biometrics* 24 (3), 557. <http://dx.doi.org/10.2307/2528318>.
- Gibbons, C.L., Mangen, M.-J.J., Plass, D., Havelaar, A.H., Brooke, R.J., Kramarz, P., et al., 2014. Measuring underreporting and under-ascertainment in infectious disease datasets: a comparison of methods. *BMC Public Health* 14 (1), 147. <http://dx.doi.org/10.1186/1471-2458-14-147>.
- Glasser, J., Taneri, D., Feng, Z., Chuang, J.-H., Tüll, P., Thompson, W., et al., 2010. Evaluation of targeted influenza vaccination strategies via population modeling. *PLoS One* 5 (9), e12777. <http://dx.doi.org/10.1371/journal.pone.0012777>.
- Gostic, K.M., Ambrose, M., Worobey, M., Lloyd-Smith, J.O., 2016. Potent protection against H5N1 and H7N9 influenza via childhood hemagglutinin imprinting. *Science* 354 (6313), 722–726. <http://dx.doi.org/10.1126/science.aag1322>.
- Gostic, K.M., Bridge, R., Brady, S., Viboud, C., Worobey, M., Lloyd-Smith, J.O., 2019. Childhood immune imprinting to influenza A shapes birth year-specific risk during seasonal H1N1 and H3N2 epidemics. *PLoS Pathog.* 15 (12), <http://dx.doi.org/10.1371/journal.ppat.1008109>.
- Grebe, K.M., Yewdell, J.W., Bennink, J.R., 2008. Heterosubtypic immunity to influenza A virus: where do we stand? *Microb. Infect.* 10 (9), 1024–1029. <http://dx.doi.org/10.1016/j.micinf.2008.07.002>.
- Grenfell, B.T., 2004. Unifying the epidemiological and evolutionary dynamics of pathogens. *Science* 303 (5656), 327–332. <http://dx.doi.org/10.1126/science.1090727>.
- Hall, I.M., Gani, R., Hughes, H.E., Leach, S., 2007. Real-time epidemic forecasting for pandemic influenza. *Epidemiol. Infect.* 135 (3), 372–385. <http://dx.doi.org/10.1017/s0950268806007084>.
- Hancock, K., Veguilla, V., Lu, X., Zhong, W., Butler, E.N., Sun, H., et al., 2009. Cross-reactive antibody responses to the 2009 pandemic H1N1 influenza virus. *N. Engl. J. Med.* 361 (20), 1945–1952. <http://dx.doi.org/10.1056/NEJMoa0906453>.
- Haq, K., McElhaney, J.E., 2014. Immunosenescence: influenza vaccination and the elderly. *Curr. Opin. Immunol.* 29, 38–42. <http://dx.doi.org/10.1016/j.coi.2014.03.008>.
- Hardelid, P., Andrews, N., Hoschler, K., Stanford, E., Baguelin, M., Waight, P., et al., 2010. Assessment of baseline age-specific antibody prevalence and incidence of infection to novel influenza A/H1N1 2009. *Health Technol. Assess.* 14 (55), 115–192. <http://dx.doi.org/10.3310/hta14550-03>.
- Hart, W.S., Maini, P.K., Yates, C.A., Thompson, R.N., 2020. A theoretical framework for transitioning from patient-level to population-scale epidemiological dynamics: influenza A as a case study. *J. R. Soc. Interface* 17 (166), 20200230. <http://dx.doi.org/10.1098/rsif.2020.0230>.
- Hastings, W.K., 1970. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* 57 (1), 97–109. <http://dx.doi.org/10.1093/biomet/57.1.97>.
- He, D., Lui, R., Wang, L., Tse, C.K., Yang, L., Stone, L., 2015. Global spatio-temporal patterns of influenza in the post-pandemic era. *Sci. Rep.* 5 (1), <http://dx.doi.org/10.1038/srep11013>.
- Hill, E.M., Petrou, S., de Lusignan, S., Yonova, I., Keeling, M.J., 2019. Seasonal influenza: Modelling approaches to capture immunity propagation. *PLoS Comput. Biol.* 15 (10), 1–26. <http://dx.doi.org/10.1371/journal.pcbi.1007096>.
- Honce, R., Schultz-Cherry, S., 2019. Impact of obesity on influenza A virus pathogenesis, immune response, and evolution. *Front. Immunol.* 10, <http://dx.doi.org/10.3389/fimmu.2019.01071>.
- Huang, S.S.H., Lin, Z., Banner, D., León, A.J., Paquette, S.G., Rubin, B., et al., 2013. Immunity toward H1N1 influenza hemagglutinin of historical and contemporary strains suggests protection and vaccine failure. *Sci. Rep.* 3 (1), <http://dx.doi.org/10.1038/srep01698>.
- Hui, S., Chu, L., Peiris, J., Chan, K., Chu, D., Tsui, W., 2006. Immune response to influenza vaccination in community-dwelling Chinese elderly persons. *Vaccine* 24 (25), 5371–5380. <http://dx.doi.org/10.1016/j.vaccine.2006.04.032>.
- Johnson, N.P., Mueller, J., 2002. Updating the accounts: global mortality of the 1918–1920 “Spanish” influenza pandemic. *Bull. Hist. Med.* 76 (1), 105–115. <http://dx.doi.org/10.1353/bhm.2002.0022>.
- Kamigaki, T., Oshitani, H., 2009. Epidemiological characteristics and low case fatality rate of pandemic (H1N1) 2009 in Japan. *PLoS Curr.* 1, RRN1139. <http://dx.doi.org/10.1371/currents.rn1139>.
- Kash, J.C., Qi, L., Dugan, V.G., Jagger, B.W., Hrabal, R.J., Memoli, M.J., et al., 2010. Prior infection with classical swine H1N1 influenza viruses is associated with protective immunity to the 2009 pandemic H1N1 virus. *Influenza Other Respir. Viruses* 4 (3), 121–127. <http://dx.doi.org/10.1111/j.1750-2659.2010.00132.x>.
- Kelly, H., Peck, H.A., Laurie, K.L., Wu, P., Nishiura, H., Cowling, B.J., 2011. The age-specific cumulative incidence of infection with pandemic influenza H1N1 2009 was similar in various countries prior to vaccination. *PLoS One* 6 (8), e21828. <http://dx.doi.org/10.1371/journal.pone.0021828>.
- Kilbourne, E.D., 2006. Influenza pandemics of the 20th century. *Emerg. Infect. Diseases* 12 (1), 9–14. <http://dx.doi.org/10.3201/eid1201.051254>.
- Kim, H., Webster, R.G., Webby, R.J., 2018. Influenza virus: dealing with a drifting and shifting pathogen. *Viral Immunol.* 31 (2), 174–183. <http://dx.doi.org/10.1089/vim.2017.0141>.
- Klepac, P., Kissler, S., Gog, J., 2018. Contagion! The BBC Four Pandemic—The model behind the documentary. *Epidemics* 24, 49–59. <http://dx.doi.org/10.1016/j.epidem.2018.03.003>.
- Koelle, K., Cobey, S., Grenfell, B., Pascual, M., 2006. Epochal evolution shapes the phylodynamics of inter-pandemic influenza A (H3N2) in humans. *Science* 314 (5807), 1898–1903. <http://dx.doi.org/10.1126/science.1132745>.
- Krammer, F., 2019. The human antibody response to influenza A virus infection and vaccination. *Nat. Rev. Immunol.* 19 (6), 383–397. <http://dx.doi.org/10.1038/s41577-019-0143-6>.
- Kreijtz, J., Fouchier, R., Rimmelzwaan, G., 2011. Immune responses to influenza virus infection. *Virus Res.* 162 (1–2), 19–30. <http://dx.doi.org/10.1016/j.virusres.2011.09.022>.
- Lambert, N.D., Ovsyannikova, I.G., Pankratz, V.S., Jacobson, R.M., Poland, G.A., 2012. Understanding the immune response to seasonal influenza vaccination in older adults: a systems biology approach. *Expert Rev. Vaccines* 11 (8), 985–994. <http://dx.doi.org/10.1586/erv.12.61>.
- Laurie, K.L., Carolan, L.A., Middleton, D., Lowther, S., Kelso, A., Barr, I.G., 2010. Multiple infections with seasonal influenza A virus induce cross-protective immunity against A(H1N1) pandemic influenza virus in a ferret model. *J. Infect. Dis.* 202 (7), 1011–1020. <http://dx.doi.org/10.1086/656188>.
- Longini Jr., I.M., Halloran, M.E., Nizam, A., Yang, Y., 2004. Containing pandemic influenza with antiviral agents. *Am. J. Epidemiol.* 159 (7), 623–633. <http://dx.doi.org/10.1093/aje/kwh092>.

- Longini Jr., I.M., Nizam, A., Xu, S., Ungchusak, K., Hanshaworakul, W., Cummings, D.A., Halloran, M.E., 2005. Containing pandemic influenza at the source. *Science* 309 (5737), 1083–1087. <http://dx.doi.org/10.1126/science.1115717>.
- Martcheva, M., Pilyugin, S.S., 2006. An epidemic model structured by host immunity. *J. Biol. Syst.* 14 (02), 185–203. <http://dx.doi.org/10.1142/s0218339006001787>.
- Medlock, J., Galvani, A.P., 2009. Optimizing influenza vaccine distribution. *Science* 325 (5948), 1705–1708. <http://dx.doi.org/10.1126/science.1175570>.
- Meyers, L.A., Newman, M., Pourbohloul, B., 2006. Predicting epidemics on directed contact networks. *J. Theoret. Biol.* 240 (3), 400–418. <http://dx.doi.org/10.1016/j.jtbi.2005.10.004>.
- Miller, E., Hoschler, K., Hardelid, P., Stanford, E., Andrews, N., Zambon, M., 2010. Incidence of 2009 pandemic influenza A H1N1 infection in England: a cross-sectional serological study. *Lancet* 375 (9720), 1100–1108. [http://dx.doi.org/10.1016/s0140-6736\(09\)62126-7](http://dx.doi.org/10.1016/s0140-6736(09)62126-7).
- Miller, J.C., Kiss, I.Z., 2014. Epidemic spread in networks: Existing methods and current challenges. *Math. Model. Nat. Phenom.* 9 (2), 4–42. <http://dx.doi.org/10.1051/mmnp/20149202>.
- Mills, C.E., Robins, J.M., Lipsitch, M., 2004. Transmissibility of 1918 pandemic influenza. *Nature* 432 (7019), 904–906. <http://dx.doi.org/10.1038/nature03063>.
- Mizumoto, K., Yamamoto, T., Nishiura, H., 2013. Age-dependent estimates of the epidemiological impact of pandemic influenza (H1N1-2009) in Japan. *Comput. Math. Methods Med.* 2013, 1–8. <http://dx.doi.org/10.1155/2013/637064>.
- Monto, A.S., 2004. Global burden of influenza: what we know and what we need to know. *Int. Congr. Ser.* 1263, 3–11. <http://dx.doi.org/10.1016/j.jics.2004.02.049>.
- Mozdzanowska, K., Maiese, K., Furchner, M., Gerhard, W., 1999. Treatment of influenza virus-infected SCID mice with nonneutralizing antibodies specific for the transmembrane proteins matrix 2 and neuraminidase reduces the pulmonary virus titer but fails to clear the infection. *Virology* 254 (1), 138–146. <http://dx.doi.org/10.1006/viro.1998.9534>.
- Murphy, B.R., Kasel, J.A., Chanock, R.M., 1972. Association of serum anti-neuraminidase antibody with resistance to influenza in man. *N. Engl. J. Med.* 286 (25), 1329–1332. <http://dx.doi.org/10.1056/nejm197206222862502>.
- National Institute of Infectious Diseases, 2020. Isolation/detection of influenza virus in Japan, 2017/18 season. <https://www.niid.go.jp/niid/en/influenza-e.html>. (Accessed 30 November 2020).
- Neumann, G., Kawaoka, Y., 2019. Predicting the next influenza pandemics. *J. Infect. Dis.* 219 (Suppl. 1), S14–S20. <http://dx.doi.org/10.1093/infdis/jiz040>.
- Nguyen, H.H., Zemlin, M., Ivanov, I.I., Andras, J., Zemlin, C., Vu, H.L., et al., 2007. Heterosubtypic immunity to influenza A virus infection requires a properly diversified antibody repertoire. *J. Virol.* 81 (17), 9331–9338. <http://dx.doi.org/10.1128/jvi.00751-07>.
- Nishiura, H., 2011. Real-time forecasting of an epidemic using a discrete time stochastic model: a case study of pandemic influenza (H1N1-2009). *Biomed. Eng. Online* 10 (1), 15. <http://dx.doi.org/10.1186/1475-925x-10-15>.
- Ohkusa, Y., Sugawara, T., Taniguchi, K., Okabe, N., 2011. Real-time estimation and prediction for pandemic A/H1N1 (2009) in Japan. *J. Infect. Chemother.* 17 (4), 468–472. <http://dx.doi.org/10.1007/s10156-010-0200-3>.
- Ohkusa, Y., Sugawara, T., et al., 2009. Simulation model of pandemic influenza in the whole of Japan. *Jpn. J. Infect. Dis.* 62 (2), 98–106.
- Omori, R., Nishiura, H., 2011. Theoretical basis to measure the impact of short-lasting control of an infectious disease on the epidemic peak. *Theor. Biol. Med. Model.* 8 (1), <http://dx.doi.org/10.1186/1742-4682-8-2>.
- Padilla-Quirarte, H.O., Lopez-Guerrero, D.V., Gutierrez-Xicotencatl, L., Esquivel-Guadarrama, F., 2019. Protective antibodies against influenza proteins. *Front. Immunol.* 10, <http://dx.doi.org/10.3389/fimmu.2019.01677>.
- Penman, B.S., Gupta, S., Shanks, G.D., 2016. Rapid mortality transition of Pacific Islands in the 19th century. *Epidemiol. Infect.* 145 (1), 1–11. <http://dx.doi.org/10.1017/s0950268816001989>.
- Petrova, V.N., Russell, C.A., 2017. The evolution of seasonal influenza viruses. *Nat. Rev. Microbiol.* 16 (1), 47–60. <http://dx.doi.org/10.1038/nrmicro.2017.118>.
- Pouillot, R., Lachenal, G., Pybus, O.G., Rousset, D., Njouom, R., 2008. Variable epidemic histories of hepatitis C virus genotype 2 infection in West Africa and Cameroon. *Infect. Genet. Evol.* 8 (5), 676–681. <http://dx.doi.org/10.1016/j.meegid.2008.06.001>.
- Pourbohloul, B., Ahued, A., Davoudi, B., Meza, R., Meyers, L.A., Skowronski, D.M., et al., 2009. Initial human transmission dynamics of the pandemic (H1N1) 2009 virus in North America. *Influenza Other Respir. Viruses* 3 (5), 215–222. <http://dx.doi.org/10.1111/j.1750-2659.2009.00100.x>.
- Rajaram, S., Wiecek, W., Lawson, R., Blak, B.T., Zhao, Y., Hackett, J., et al., 2017. Impact of increased influenza vaccination in 2–3-year-old children on disease burden within the general population: A Bayesian model-based approach. *PLoS One* 12 (12), e0186739. <http://dx.doi.org/10.1371/journal.pone.0186739>.
- Reichert, T., Chowell, G., McCullers, J.A., 2012. The age distribution of mortality due to influenza: pandemic and peri-pandemic. *BMC Med.* 10 (1), 162. <http://dx.doi.org/10.1186/1741-7015-10-162>.
- Reluga, T.C., Medlock, J., Perelson, A.S., 2008. Backward bifurcations and multiple equilibria in epidemic models with structured immunity. *J. Theoret. Biol.* 252 (1), 155–165. <http://dx.doi.org/10.1016/j.jtbi.2008.01.014>.
- Schulman, J.L., Khakpour, M., Kilbourne, E.D., 1968. Protective effects of specific immunity to viral neuraminidase on influenza virus infection of mice. *J. Virol.* 2 (8), 778–786. <http://dx.doi.org/10.1128/jvi.2.8.778-786.1968>.
- Skountzou, I., Koutsouanos, D.G., Kim, J.H., Powers, R., Satyabhaman, L., Masseoud, F., et al., 2010. Immunity to pre-1950 H1N1 influenza viruses confers cross-protection against the pandemic swine-origin 2009 A (H1N1) influenza virus. *J. Immunol.* 185 (3), 1642–1649. <http://dx.doi.org/10.4049/jimmunol.1000091>.
- Spiegelhalter, D.J., Best, N.G., Carlin, B.P., van der Linde, A., 2002. Bayesian measures of model complexity and fit. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 64 (4), 583–639. <http://dx.doi.org/10.1111/1467-9868.00353>.
- Sym, D., Patel, P.N., Ei-Chaar, G.M., 2009. Seasonal, avian, and novel H1N1 influenza: Prevention and treatment modalities. *Ann. Pharmacother.* 43 (12), 2001–2011. <http://dx.doi.org/10.1345/aph.1m557>.
- Taubenberger, J.K., Morens, D.M., 2006. 1918 influenza: the mother of all pandemics. *Emerg. Infect. Diseases* 12 (1), 15–22. <http://dx.doi.org/10.3201/eid1209.05-0979>.
- Thompson, R.N., 2020b. Epidemiological models are important tools for guiding COVID-19 interventions. *BMC Med.* 18 (1), <http://dx.doi.org/10.1186/s12916-020-01628-4>.
- Thompson, R.N., 2020c. Novel coronavirus outbreak in Wuhan, China, 2020: Intense surveillance is vital for preventing sustained transmission in new locations. *J. Clin. Med.* 9 (2), 498. <http://dx.doi.org/10.3390/jcm9020498>.
- Thompson, R.N., Brooks-Pollock, E., 2019. Detection, forecasting and control of infectious disease epidemics: modelling outbreaks in humans, animals and plants. *Philos. Trans. R. Soc. B* 374 (1775), 20190038. <http://dx.doi.org/10.1098/rstb.2019.0038>.
- Thompson, R.N., Gilligan, C.A., Cuniffe, N.J., 2016. Detecting presymptomatic infection is necessary to forecast major epidemics in the earliest stages of infectious disease outbreaks. *PLoS Comput. Biol.* 12 (4), e1004836. <http://dx.doi.org/10.1371/journal.pcbi.1004836>.
- Thompson, R.N., Gilligan, C.A., Cuniffe, N.J., 2020a. 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. <http://dx.doi.org/10.1098/rsif.2020.0690>.
- Thompson, R.N., Hart, W.S., 2018. Effect of confusing symptoms and infectiousness on forecasting and control of Ebola outbreaks. *Clin. Inf. Dis.* 67 (9), 1472–1474. <http://dx.doi.org/10.1093/cid/ciy248>.
- Thompson, R.N., Hollingsworth, T.D., Isham, V., Arribas-Bel, D., Ashby, B., Britton, T., et al., 2020d. Key questions for modelling COVID-19 exit strategies. *Proc. R. Soc. B* 287 (1932), 20201405. <http://dx.doi.org/10.1098/rspb.2020.1405>.
- Thompson, R.N., Thompson, C.P., Pelerman, O., Gupta, S., Obolski, U., 2019. Increased frequency of travel in the presence of cross-immunity may act to decrease the chance of a global pandemic. *Philos. Trans. R. Soc. B* 374 (1775), 20180274. <http://dx.doi.org/10.1098/rstb.2018.0274>.
- Tizzoni, M., Bajardi, P., Poletto, C., Ramasco, J.J., Balcan, D., Gonçalves, B., et al., 2012. Real-time numerical forecast of global epidemic spreading: case study of 2009 A/H1N1pdm. *BMC Med.* 10 (1), <http://dx.doi.org/10.1186/1741-7015-10-165>.
- Trifonov, V., Khabiabnian, H., Rabadan, R., 2009. Geographic dependence, surveillance, and origins of the 2009 influenza A (H1N1) virus. *N. Engl. J. Med.* 361 (2), 115–119. <http://dx.doi.org/10.1056/nejmp0904572>.
- Tuite, A.R., Greer, A.L., Whelan, M., Winter, A.-L., Lee, B., Yan, P., et al., 2009. Estimated epidemiologic parameters and morbidity associated with pandemic H1N1 influenza. *Can. Med. Assoc. J.* 182 (2), 131–136. <http://dx.doi.org/10.1503/cmaj.091807>.
- Volz, E., Meyers, L.A., 2007. Susceptible–infected–recovered epidemics in dynamic contact networks. *Proc. R. Soc. B* 274 (1628), 2925–2934. <http://dx.doi.org/10.1098/rspb.2007.1159>.
- Volz, E.M., Miller, J.C., Galvani, A., Meyers, L.A., 2011. Effects of heterogeneous and clustered contact patterns on infectious disease dynamics. *PLoS Comput. Biol.* 7 (6), e1002042. <http://dx.doi.org/10.1371/journal.pcbi.1002042>.
- Webster, R., Reay, P., Laver, W., 1988. Protection against lethal influenza with neuraminidase. *Virology* 164 (1), 230–237. [http://dx.doi.org/10.1016/0042-6822\(88\)90640-x](http://dx.doi.org/10.1016/0042-6822(88)90640-x).
- Weiss, H.H., 2013. The SIR model and the foundations of public health. *Mater. Mat.* (3), 1–17. <https://ddd.uab.cat/record/108432>.
- Xing, Z., Cardona, C.J., 2009. Preexisting immunity to pandemic (H1N1) 2009. *Emerg. Infect. Diseases* 15 (11), 1847–1849. <http://dx.doi.org/10.3201/eid1511.090685>.
- Xu, R., Ekiert, D.C., Krause, J.C., Hai, R., Crowe, J.E., Wilson, I.A., 2010. Structural basis of preexisting immunity to the 2009 H1N1 pandemic influenza virus. *Science* 328 (5976), 357–360. <http://dx.doi.org/10.1126/science.1186430>.
- Yang, W., Lipsitch, M., Shaman, J., 2015. Inference of seasonal and pandemic influenza transmission dynamics. *Proc. Natl. Acad. Sci. USA* 112 (9), 2723–2728. <http://dx.doi.org/10.1073/pnas.1415012112>.
- Yu, X., Tsibane, T., McGraw, P.A., House, F.S., Keefer, C.J., Hicar, M.D., et al., 2008. Neutralizing antibodies derived from the B cells of 1918 influenza pandemic survivors. *Nature* 455 (7212), 532–536. <http://dx.doi.org/10.1038/nature07231>.

Bibliography

- [1] Abrams, S., Wambua, J., Santermans, E., Willem, L., Kuylen, E., Coletti, P., et al. Modelling the early phase of the belgian COVID-19 epidemic using a stochastic compartmental model and studying its implied future trajectories. *Epidemics*, 35:100449, 2021.
- [2] ACAPS. Secondary impacts of COVID-19. <https://www.acaps.org/projects/secondary-impacts-covid19/data>, 2021. Accessed on 08/06/21.
- [3] Adam, D. C., Wu, P., Wong, J. Y., Lau, E. H. Y., Tsang, T. K., Cauchemez, S., et al. Clustering and superspreading potential of SARS-CoV-2 infections in Hong Kong. *Nature Medicine*, 26(11):1714–1719, 2020.
- [4] Ajelli, M., Zhang, Q., Sun, K., Merler, S., Fumanelli, L., Chowell, G., et al. The RAPIDD Ebola forecasting challenge: Model description and synthetic data generation. *Epidemics*, 22:3–12, 2018.
- [5] Akhmetzhanov, A. R., Jung, S.-m., Cheng, H.-Y., and Thompson, R. N. A hospital-related outbreak of SARS-CoV-2 associated with variant Epsilon (B.1.429) in Taiwan: transmission potential and outbreak containment under intensified contact tracing, January–February 2021. *International Journal of Infectious Diseases*, 110:15–20, 2021.
- [6] Aldrich, J. R.A. Fisher and the making of maximum likelihood 1912-1922. *Statistical Science*, 12(3), 1997.
- [7] Aledort, J. E., Lurie, N., Wasserman, J., and Bozzette, S. A. Non-pharmaceutical public health interventions for pandemic influenza: an evaluation of the evidence base. *BMC Public Health*, 7(1), 2007.
- [8] Allen, L. J. S. An introduction to stochastic epidemic models. In *Mathematical Epidemiology*, pages 81–130. Springer Berlin Heidelberg, 2008.
- [9] Allen, L. J. S. *Stochastic Population and Epidemic Models*. Springer International Publishing, 2015.
- [10] Allen, L. J. S. A primer on stochastic epidemic models: Formulation, numerical simulation, and analysis. *Infectious Disease Modelling*, 2(2):128–142, 2017.

- [11] Althaus, C., Low, N., Musa, E., Shuaib, F., and Gsteiger, S. Ebola virus disease outbreak in Nigeria: Transmission dynamics and rapid control. *Epidemics*, 11:80–84, 2015.
- [12] Anderson, R. M. and May, R. M. *Infectious Diseases of Humans: Dynamics and Control*. Oxford University Press, 1991.
- [13] Anderson, R. M., Fraser, C., Ghani, A. C., Donnelly, C. A., Riley, S., Ferguson, N. M., et al. Epidemiology, transmission dynamics and control of SARS: the 2002–2003 epidemic. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 359(1447):1091–1105, 2004.
- [14] Anderson, R. M., Heesterbeek, H., Klinkenberg, D., and Hollingsworth, T. D. How will country-based mitigation measures influence the course of the COVID-19 epidemic? *The Lancet*, 395(10228):931–934, 2020.
- [15] Andreasen, V., Lin, J., and Levin, S. A. The dynamics of cocirculating influenza strains conferring partial cross-immunity. *Journal of Mathematical Biology*, 35(7):825–842, 1997.
- [16] Arenas, A., Cota, W., Gómez-Gardeñes, J., Gómez, S., Granell, C., Matamalas, J. T., et al. Modeling the spatiotemporal epidemic spreading of COVID-19 and the impact of mobility and social distancing interventions. *Physical Review X*, 10(4), 2020.
- [17] Bacaër, N. and Dads, E. H. A. On the probability of extinction in a periodic environment. *Journal of Mathematical Biology*, 68(3):533–548, 2012.
- [18] Bacaër, N. and Ed-Darraz, A. On linear birth-and-death processes in a random environment. *Journal of Mathematical Biology*, 69(1):73–90, 2013.
- [19] Baccam, P., Beauchemin, C., Macken, C. A., Hayden, F. G., and Perelson, A. S. Kinetics of influenza A virus infection in humans. *Journal of Virology*, 80(15):7590–7599, 2006.
- [20] Backer, J., Wallinga, J., Meijer, A., Donker, G., Van der Hoek, W., and Van Boven, M. The impact of influenza vaccination on infection, hospitalisation and mortality in the Netherlands between 2003 and 2015. *Epidemics*, 26:77–85, 2019.
- [21] Baden, L. R., Sahly, H. M. E., Essink, B., Kotloff, K., Frey, S., Novak, R., et al. Efficacy and safety of the mRNA-1273 SARS-CoV-2 vaccine. *New England Journal of Medicine*, 384(5):403–416, 2021.
- [22] Baguelin, M., Flasche, S., Camacho, A., Demiris, N., Miller, E., and Edmunds, W. J. Assessing optimal target populations for influenza vaccination programmes: an evidence synthesis and modelling study. *PLoS Medicine*, 10(10):e1001527, 2013.

- [23] Bai, F., Huff, K. E. S., and Allen, L. J. S. The effect of delay in viral production in within-host models during early infection. *Journal of Biological Dynamics*, 13(sup1): 47–73, 2018.
- [24] Bajardi, P., Poletto, C., Ramasco, J. J., Tizzoni, M., Colizza, V., and Vespignani, A. Human mobility networks, travel restrictions, and the global spread of 2009 H1N1 pandemic. *PLoS One*, 6(1):e16591, 2011.
- [25] Baker, M. G., Wilson, N., and Anglemeyer, A. Successful elimination of Covid-19 transmission in New Zealand. *New England Journal of Medicine*, 383(8):e56, 2020.
- [26] Baker, M. G., Wilson, N., and Blakely, T. Elimination could be the optimal response strategy for covid-19 and other emerging pandemic diseases. *BMJ*, page m4907, 2020.
- [27] Ball, F. The threshold behaviour of epidemic models. *Journal of Applied Probability*, 20(2):227–241, 1983.
- [28] Bandaranayake, D., Huang, Q. S., Bissielo, A., Wood, T., Mackereth, G., Baker, M. G., et al. Risk factors and immunity in a nationally representative population following the 2009 influenza A(H1N1) pandemic. *PLoS One*, 5(10):e13211, 2010.
- [29] Banerjee, S., Guedj, J., Ribeiro, R. M., Moses, M., and Perelson, A. S. Estimating biologically relevant parameters under uncertainty for experimental within-host murine West Nile virus infection. *Journal of The Royal Society Interface*, 13(117):20160130, 2016.
- [30] Basta, N. E., Chao, D. L., Halloran, M. E., Matrajt, L., and Longini, I. M. Strategies for pandemic and seasonal influenza vaccination of schoolchildren in the United States. *American Journal of Epidemiology*, 170(6):679–686, 2009.
- [31] Beauchemin, C. A. and Handel, A. A review of mathematical models of influenza A infections within a host or cell culture: lessons learned and challenges ahead. *BMC Public Health*, 11(Suppl 1):S7, 2011.
- [32] Beauchemin, C. A., McSharry, J. J., Drusano, G. L., Nguyen, J. T., Went, G. T., Ribeiro, R. M., and Perelson, A. S. Modeling amantadine treatment of influenza A virus in vitro. *Journal of Theoretical Biology*, 254(2):439–451, 2008.
- [33] Bedford, J., Farrar, J., Ihekweazu, C., Kang, G., Koopmans, M., and Nkengasong, J. A new twenty-first century science for effective epidemic response. *Nature*, 575(7781): 130–136, 2019.
- [34] Bellman, R. and Åström, K. On structural identifiability. *Mathematical Biosciences*, 7 (3-4):329–339, 1970.

- [35] Bender, B. S., Croghan, T., Zhang, L., and Small, P. A. Transgenic mice lacking class I major histocompatibility complex-restricted T cells have delayed viral clearance and increased mortality after influenza virus challenge. *Journal of Experimental Medicine*, 175(4):1143–1145, 1992.
- [36] Best, K. and Perelson, A. S. Mathematical modeling of within-host Zika virus dynamics. *Immunological Reviews*, 285(1):81–96, 2018.
- [37] Bettencourt, L. M. and Ribeiro, R. M. Real time bayesian estimation of the epidemic potential of emerging infectious diseases. *PLoS One*, 3(5):e2185, 2008.
- [38] Betti, M. I. and Heffernan, J. M. A simple model for fitting mild, severe, and known cases during an epidemic with an application to the current SARS-CoV-2 pandemic. *Infectious Disease Modelling*, 6:313–323, 2021.
- [39] Biggerstaff, M., Cauchemez, S., Reed, C., Gambhir, M., and Finelli, L. Estimates of the reproduction number for seasonal, pandemic, and zoonotic influenza: a systematic review of the literature. *BMC Infectious Diseases*, 14(1), 2014.
- [40] Biggerstaff, M., Alper, D., Dredze, M., Fox, S., Fung, I. C.-H., Hickmann, K. S., et al. Results from the centers for disease control and prevention’s predict the 2013–2014 influenza season challenge. *BMC Infectious Diseases*, 16(1), 2016.
- [41] Biggerstaff, M., Johansson, M., Alper, D., Brooks, L. C., Chakraborty, P., Farrow, D. C., et al. Results from the second year of a collaborative effort to forecast influenza seasons in the United States. *Epidemics*, 24:26–33, 2018.
- [42] Billah, M. A., Miah, M. M., and Khan, M. N. Reproductive number of coronavirus: A systematic review and meta-analysis based on global level evidence. *PLoS One*, 15(11): e0242128, 2020.
- [43] Blomberg, B. B. and Frasca, D. Quantity, not quality, of antibody response decreased in the elderly. *Journal of Clinical Investigation*, 121(8):2981–2983, 2011.
- [44] Bocharov, G. and Romanyukha, A. Mathematical model of antiviral immune response III. influenza A virus infection. *Journal of Theoretical Biology*, 167(4):323–360, 1994.
- [45] Boianelli, A., Nguyen, V. K., Ebsensen, T., Schulze, K., Wilk, E., Sharma, N., et al. Modeling influenza virus infection: A roadmap for influenza research. *Viruses*, 7(10): 5274–5304, 2015.
- [46] Bouvier, N. M. and Palese, P. The biology of influenza viruses. *Vaccine*, 26:D49–D53, 2008.

- [47] Britton, T. Stochastic epidemic models: A survey. *Mathematical Biosciences*, 225(1): 24–35, 2010.
- [48] Britton, T., Ball, F., and Trapman, P. A mathematical model reveals the influence of population heterogeneity on herd immunity to SARS-CoV-2. *Science*, 369(6505): 846–849, 2020.
- [49] Bromiley, J., Hart, W., Iwami, S., and Thompson, R. Assessing the threat from emerging infectious diseases in seasonally varying environments. *Preprint*, 2021.
- [50] Brookmeyer, R. Measuring the HIV/AIDS epidemic: Approaches and challenges. *Epidemiologic Reviews*, 32(1):26–37, 2010.
- [51] Bubar, K. M., Reinholt, K., Kissler, S. M., Lipsitch, M., Cobey, S., Grad, Y. H., and Larremore, D. B. Model-informed COVID-19 vaccine prioritization strategies by age and serostatus. *Science*, 371(6532):916–921, 2021.
- [52] Buckeridge, D. L., Burkom, H., Campbell, M., Hogan, W. R., and Moore, A. W. Algorithms for rapid outbreak detection: a research synthesis. *Journal of Biomedical Informatics*, 38(2):99–113, 2005.
- [53] Burnham, K. P. and Anderson, D. R. *Model Selection and Multimodel Inference*. Springer New York, 2004.
- [54] Callaway, E. The coronavirus is mutating — does it matter? *Nature*, 585(7824):174–177, 2020.
- [55] Canini, L. and Carrat, F. Population modeling of influenza A/H1N1 virus kinetics and symptom dynamics. *Journal of Virology*, 85(6):2764–2770, 2010.
- [56] Canini, L., Conway, J. M., Perelson, A. S., and Carrat, F. Impact of different oseltamivir regimens on treating influenza A virus infection and resistance emergence: Insights from a modelling study. *PLoS Computational Biology*, 10(4):e1003568, 2014.
- [57] Canini, L., Woolhouse, M. E. J., Maines, T. R., and Carrat, F. Heterogeneous shedding of influenza by human subjects and its implications for epidemiology and control. *Scientific Reports*, 6(1), 2016.
- [58] Cao, P. and McCaw, J. The mechanisms for within-host influenza virus control affect model-based assessment and prediction of antiviral treatment. *Viruses*, 9(8):197, 2017.
- [59] Carmona, P. and Gandon, S. Winter is coming: Pathogen emergence in seasonal environments. *PLoS Computational Biology*, 16(7):e1007954, 2020.

- [60] Carrat, F., Vergu, E., Ferguson, N. M., Lemaître, M., Cauchemez, S., Leach, S., and Valleron, A.-J. Time lines of infection and disease in human influenza: A review of volunteer challenge studies. *American Journal of Epidemiology*, 167(7):775–785, 2008.
- [61] Centers for Disease Control and Prevention (CDC). How CDC Uses Flu Forecasting. <https://www.cdc.gov/flu/weekly/flusight/how-flu-forecasting.htm>, 2020. Accessed on 27/10/20.
- [62] Centers for Disease Control and Prevention (CDC). Interim public health recommendations for fully vaccinated people. <https://www.cdc.gov/coronavirus/2019-ncov/vaccines/fully-vaccinated-guidance.html>, 2021. Accessed on 28/05/21.
- [63] Challen, R., Brooks-Pollock, E., Read, J. M., Dyson, L., Tsaneva-Atanasova, K., and Danon, L. Risk of mortality in patients infected with SARS-CoV-2 variant of concern 202012/1: matched cohort study. *BMJ*, page n579, 2021.
- [64] Chappell, M. J., Godfrey, K. R., and Vajda, S. Global identifiability of the parameters of nonlinear systems with specified inputs: A comparison of methods. *Mathematical Biosciences*, 102(1):41–73, 1990.
- [65] Chen, M. I. C. 2009 influenza A(H1N1) seroconversion rates and risk factors among distinct adult cohorts in Singapore. *JAMA*, 303(14):1383, 2010.
- [66] Chen, S.-C. and Liao, C.-M. Modelling control measures to reduce the impact of pandemic influenza among schoolchildren. *Epidemiology and Infection*, 136(8):1035–1045, 2008.
- [67] Cheng, Y.-H., You, S.-H., Lin, Y.-J., Chen, S.-C., Chen, W.-Y., Chou, W.-C., et al. Mathematical modeling of postcoinfection with influenza A virus and *Streptococcus pneumoniae*, with implications for pneumonia and COPD-risk assessment. *International Journal of Chronic Obstructive Pulmonary Disease*, Volume 12:1973–1988, 2017.
- [68] Childs, L. M., Moustaid, F. E., Gajewski, Z., Kadelka, S., Nikin-Beers, R., John W. Smith, J., et al. Linked within-host and between-host models and data for infectious diseases: a systematic review. *PeerJ*, 7:e7057, 2019.
- [69] Chinazzi, M., Davis, J. T., Ajelli, M., Gioannini, C., Litvinova, M., Merler, S., et al. The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. *Science*, 368(6489):395–400, 2020.
- [70] Chis, O., Banga, J. R., and Balsa-Canto, E. GenSSI: a software toolbox for structural identifiability analysis of biological models. *Bioinformatics*, 2011.

- [71] Chis, O.-T., Banga, J. R., and Balsa-Canto, E. Structural identifiability of systems biology models: A critical comparison of methods. *PLoS One*, 6(11):e27755, 2011.
- [72] Chowell, G., Nishiura, H., and Bettencourt, L. M. Comparative estimation of the reproduction number for pandemic influenza from daily case notification data. *Journal of The Royal Society Interface*, 4(12):155–166, 2006.
- [73] Chowell, G., Viboud, C., Wang, X., Bertozzi, S. M., and Miller, M. A. Adaptive vaccination strategies to mitigate pandemic influenza: Mexico as a case study. *PLoS One*, 4(12):e8164, 2009.
- [74] Chowell, G., Sattenspiel, L., Bansal, S., and Viboud, C. Mathematical models to characterize early epidemic growth: A review. *Physics of Life Reviews*, 18:66–97, 2016.
- [75] Christman, M. C., Kedwaih, A., Xu, J., Donis, R. O., and Lu, G. Pandemic (H1N1) 2009 virus revisited: an evolutionary retrospective. *Infection, Genetics and Evolution*, 11(5):803–811, 2011.
- [76] Cintrón-Arias, A., Castillo-Chávez, C., Bettencourt, L. M., Lloyd, A. L., and Banks, H. The estimation of the effective reproductive number from disease outbreak data. *Mathematical Biosciences and Engineering*, 6(2):261–282, 2009.
- [77] Ciupe, S. M. and Heffernan, J. M. In-host modeling. *Infectious Disease Modelling*, 2(2):188–202, 2017.
- [78] Cohen, S., Doyle, W. J., and Skoner, D. P. Psychological stress, cytokine production, and severity of upper respiratory illness. *Psychosomatic medicine*, 61(2):175–180, 1999.
- [79] Couch, R. B., Kasel, J. A., Gerin, J. L., Schulman, J. L., and Kilbourne, E. D. Induction of partial immunity to influenza by a neuraminidase-specific vaccine. *Journal of Infectious Diseases*, 129(4):411–420, 1974.
- [80] Cowling, B. J., Ali, S. T., Ng, T. W. Y., Tsang, T. K., Li, J. C. M., Fong, M. W., et al. Impact assessment of non-pharmaceutical interventions against coronavirus disease 2019 and influenza in Hong Kong: an observational study. *The Lancet Public Health*, 5(5): e279–e288, 2020.
- [81] Dagan, N., Barda, N., Kepten, E., Miron, O., Perchik, S., Katz, M. A., and others, M. A. H. BNT162b2 mRNA Covid-19 vaccine in a nationwide mass vaccination setting. *New England Journal of Medicine*, 2021.
- [82] Daniels, B. C. and Nemenman, I. Automated adaptive inference of phenomenological dynamical models. *Nature Communications*, 6(1), 2015.

- [83] Daon, Y., Thompson, R. N., and Obolski, U. Estimating COVID-19 outbreak risk through air travel. *Journal of Travel Medicine*, 27(5), 2020.
- [84] Davies, N. G., Klepac, P., Liu, Y., Prem, K., Jit, M., Pearson, C. A. B., et al. Age-dependent effects in the transmission and control of COVID-19 epidemics. *Nature Medicine*, 26(8):1205–1211, 2020.
- [85] Davies, N. G., Kucharski, A. J., Eggo, R. M., Gimma, A., Edmunds, W. J., Jombart, T., et al. Effects of non-pharmaceutical interventions on COVID-19 cases, deaths, and demand for hospital services in the UK: a modelling study. *The Lancet Public Health*, 5(7):e375–e385, 2020.
- [86] Davies, N. G., Abbott, S., Barnard, R. C., Jarvis, C. I., Kucharski, A. J., Munday, J. D., et al. Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England. *Science*, 372(6538):eabg3055, 2021.
- [87] de Jong, M. D., Simmons, C. P., Thanh, T. T., Hien, V. M., Smith, G. J. D., Chau, T. N. B., et al. Fatal outcome of human influenza A (H5N1) is associated with high viral load and hypercytokinemia. *Nature Medicine*, 12(10):1203–1207, 2006.
- [88] Department of Health and Social Care. Priority groups for coronavirus (COVID-19) vaccination: advice from the JCVI. <https://www.gov.uk/government/publications/priority-groups-for-coronavirus-covid-19-vaccination-advice-from-the-jcvi-30-december-2020>, 2020. Accessed on 03/04/21.
- [89] Diekmann, O., Heesterbeek, J., and Metz, J. On the definition and the computation of the basic reproduction ratio R_0 in models for infectious diseases in heterogeneous populations. *Journal of Mathematical Biology*, 28(4), 1990.
- [90] Dietzen, D. J. Amino acids, peptides, and proteins. In *Principles and Applications of Molecular Diagnostics*, pages 345–380. Elsevier, 2018.
- [91] Dobrovolny, H. M., Baron, M. J., Gieschke, R., Davies, B. E., Jumbe, N. L., and Beauchemin, C. A. A. Exploring cell tropism as a possible contributor to influenza infection severity. *PLoS One*, 5(11):e13811, 2010.
- [92] Dobrovolny, H. M., Gieschke, R., Davies, B. E., Jumbe, N. L., and Beauchemin, C. A. Neuraminidase inhibitors for treatment of human and avian strain influenza: A comparative modeling study. *Journal of Theoretical Biology*, 269(1):234–244, 2011.
- [93] Dobrovolny, H. M., Reddy, M. B., Kamal, M. A., Rayner, C. R., and Beauchemin, C. A. A. Assessing mathematical models of influenza infections using features of the immune response. *PLoS One*, 8(2):e57088, 2013.

- [94] Donaldson, L. J., Rutter, P. D., Ellis, B. M., Greaves, F. E. C., Mytton, O. T., Pebody, R. G., and Yardley, I. E. Mortality from pandemic A/H1N1 2009 influenza in England: public health surveillance study. *BMJ*, 339(dec10 1):b5213–b5213, 2009.
- [95] Dou, D., Revol, R., Östbye, H., Wang, H., and Daniels, R. Influenza A virus cell entry, replication, virion assembly and movement. *Frontiers in Immunology*, 9, 2018.
- [96] Du, S. Q. and Yuan, W. Mathematical modeling of interaction between innate and adaptive immune responses in COVID-19 and implications for viral pathogenesis. *Journal of Medical Virology*, 92(9):1615–1628, 2020.
- [97] Du, X., King, A. A., Woods, R. J., and Pascual, M. Evolution-informed forecasting of seasonal influenza A (H3N2). *Science Translational Medicine*, 9(413):eaan5325, 2017.
- [98] Dushoff, J., Plotkin, J. B., Viboud, C., Simonsen, L., Miller, M., Loeb, M., and David, J. Vaccinating to protect a vulnerable subpopulation. *PLoS Medicine*, 4(5):e174, 2007.
- [99] Earn, D. J. Effects of school closure on incidence of pandemic influenza in Alberta, Canada. *Annals of Internal Medicine*, 156(3):173, 2012.
- [100] Eccles, R. Understanding the symptoms of the common cold and influenza. *The Lancet infectious diseases*, 5(11):718–725, 2005.
- [101] Eichelberger, M., Allan, W., Zijlstra, M., Jaenisch, R., and Doherty, P. C. Clearance of influenza virus respiratory infection in mice lacking class i major histocompatibility complex-restricted CD8+ T cells. *Journal of Experimental Medicine*, 174(4):875–880, 1991.
- [102] Ellebedy, A., Ducatez, M., Duan, S., Stigger-Rosser, E., Rubrum, A., Govorkova, E., et al. Impact of prior seasonal influenza vaccination and infection on pandemic A (H1N1) influenza virus replication in ferrets. *Vaccine*, 29(17):3335–3339, 2011.
- [103] Ellebedy, A. H., Fabrizio, T. P., Kayali, G., Oguin, T. H., Brown, S. A., Rehg, J., et al. Contemporary seasonal influenza A (H1N1) virus infection primes for a more robust response to split inactivated pandemic influenza A (H1N1) virus vaccination in ferrets. *Clinical and Vaccine Immunology*, 17(12):1998–2006, 2010.
- [104] Emary, K. R. W., Golubchik, T., Aley, P. K., Ariani, C. V., Angus, B. J., Bibi, S., et al. Efficacy of ChAdOx1 nCoV-19 (AZD1222) vaccine against SARS-CoV-2 VOC 202012/01 (B.1.1.7). *SSRN Electronic Journal*, 2021.
- [105] Endo, A., Abbott, S., Kucharski, A. J., and and, S. F. Estimating the overdispersion in COVID-19 transmission using outbreak sizes outside China. *Wellcome Open Research*, 5:67, 2020.

- [106] Epstein, S., Misplon, J., Lawson, C., Subbarao, E., Connors, M., and Murphy, B. Beta 2-microglobulin-deficient mice can be protected against influenza A infection by vaccination with vaccinia-influenza recombinants expressing hemagglutinin and neuraminidase. *The Journal of Immunology*, 150(12):5484–5493, 1993.
- [107] Fadai, N. T., Sachak-Patwa, R., Byrne, H. M., Maini, P. K., Bafadhel, M., and Nicolau, D. V. Infection, inflammation and intervention: mechanistic modelling of epithelial cells in COVID-19. *Journal of The Royal Society Interface*, 18(175), 2021.
- [108] Fan, G., Tu, C., Zhou, F., Liu, Z., Wang, Y., Song, B., et al. Comparison of severity scores for COVID-19 patients with pneumonia: a retrospective study. *European Respiratory Journal*, 56(3):2002113, 2020.
- [109] Fang, Y., Banner, D., Kelvin, A. A., Huang, S. S. H., Paige, C. J., Corfe, S. A., et al. Seasonal H1N1 influenza virus infection induces cross-protective pandemic H1N1 virus immunity through a CD8-independent, B cell-dependent mechanism. *Journal of Virology*, 86(4):2229–2238, 2011.
- [110] Ferguson, N., Laydon, D., Nedjati Gilani, G., Imai, N., Ainslie, K., Baguelin, M., et al. Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand. 2020. Accessed on 01/05/21.
- [111] Fesq, H., Bacher, M., Nain, M., and Gemsa, D. Programmed cell death (apoptosis) in human monocytes infected by influenza A virus. *Immunobiology*, 190(1-2):175–182, 1994.
- [112] Fielding, J. E., Kelly, H. A., Mercer, G. N., and Glass, K. Systematic review of influenza A(H1N1)pdm09 virus shedding: duration is affected by severity, but not age. *Influenza and Other Respiratory Viruses*, 8(2):142–150, 2013.
- [113] Fisher, R. A. On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, 222(594-604):309–368, 1922.
- [114] Fisman, D., Rivers, C., Lofgren, E., and Majumder, M. S. Estimation of MERS-coronavirus reproductive number and case fatality rate for the spring 2014 Saudi Arabia outbreak: Insights from publicly available data. *PLoS Currents*, 6, 2014.
- [115] Flaxman, S., Mishra, S., Gandy, A., Unwin, H. J. T., Mellan, T. A., Coupland, H., et al. Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. *Nature*, 584(7820):257–261, 2020.

- [116] Fontanet, A., Autran, B., Lina, B., Kieny, M. P., Karim, S. S. A., and Sridhar, D. SARS-CoV-2 variants and ending the COVID-19 pandemic. *The Lancet*, 397(10278): 952–954, 2021.
- [117] Fox, S. J., Miller, J. C., and Meyers, L. A. Seasonality in risk of pandemic influenza emergence. *PLoS Computational Biology*, 13(10):e1005749, 2017.
- [118] Fraser, C. Estimating individual and household reproduction numbers in an emerging epidemic. *PLoS One*, 2(8):e758, 2007.
- [119] Fraser, C., Donnelly, C. A., Cauchemez, S., Hanage, W. P., Van Kerkhove, M. D., Hollingsworth, T. D., et al. Pandemic potential of a strain of influenza A (H1N1): early findings. *Science*, 324(5934):1557–1561, 2009.
- [120] Fritz, R. S., Hayden, F. G., Calfee, D. P., Cass, L. M. R., Peng, A. W., Alvord, W. G., et al. Nasal cytokine and chemokine responses in experimental influenza A virus infection: Results of a placebo-controlled trial of intravenous zanamivir treatment. *The Journal of Infectious Diseases*, 180(3):586–593, 1999.
- [121] Frost, S. D., Pybus, O. G., Gog, J. R., Viboud, C., Bonhoeffer, S., and Bedford, T. Eight challenges in phylodynamic inference. *Epidemics*, 10:88–92, 2015.
- [122] Fukuyama, S. and Kawaoka, Y. The pathogenesis of influenza virus infections: the contributions of virus and host factors. *Current Opinion in Immunology*, 23(4):481–486, 2011.
- [123] Fulton, R. B. and Varga, S. M. Effects of aging on the adaptive immune response to respiratory virus infections. *Aging Health*, 5(6):775–787, 2009.
- [124] Funk, S., Camacho, A., Kucharski, A. J., Lowe, R., Eggo, R. M., and Edmunds, W. J. Assessing the performance of real-time epidemic forecasts: A case study of Ebola in the Western Area region of Sierra Leone, 2014-15. *PLoS Computational Biology*, 15(2): e1006785, 2019.
- [125] Gani, R., Hughes, H., Fleming, D., Griffin, T., Medlock, J., and Leach, S. Potential impact of antiviral drug use during influenza pandemic. *Emerging Infectious Diseases*, 11(9):1355, 2005.
- [126] García-Sastre, A. Inhibition of interferon-mediated antiviral responses by influenza A viruses and other negative-strand RNA viruses. *Virology*, 279(2):375–384, 2001.
- [127] Gart, J. J. The mathematical analysis of an epidemic with two kinds of susceptibles. *Biometrics*, 24(3):557, 1968.

- [128] Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. *Bayesian data analysis*. CRC Press, 2013.
- [129] Gibbons, C. L., Mangen, M.-J. J., Plass, D., Havelaar, A. H., Brooke, R. J., Kramarz, P., et al. Measuring underreporting and under-ascertainment in infectious disease datasets: a comparison of methods. *BMC Public Health*, 14(1):147, 2014.
- [130] Gilks, W., Richardson, S., and Spiegelhalter, D. *Markov Chain Monte Carlo in Practice*. Chapman and Hall/CRC, 1995.
- [131] Gillespie, D. T. Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry*, 81(25):2340–2361, 1977.
- [132] Ginaldi, L., Loreto, M. F., Corsi, M. P., Modesti, M., and Martinis, M. D. Immunosenescence and infectious diseases. *Microbes and Infection*, 3(10):851–857, 2001.
- [133] Glasser, J., Taneri, D., Feng, Z., Chuang, J.-H., Tüll, P., Thompson, W., et al. Evaluation of targeted influenza vaccination strategies via population modeling. *PLoS One*, 5(9):e12777, 2010.
- [134] Gog, J. R. and Hollingsworth, T. D. Epidemic interventions: insights from classic results. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 376(1829):20200263, 2021.
- [135] Gog, J. R., Pellis, L., Wood, J. L., McLean, A. R., Arinaminpathy, N., and Lloyd-Smith, J. O. Seven challenges in modeling pathogen dynamics within-host and across scales. *Epidemics*, 10:45–48, 2015.
- [136] Gog, J. R., Hill, E. M., Danon, L., and Thompson, R. Vaccine escape in a heterogeneous population: insights for SARS-CoV-2 from a simple model. *medRxiv*, 2021.
- [137] Goncalves, A., Ray, P., Soper, B., Stevens, J., Coyle, L., and Sales, A. P. Generation and evaluation of synthetic patient data. *BMC Medical Research Methodology*, 20(1), 2020.
- [138] González-Parra, G. and Dobrovolny, H. M. Assessing uncertainty in A2 respiratory syncytial virus viral dynamics. *Computational and Mathematical Methods in Medicine*, 2015:1–9, 2015.
- [139] Gostic, K. M., Ambrose, M., Worobey, M., and Lloyd-Smith, J. O. Potent protection against H5N1 and H7N9 influenza via childhood hemagglutinin imprinting. *Science*, 354(6313):722–726, 2016.

- [140] Gostic, K. M., Bridge, R., Brady, S., Viboud, C., Worobey, M., and Lloyd-Smith, J. O. Childhood immune imprinting to influenza A shapes birth year-specific risk during seasonal H1N1 and H3N2 epidemics. *PLoS Pathogens*, 15(12), 2019.
- [141] Grant, E. J., Chen, L., Quinones-Parra, S., Pang, K., Kedzierska, K., and Chen, W. T-Cell immunity to influenza A viruses. *Critical Reviews in Immunology*, 34(1):15–39, 2014.
- [142] Grebe, K. M., Yewdell, J. W., and Bennink, J. R. Heterosubtypic immunity to influenza A virus: where do we stand? *Microbes and Infection*, 10(9):1024–1029, 2008.
- [143] Grenfell, B. T. Unifying the epidemiological and evolutionary dynamics of pathogens. *Science*, 303(5656):327–332, 2004.
- [144] Gu, Y. Pandemic (H1N1) 2009 transmission during presymptomatic phase, Japan. *Emerging Infectious Diseases*, 17(9):1737–1739, 2011.
- [145] Gubareva, L. V., Kaiser, L., and Hayden, F. G. Influenza virus neuraminidase inhibitors. *The Lancet*, 355(9206):827–835, 2000.
- [146] Gumel, A. B., Ruan, S., Day, T., Watmough, J., Brauer, F., van den Driessche, P., et al. Modelling strategies for controlling SARS outbreaks. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 271(1554):2223–2232, 2004.
- [147] Hadjichrysanthou, C., Cauët, E., Lawrence, E., Vegvari, C., de Wolf, F., and Anderson, R. M. Understanding the within-host dynamics of influenza A virus: from theory to clinical implications. *Journal of The Royal Society Interface*, 13(119):20160289, 2016.
- [148] Hagau, N., Slavcovici, A., Gonganau, D. N., Oltean, S., Dirzu, D. S., Brezoszki, E. S., et al. Clinical aspects and cytokine response in severe H1N1 influenza A virus infection. *Critical Care*, 14(6):R203, 2010.
- [149] Hall, I. M., Gani, R., Hughes, H. E., and Leach, S. Real-time epidemic forecasting for pandemic influenza. *Epidemiology and Infection*, 135(3):372–385, 2007.
- [150] Hancioglu, B., Swigon, D., and Clermont, G. A dynamical model of human immune response to influenza A virus infection. *Journal of Theoretical Biology*, 246(1):70–86, 2007.
- [151] Hancock, K., Veguilla, V., Lu, X., Zhong, W., Butler, E. N., Sun, H., et al. Cross-reactive antibody responses to the 2009 pandemic H1N1 influenza virus. *New England Journal of Medicine*, 361(20):1945–1952, 2009.

- [152] Handel, A., Longini, I. M., and Antia, R. Neuraminidase inhibitor resistance in influenza: Assessing the danger of its generation and spread. *PLoS Computational Biology*, 3(12):e240, 2007.
- [153] Handel, A., Longini, I. M., and Antia, R. Towards a quantitative understanding of the within-host dynamics of influenza A infections. *Journal of The Royal Society Interface*, 7(42):35–47, 2009.
- [154] Handel, A., Liao, L. E., and Beauchemin, C. A. Progress and trends in mathematical modelling of influenza A virus infections. *Current Opinion in Systems Biology*, 12:30–36, 2018.
- [155] Haq, K. and McElhaney, J. E. Immunosenescence: influenza vaccination and the elderly. *Current Opinion in Immunology*, 29:38–42, 2014.
- [156] Hardelid, P., Andrews, N., Hoschler, K., Stanford, E., Baguelin, M., Waight, P., et al. Assessment of baseline age-specific antibody prevalence and incidence of infection to novel influenza A/H1N1 2009. *Health Technology Assessment*, 14(55):115–92, 2010.
- [157] Harris, R. J., Hall, J. A., Zaidi, A., Andrews, N. J., Dunbar, J. K., and Dabrera, G. Impact of vaccination on household transmission of SARS-COV-2 in England. *medRxiv*, 2021.
- [158] Hart, W. S., Maini, P. K., Yates, C. A., and Thompson, R. N. A theoretical framework for transitioning from patient-level to population-scale epidemiological dynamics: influenza A as a case study. *Journal of The Royal Society Interface*, 17(166):20200230, 2020.
- [159] Hashimoto, Y., Moki, T., Takizawa, T., Shiratsuchi, A., and Nakanishi, Y. Evidence for phagocytosis of influenza virus-infected, apoptotic cells by neutrophils and macrophages in mice. *The Journal of Immunology*, 178(4):2448–2457, 2007.
- [160] Hastings, W. K. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1):97–109, 1970.
- [161] Hatchett, R. J., Mecher, C. E., and Lipsitch, M. Public health interventions and epidemic intensity during the 1918 influenza pandemic. *Proceedings of the National Academy of Sciences*, 104(18):7582–7587, 2007.
- [162] Hayden, F. G., Fritz, R., Lobo, M. C., Alvord, W., Strober, W., and Straus, S. E. Local and systemic cytokine responses during experimental human influenza A virus infection. relation to symptom formation and host defense. *Journal of Clinical Investigation*, 101(3):643–649, 1998.

- [163] Hayden, F. and Gwaltney, J. *Viral infections*. WB Saunders Co, 1994.
- [164] Hayden, F. G. and Palese, P. Influenza virus. In *Clinical Virology, Third Edition*, pages 943–976. American Society of Microbiology, 2009.
- [165] Hayden, F. G., Treanor, J. J., Betts, R. F., Lobo, M., Esinhart, J. D., and Hussey, E. K. Safety and efficacy of the neuraminidase inhibitor GG167 in experimental human influenza. *JAMA*, 275(4):295–299, 1996.
- [166] He, D., Lui, R., Wang, L., Tse, C. K., Yang, L., and Stone, L. Global spatio-temporal patterns of influenza in the post-pandemic era. *Scientific Reports*, 5(1), 2015.
- [167] He, S., Tang, S., and Rong, L. A discrete stochastic model of the COVID-19 outbreak: Forecast and control. *Mathematical Biosciences and Engineering*, 17(4):2792–2804, 2020.
- [168] Heffernan, J. M. and Wahl, L. M. Monte Carlo estimates of natural variation in HIV infection. *Journal of Theoretical Biology*, 236(2):137–153, 2005.
- [169] Heffernan, J. and Keeling, M. Implications of vaccination and waning immunity. *Proceedings of the Royal Society B: Biological Sciences*, 276(1664):2071–2080, 2009.
- [170] Heffernan, J., Smith, R., and Wahl, L. Perspectives on the basic reproductive ratio. *Journal of The Royal Society Interface*, 2(4):281–293, 2005.
- [171] Hellewell, J., Abbott, S., Gimma, A., Bosse, N. I., Jarvis, C. I., Russell, T. W., et al. Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts. *The Lancet Global Health*, 8(4):e488–e496, 2020.
- [172] Hernandez-Vargas, E. A., Wilk, E., Canini, L., Toapanta, F. R., Binder, S. C., Uvarovskii, A., et al. Effects of aging on influenza virus infection dynamics. *Journal of Virology*, 88(8):4123–4131, 2014.
- [173] Hill, E. M., Petrou, S., de Lusignan, S., Yonova, I., and Keeling, M. J. Seasonal influenza: Modelling approaches to capture immunity propagation. *PLoS Computational Biology*, 15(10):1–26, 2019.
- [174] Hoertel, N., Blachier, M., Blanco, C., Olsson, M., Massetti, M., Rico, M. S., et al. A stochastic agent-based model of the SARS-CoV-2 epidemic in France. *Nature Medicine*, 26(9):1417–1421, 2020.
- [175] Holder, B. P. and Beauchemin, C. A. Exploring the effect of biological delays in kinetic models of influenza within a host or cell culture. *BMC Public Health*, 11(Suppl 1):S10, 2011.

- [176] Hollingsworth, T. D. Controlling infectious disease outbreaks: Lessons from mathematical modelling. *Journal of Public Health Policy*, 30(3):328–341, 2009.
- [177] Hollingsworth, T. D., Ferguson, N. M., and Anderson, R. M. Will travel restrictions control the international spread of pandemic influenza? *Nature Medicine*, 12(5):497–499, 2006.
- [178] Hollingsworth, T. D., Ferguson, N. M., and Anderson, R. M. Frequent travelers and rate of spread of epidemics. *Emerging Infectious Diseases*, 13(9):1288–1294, 2007.
- [179] Hollingsworth, T. D., Klinkenberg, D., Heesterbeek, H., and Anderson, R. M. Mitigation strategies for pandemic influenza A: Balancing conflicting policy objectives. *PLoS Computational Biology*, 7(2):e1001076, 2011.
- [180] Honce, R. and Schultz-Cherry, S. Impact of obesity on influenza A virus pathogenesis, immune response, and evolution. *Frontiers in Immunology*, 10, 2019.
- [181] Huang, Q. S., Wood, T., Jelley, L., Jennings, T., Jefferies, S., Daniells, K., et al. Impact of the COVID-19 nonpharmaceutical interventions on influenza and other respiratory viral infections in New Zealand. *Nature Communications*, 12(1), 2021.
- [182] Huang, S. S. H., Lin, Z., Banner, D., León, A. J., Paquette, S. G., Rubin, B., et al. Immunity toward H1N1 influenza hemagglutinin of historical and contemporary strains suggests protection and vaccine failure. *Scientific Reports*, 3(1), 2013.
- [183] Hufford, M. M., Kim, T. S., Sun, J., and Braciale, T. J. The effector T cell response to influenza infection. In *Current Topics in Microbiology and Immunology*, pages 423–455. Springer International Publishing, 2014.
- [184] Hui, S., Chu, L., Peiris, J., Chan, K., Chu, D., and Tsui, W. Immune response to influenza vaccination in community-dwelling Chinese elderly persons. *Vaccine*, 24(25): 5371–5380, 2006.
- [185] Isle of Man Government. COVID-19 Vaccination Programme. <https://covid19.gov.im/about-coronavirus/open-data-downloads/>, 2021. Accessed on 15/04/21.
- [186] Isle of Man Government. Frequently asked questions about the vaccine. <https://covid19.gov.im/vaccination/frequently-asked-questions-about-the-vaccine/>, 2021. Accessed on 21/04/21.
- [187] Iuliano, A. D., Roguski, K. M., Chang, H. H., Muscatello, D. J., Palekar, R., Tempia, S., et al. Estimates of global seasonal influenza-associated respiratory mortality: a modelling study. *The Lancet*, 391(10127):1285–1300, 2018.

- [188] Iwasaki, A. and Pillai, P. S. Innate immunity to influenza virus infection. *Nature Reviews Immunology*, 14(5):315–328, 2014.
- [189] Iwasaki, T. and Nozima, T. Defense mechanisms against primary influenza virus infection in mice: I. the roles of interferon and neutralizing antibodies and thymus dependence of interferon and antibody production. *The Journal of Immunology*, 118(1): 256–263, 1977.
- [190] Jagers, P. and Nerman, O. Branching processes in periodically varying environment. *The Annals of Probability*, pages 254–268, 1985.
- [191] Jahedi, S. and Yorke, J. A. When the best pandemic models are the simplest. *Biology*, 9(11):353, 2020.
- [192] Janeway Jr, C. A., Travers, P., Walport, M., and Shlomchik, M. J. Principles of innate and adaptive immunity. In *Immunobiology: The Immune System in Health and Disease. 5th edition*. Garland Science, 2001.
- [193] Johnson, N. P. and Mueller, J. Updating the accounts: global mortality of the 1918–1920 “Spanish” influenza pandemic. *Bulletin of the History of Medicine*, 76(1):105–115, 2002.
- [194] Kamal, M. A., Gieschke, R., Lemenuel-Diot, A., Beauchemin, C. A. A., Smith, P. F., and Rayner, C. R. A drug-disease model describing the effect of oseltamivir neuraminidase inhibition on influenza virus progression. *Antimicrobial Agents and Chemotherapy*, 59(9):5388–5395, 2015.
- [195] Kamigaki, T. and Oshitani, H. Epidemiological characteristics and low case fatality rate of pandemic (H1N1) 2009 in Japan. *PLoS Currents*, 1:RRN1139, 2009.
- [196] Kash, J. C., Qi, L., Dugan, V. G., Jagger, B. W., Hrabal, R. J., Memoli, M. J., et al. Prior infection with classical swine H1N1 influenza viruses is associated with protective immunity to the 2009 pandemic H1N1 virus. *Influenza and Other Respiratory Viruses*, 4(3):121–127, 2010.
- [197] Keeling, M. J. and Rohani, P. *Modeling Infectious Diseases in Humans and Animals*. Princeton University Press, 2011.
- [198] Keeling, M. and Ross, J. On methods for studying stochastic disease dynamics. *Journal of The Royal Society Interface*, 5(19):171–181, 2007.
- [199] Keeling, M., Moore, S., L Dyson L, M. T., and Hill, E. Road map scenarios and sensitivity. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/975911/S1184_SPI-M_University_of_Warwick_Road_Map_Scenarios_and_Sensitivity.pdf, 2021. Accessed on 01/05/21.

- [200] Kelly, H., Peck, H. A., Laurie, K. L., Wu, P., Nishiura, H., and Cowling, B. J. The age-specific cumulative incidence of infection with pandemic influenza H1N1 2009 was similar in various countries prior to vaccination. *PLoS One*, 6(8):e21828, 2011.
- [201] Kermack, W. O. and McKendrick, A. G. A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, 115(772):700–721, 1927.
- [202] Kilbourne, E. D. Influenza pandemics of the 20th century. *Emerging Infectious Diseases*, 12(1):9–14, 2006.
- [203] Killingley, B. and Nguyen-Van-Tam, J. Routes of influenza transmission. *Influenza and Other Respiratory Viruses*, 7:42–51, 2013.
- [204] Kim, H., Webster, R. G., and Webby, R. J. Influenza virus: dealing with a drifting and shifting pathogen. *Viral Immunology*, 31(2):174–183, 2018.
- [205] Kim, K. S., Ejima, K., Iwanami, S., Fujita, Y., Ohashi, H., Koizumi, Y., et al. A quantitative model used to compare within-host SARS-CoV-2, MERS-CoV, and SARS-CoV dynamics provides insights into the pathogenesis and treatment of SARS-CoV-2. *PLoS Biology*, 19(3):e3001128, 2021.
- [206] Kissler, S. M., Tedijanto, C., Goldstein, E., Grad, Y. H., and Lipsitch, M. Projecting the transmission dynamics of SARS-CoV-2 through the postpandemic period. *Science*, 368(6493):860–868, 2020.
- [207] Klein, B. and MacDonald, P. D. M. The multitype continuous-time Markov branching process in a periodic environment. *Advances in Applied Probability*, 12(1):81–93, 1980.
- [208] Klepac, P., Kissler, S., and Gog, J. Contagion! the BBC Four Pandemic—the model behind the documentary. *Epidemics*, 24:49–59, 2018.
- [209] Koelle, K., Cobey, S., Grenfell, B., and Pascual, M. Epochal evolution shapes the phylodynamics of interpandemic influenza A (H3N2) in humans. *Science*, 314(5807):1898–1903, 2006.
- [210] Krammer, F. The human antibody response to influenza A virus infection and vaccination. *Nature Reviews Immunology*, 19(6):383–397, 2019.
- [211] Kreijtz, J., Fouchier, R., and Rimmelzwaan, G. Immune responses to influenza virus infection. *Virus Research*, 162(1-2):19–30, 2011.
- [212] Kucharski, A. J., Klepac, P., Conlan, A. J. K., Kissler, S. M., Tang, M. L., Fry, H., et al. Effectiveness of isolation, testing, contact tracing, and physical distancing on reducing

transmission of SARS-CoV-2 in different settings: a mathematical modelling study. *The Lancet Infectious Diseases*, 20(10):1151–1160, 2020.

- [213] La Gruta, N. L., Kedzierska, K., Stambas, J., and Doherty, P. C. A question of self-preservation: immunopathology in influenza virus infection. *Immunology & Cell Biology*, 85(2):85–92, 2007.
- [214] Lahodny, G. E. and Allen, L. J. S. Probability of a disease outbreak in stochastic multipatch epidemic models. *Bulletin of Mathematical Biology*, 75(7):1157–1180, 2013.
- [215] Lai, S., Ruktanonchai, N. W., Zhou, L., Prosper, O., Luo, W., Floyd, J. R., et al. Effect of non-pharmaceutical interventions to contain COVID-19 in China. *Nature*, 585(7825):410–413, 2020.
- [216] Lambert, N. D., Ovsyannikova, I. G., Pankratz, V. S., Jacobson, R. M., and Poland, G. A. Understanding the immune response to seasonal influenza vaccination in older adults: a systems biology approach. *Expert Review of Vaccines*, 11(8):985–994, 2012.
- [217] Laurie, K. L., Carolan, L. A., Middleton, D., Lowther, S., Kelso, A., and Barr, I. G. Multiple infections with seasonal influenza A virus induce cross-protective immunity against A(H1N1) pandemic influenza virus in a ferret model. *The Journal of Infectious Diseases*, 202(7):1011–1020, 2010.
- [218] Luring, A. S. and Hodcroft, E. B. Genetic variants of SARS-CoV-2—what do they mean? *JAMA*, 325(6):529, 2021.
- [219] Lavezzo, E., Franchin, E., Ciavarella, C., Cuomo-Dannenburg, G., Barzon, L., Vecchio, C. D., et al. Suppression of a SARS-CoV-2 outbreak in the Italian municipality of Vo’. *Nature*, 584(7821):425–429, 2020.
- [220] Lee, H. Y., Topham, D. J., Park, S. Y., Hollenbaugh, J., Treanor, J., Mosmann, T. R., et al. Simulation and prediction of the adaptive immune response to influenza A virus infection. *Journal of Virology*, 83(14):7151–7165, 2009.
- [221] Leung, K., Wu, J. T., Liu, D., and Leung, G. M. First-wave COVID-19 transmissibility and severity in China outside Hubei after control measures, and second-wave scenario planning: a modelling impact assessment. *The Lancet*, 395(10233):1382–1393, 2020.
- [222] Levins, R., Awerbuch, T., Brinkmann, U., Eckardt, I., Epstein, P., Makhoul, N., et al. The emergence of new diseases. *American Scientist*, 82(1):52–60, 1994.
- [223] Lewnard, J. A., Mbah, M. L. N., Alfaro-Murillo, J. A., Altice, F. L., Bawo, L., Nyenswah, T. G., and Galvani, A. P. Dynamics and control of Ebola virus transmission in Montserrado, Liberia: a mathematical modelling analysis. *The Lancet Infectious Diseases*, 14(12):1189–1195, 2014.

- [224] Li, R., Pei, S., Chen, B., Song, Y., Zhang, T., Yang, W., and Shaman, J. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). *Science*, 368(6490):489–493, 2020.
- [225] Li, Y. and Handel, A. Modeling inoculum dose dependent patterns of acute virus infections. *Journal of Theoretical Biology*, 347:63–73, 2014.
- [226] Liu, Q., hong Zhou, Y., and qiu Yang, Z. The cytokine storm of severe influenza and development of immunomodulatory therapy. *Cellular & Molecular Immunology*, 13(1): 3–10, 2015.
- [227] Liu, Y., Morgenstern, C., Kelly, J., Lowe, R., and Jit, M. The impact of non-pharmaceutical interventions on SARS-CoV-2 transmission across 130 countries and territories. *BMC Medicine*, 19(1), 2021.
- [228] Lloyd-Smith, J. O., Schreiber, S. J., Kopp, P. E., and Getz, W. M. Superspreading and the effect of individual variation on disease emergence. *Nature*, 438(7066):355–359, 2005.
- [229] Longini, I. M., Nizam, A., Xu, S., Ungchusak, K., Hanshaoworakul, W., Cummings, D. A., and Halloran, M. E. Containing pandemic influenza at the source. *Science*, 309 (5737):1083–1087, 2005.
- [230] Longini Jr, I. M., Halloran, M. E., Nizam, A., and Yang, Y. Containing pandemic influenza with antiviral agents. *American Journal of Epidemiology*, 159(7):623–633, 2004.
- [231] Lovell-Read, F. A., Funk, S., Obolski, U., Donnelly, C. A., and Thompson, R. N. Interventions targeting nonsymptomatic cases can be important to prevent local outbreaks: SARS-CoV-2 as a case-study. *Journal of the Royal Society Interface*, 18(178):20201014, 2021.
- [232] Lowe, R., Barcellos, C., Brasil, P., Cruz, O., Honório, N., Kuper, H., and Carvalho, M. The Zika virus epidemic in Brazil: From discovery to future implications. *International Journal of Environmental Research and Public Health*, 15(1):96, 2018.
- [233] Lukens, S., DePasse, J., Rosenfeld, R., Ghedin, E., Mochan, E., Brown, S. T., et al. A large-scale immuno-epidemiological simulation of influenza A epidemics. *BMC Public Health*, 14(1), 2014.
- [234] Lumen Learning. The Viral Life Cycle. <https://courses.lumenlearning.com/microbiology/chapter/the-viral-life-cycle/>, 2021. Accessed on 08/05/21.

- [235] Mackenzie, C., Taylor, P., and Askonas, B. Rapid recovery of lung histology correlates with clearance of influenza virus by specific CD8+ cytotoxic T cells. *Immunology*, 67(3):375, 1989.
- [236] Mahase, E. Covid-19: UK offers under 40s alternative to AstraZeneca vaccine to boost confidence. *BMJ*, page n1185, 2021.
- [237] Mahase, E. Covid-19: Booster dose will be needed in autumn to avoid winter surge, says government adviser. *BMJ*, page n664, 2021.
- [238] Mahase, E. Covid-19: Novavax vaccine efficacy is 86% against UK variant and 60% against South African variant. *BMJ*, page n296, 2021.
- [239] Maines, T. R., Szretter, K. J., Perrone, L., Belser, J. A., Bright, R. A., Zeng, H., et al. Pathogenesis of emerging avian influenza viruses in mammals and the host innate immune response. *Immunological Reviews*, 225(1):68–84, 2008.
- [240] Manchanda, H., Seidel, N., Krumbholz, A., Sauerbrei, A., Schmidtke, M., and Guthke, R. Within-host influenza dynamics: A small-scale mathematical modeling approach. *Biosystems*, 118:51–59, 2014.
- [241] Mancuso, P. Obesity and respiratory infections: Does excess adiposity weigh down host defense? *Pulmonary Pharmacology & Therapeutics*, 26(4):412–419, 2013.
- [242] Marino, S., Hogue, I. B., Ray, C. J., and Kirschner, D. E. A methodology for performing global uncertainty and sensitivity analysis in systems biology. *Journal of Theoretical Biology*, 254(1):178–196, 2008.
- [243] Martcheva, M. and Pilyugin, S. S. An epidemic model structured by host immunity. *Journal of Biological Systems*, 14(02):185–203, 2006.
- [244] McClain, M. T., Henao, R., Williams, J., Nicholson, B., Veldman, T., Hudson, L., et al. Differential evolution of peripheral cytokine levels in symptomatic and asymptomatic responses to experimental influenza virus challenge. *Clinical & Experimental Immunology*, 183(3):441–451, 2015.
- [245] McGill COVID-19 Vaccine Tracker Team. COVID-19 Vaccine Tracker. <https://covid19.trackvaccines.org/vaccines/>, 2021. Accessed on 08/06/21.
- [246] Medlock, J. and Galvani, A. P. Optimizing influenza vaccine distribution. *Science*, 325(5948):1705–1708, 2009.
- [247] Meeker, W. Q. and Escobar, L. A. Teaching about approximate confidence regions based on maximum likelihood estimation. *The American Statistician*, 49(1):48, 1995.

- [248] Menni, C., Valdes, A. M., Freidin, M. B., Sudre, C. H., Nguyen, L. H., Drew, D. A., et al. Real-time tracking of self-reported symptoms to predict potential COVID-19. *Nature Medicine*, 26(7):1037–1040, 2020.
- [249] Merler, S., Ajelli, M., Fumanelli, L., Parlamento, S., y Piontti, A. P., Dean, N. E., et al. Containing Ebola at the source with ring vaccination. *PLoS Neglected Tropical Diseases*, 10(11):e0005093, 2016.
- [250] Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21(6):1087–1092, 1953.
- [251] Meyers, L. A., Newman, M., and Pourbohloul, B. Predicting epidemics on directed contact networks. *Journal of Theoretical Biology*, 240(3):400–418, 2006.
- [252] Miao, H., Hollenbaugh, J. A., Zand, M. S., Holden-Wiltse, J., Mosmann, T. R., Perelson, A. S., et al. Quantifying the early immune response and adaptive immune response kinetics in mice infected with influenza A virus. *Journal of Virology*, 84(13):6687–6698, 2010.
- [253] Miao, H., Xia, X., Perelson, A. S., and Wu, H. On identifiability of nonlinear ODE models and applications in viral dynamics. *SIAM Review*, 53(1):3–39, 2011.
- [254] Miller, E., Hoschler, K., Hardelid, P., Stanford, E., Andrews, N., and Zambon, M. Incidence of 2009 pandemic influenza A H1N1 infection in England: a cross-sectional serological study. *The Lancet*, 375(9720):1100–1108, 2010.
- [255] Miller, J. C. and Kiss, I. Z. Epidemic spread in networks: Existing methods and current challenges. *Mathematical Modelling of Natural Phenomena*, 9(2):4–42, 2014.
- [256] Mills, C. E., Robins, J. M., and Lipsitch, M. Transmissibility of 1918 pandemic influenza. *Nature*, 432(7019):904–906, 2004.
- [257] Miotto, R., Li, L., Kidd, B. A., and Dudley, J. T. Deep patient: An unsupervised representation to predict the future of patients from the electronic health records. *Scientific Reports*, 6(1), 2016.
- [258] Mizumoto, K., Yamamoto, T., and Nishiura, H. Age-dependent estimates of the epidemiological impact of pandemic influenza (H1N1-2009) in Japan. *Computational and Mathematical Methods in Medicine*, 2013:1–8, 2013.
- [259] Montaseri, G., Boianelli, A., Hernandez-Vargas, E. A., and Meyer-Hermann, M. PK/PD-based adaptive tailoring of oseltamivir doses to treat within-host influenza viral infections. *Progress in Biophysics and Molecular Biology*, 139:31–42, 2018.

- [260] Monto, A. S. Global burden of influenza: what we know and what we need to know. *International Congress Series*, 1263:3–11, 2004.
- [261] Moore, J., Ahmed, H., Jia, J., Akondy, R., Ahmed, R., and Antia, R. What controls the acute viral infection following yellow fever vaccination? *Bulletin of Mathematical Biology*, 80(1):46–63, 2017.
- [262] Morrill, J. H., Kormilitzin, A., Nevado-Holgado, A. J., Swaminathan, S., Howison, S. D., and Lyons, T. J. Utilization of the signature method to identify the early onset of sepsis from multivariate physiological time series in critical care monitoring. *Critical Care Medicine*, 48(10):e976–e981, 2020.
- [263] Moxnes, J. F. and Hausken, K. Mathematical modelling of acute virus influenza A infections. *Mathematical and Computer Modelling of Dynamical Systems*, 18(5):521–538, 2012.
- [264] Mozdzanowska, K., Maiese, K., Furchner, M., and Gerhard, W. Treatment of influenza virus-infected SCID mice with nonneutralizing antibodies specific for the transmembrane proteins matrix 2 and neuraminidase reduces the pulmonary virus titer but fails to clear the infection. *Virology*, 254(1):138–146, 1999.
- [265] Mummert, A. and Otunuga, O. M. Parameter identification for a stochastic SEIRS epidemic model: case study influenza. *Journal of Mathematical Biology*, 79(2):705–729, 2019.
- [266] Murillo, L. N., Murillo, M. S., and Perelson, A. S. Towards multiscale modeling of influenza infection. *Journal of Theoretical Biology*, 332:267–290, 2013.
- [267] Murphy, B. R., Kasel, J. A., and Chanock, R. M. Association of serum anti-neuraminidase antibody with resistance to influenza in man. *New England Journal of Medicine*, 286(25):1329–1332, 1972.
- [268] Myers, M. A., Smith, A. P., Lane, L. C., Moquin, D. J., Vogel, P., Woolard, S., and Smith, A. M. Dynamically linking influenza virus infection with lung injury to predict disease severity. *bioRxiv*, 2019.
- [269] Nabel, G. J. and Fauci, A. S. Induction of unnatural immunity: prospects for a broadly protective universal influenza vaccine. *Nature Medicine*, 16(12):1389–1391, 2010.
- [270] National Institute of Infectious Diseases. Isolation/detection of influenza virus in Japan, 2017/18 season. <https://www.niid.go.jp/niid/en/influenza-e.html>, 2020. Accessed on 30/11/20.

- [271] National Institute of Population and Social Security Research. All Japan: Life Table Data Series | Japanese Mortality Database | National Institute of Population and Social Security Research. <http://www.ipss.go.jp/p-toukei/JMD/00/index-en.html>, 2020. Accessed on 22/01/20.
- [272] Neumann, G. and Kawaoka, Y. Predicting the next influenza pandemics. *The Journal of Infectious Diseases*, 219(Suppl.1):S14–S20, 2019.
- [273] Nguyen, H. H., Zemlin, M., Ivanov, I. I., Andradi, J., Zemlin, C., Vu, H. L., et al. Heterosubtypic immunity to influenza A virus infection requires a properly diversified antibody repertoire. *Journal of Virology*, 81(17):9331–9338, 2007.
- [274] Nguyen, V. K., Klawonn, F., Mikolajczyk, R., and Hernandez-Vargas, E. A. Analysis of practical identifiability of a viral infection model. *PLoS One*, 11(12):e0167568, 2016.
- [275] Nguyen, V. K., Mikolajczyk, R., and Hernandez-Vargas, E. A. High-resolution epidemic simulation using within-host infection and contact data. *BMC Public Health*, 18(1), 2018.
- [276] Nicholson, K. G., Wood, J. M., and Zambon, M. Influenza. *The Lancet*, 362(9397): 1733–1745, 2003.
- [277] Nipa, K. F. and Allen, L. J. S. Disease emergence in multi-patch stochastic epidemic models with demographic and seasonal variability. *Bulletin of Mathematical Biology*, 82 (12), 2020.
- [278] Nishiura, H. Real-time forecasting of an epidemic using a discrete time stochastic model: a case study of pandemic influenza (H1N1-2009). *BioMedical Engineering OnLine*, 10 (1):15, 2011.
- [279] Nishiura, H., Yan, P., Sleeman, C. K., and Mode, C. J. Estimating the transmission potential of supercritical processes based on the final size distribution of minor outbreaks. *Journal of Theoretical Biology*, 294:48–55, 2012.
- [280] Nishiura, H., Jung, S.-m., Linton, N. M., Kinoshita, R., Yang, Y., Hayashi, K., et al. The extent of transmission of novel coronavirus in Wuhan, China, 2020. *Journal of Clinical Medicine*, 9(2):330, 2020.
- [281] Norris, J. R. *Markov chains*. Cambridge University Press, 1997.
- [282] Nowak, M. and May, R. M. *Virus dynamics: mathematical principles of immunology and virology: mathematical principles of immunology and virology*. Oxford University Press, 2000.

- [283] Ohkusa, Y., Sugawara, T., et al. Simulation model of pandemic influenza in the whole of Japan. *Japanese Journal of Infectious Diseases*, 62(2):98–106, 2009.
- [284] Omori, R. and Nishiura, H. Theoretical basis to measure the impact of short-lasting control of an infectious disease on the epidemic peak. *Theoretical Biology and Medical Modelling*, 8(1), 2011.
- [285] O’Neill, P. D. A tutorial introduction to Bayesian inference for stochastic epidemic models using Markov chain Monte Carlo methods. *Mathematical Biosciences*, 180(1-2): 103–114, 2002.
- [286] Our World in Data. Coronavirus (COVID-19) vaccinations. <https://ourworldindata.org/covid-vaccinations>, 2021. Accessed on 23/04/21.
- [287] Outlaw, M. C. and Dimmock, N. J. Insights into neutralization of animal viruses gained from study of influenza virus. *Epidemiology and Infection*, 106(2):205–220, 1991.
- [288] Padilla-Quirarte, H. O., Lopez-Guerrero, D. V., Gutierrez-Xicotencatl, L., and Esquivel-Guadarrama, F. Protective antibodies against influenza proteins. *Frontiers in Immunology*, 10, 2019.
- [289] Parkin, J. and Cohen, B. An overview of the immune system. *The Lancet*, 357(9270): 1777–1789, 2001.
- [290] Parra-Rojas, C., Nguyen, V. K., Hernandez-Mejia, G., Hernandez-Vargas, E. A., et al. Neuraminidase inhibitors in influenza treatment and prevention—is it time to call it a day? *Viruses*, 10(9):454, 2018.
- [291] Patrozou, E. and Mermel, L. A. Does influenza transmission occur from asymptomatic infection or prior to symptom onset? *Public Health Reports*, 124(2):193–196, 2009.
- [292] Pawelek, K. A., Huynh, G. T., Quinlivan, M., Cullinane, A., Rong, L., and Perelson, A. S. Modeling within-host dynamics of influenza virus infection including immune responses. *PLoS Computational Biology*, 8(6):e1002588, 2012.
- [293] Pawelek, K. A., Dor, D., Salmeron, C., and Handel, A. Within-host models of high and low pathogenic influenza virus infections: The role of macrophages. *PLoS One*, 11(2): e0150568, 2016.
- [294] Pearson, J. E., Krapivsky, P., and Perelson, A. S. Stochastic theory of early viral infection: Continuous versus burst production of virions. *PLoS Computational Biology*, 7(2):e1001058, 2011.
- [295] Peiris, J. S. M., Guan, Y., and Yuen, K. Y. Severe acute respiratory syndrome. *Nature Medicine*, 10(S12):S88–S97, 2004.

- [296] Peiris, J., Hui, K. P., and Yen, H.-L. Host response to influenza virus: protection versus immunopathology. *Current Opinion in Immunology*, 22(4):475–481, 2010.
- [297] Penman, B. S., Gupta, S., and Shanks, G. D. Rapid mortality transition of Pacific Islands in the 19th century. *Epidemiology and Infection*, 145(1):1–11, 2016.
- [298] Perelson, A. S. Modelling viral and immune system dynamics. *Nature Reviews Immunology*, 2(1):28–36, 2002.
- [299] Perez, L. and Dragicevic, S. An agent-based approach for modeling dynamics of contagious disease spread. *International Journal of Health Geographics*, 8(1):50, 2009.
- [300] Perlman, S. Another decade, another coronavirus. *New England Journal of Medicine*, 382(8):760–762, 2020.
- [301] Petrova, V. N. and Russell, C. A. The evolution of seasonal influenza viruses. *Nature Reviews Microbiology*, 16(1):47–60, 2017.
- [302] Pillai, P. S., Molony, R. D., Martinod, K., Dong, H., Pang, I. K., Tal, M. C., et al. Mx1 reveals innate pathways to antiviral resistance and lethal influenza disease. *Science*, 352(6284):463–466, 2016.
- [303] Pinky, L. and Dobrovolny, H. M. Coinfections of the respiratory tract: Viral competition for resources. *PLoS One*, 11(5):e0155589, 2016.
- [304] Pinky, L. and Dobrovolny, H. M. The impact of cell regeneration on the dynamics of viral coinfection. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 27(6):063109, 2017.
- [305] Pinky, L., Gonzalez-Parra, G., and Dobrovolny, H. M. Effect of stochasticity on coinfection dynamics of respiratory viruses. *BMC Bioinformatics*, 20(1), 2019.
- [306] Pohjanpalo, H. System identifiability based on the power series expansion of the solution. *Mathematical Biosciences*, 41(1-2):21–33, 1978.
- [307] Polack, F. P., Thomas, S. J., Kitchin, N., Absalon, J., Gurtman, A., Lockhart, S., et al. Safety and efficacy of the BNT162b2 mRNA Covid-19 vaccine. *New England Journal of Medicine*, 383(27):2603–2615, 2020.
- [308] Pouillot, R., Lachenal, G., Pybus, O. G., Rousset, D., and Njouom, R. Variable epidemic histories of hepatitis C virus genotype 2 infection in West Africa and Cameroon. *Infection, Genetics and Evolution*, 8(5):676–681, 2008.
- [309] Pourbohloul, B., Ahued, A., Davoudi, B., Meza, R., Meyers, L. A., Skowronski, D. M., et al. Initial human transmission dynamics of the pandemic (H1N1) 2009 virus in North America. *Influenza and Other Respiratory Viruses*, 3(5):215–222, 2009.

- [310] Poynard, T., Yuen, M.-F., Ratzin, V., and Lai, C. L. Viral hepatitis C. *The Lancet*, 362(9401):2095–2100, 2003.
- [311] Prem, K., Liu, Y., Russell, T. W., Kucharski, A. J., Eggo, R. M., Davies, N., et al. The effect of control strategies to reduce social mixing on outcomes of the COVID-19 epidemic in Wuhan, China: a modelling study. *The Lancet Public Health*, 5(5):e261–e270, 2020.
- [312] Pritchard, E., Matthews, P. C., Stoesser, N., Eyre, D. W., Gethings, O., Vihta, K.-D., et al. Impact of vaccination on SARS-CoV-2 cases in the community: a population-based study using the UK’s COVID-19 infection survey. *medRxiv*, 2021.
- [313] Rajaram, S., Wiecek, W., Lawson, R., Blak, B. T., Zhao, Y., Hackett, J., et al. Impact of increased influenza vaccination in 2–3-year-old children on disease burden within the general population: A Bayesian model-based approach. *PLoS One*, 12(12):e0186739, 2017.
- [314] Raue, A., Kreutz, C., Maiwald, T., Bachmann, J., Schilling, M., Klingmüller, U., and Timmer, J. Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. *Bioinformatics*, 25(15):1923–1929, 2009.
- [315] Raue, A., Karlsson, J., Saccomani, M. P., Jirstrand, M., and Timmer, J. Comparison of approaches for parameter identifiability analysis of biological systems. *Bioinformatics*, 30(10):1440–1448, 2014.
- [316] Rawlins, E. L. and Hogan, B. L. M. Ciliated epithelial cell lifespan in the mouse trachea and lung. *American Journal of Physiology-Lung Cellular and Molecular Physiology*, 295(1):L231–L234, 2008.
- [317] Ray, E. L., Wattanachit, N., Niemi, J., Kanji, A. H., House, K., Cramer, E. Y., et al. Ensemble forecasts of coronavirus disease 2019 (COVID-19) in the U.S. *medRxiv*, 2020.
- [318] Razai, M. S., Osama, T., McKechnie, D. G. J., and Majeed, A. Covid-19 vaccine hesitancy among ethnic minority groups. *BMJ*, page n513, 2021.
- [319] Reichert, T., Chowell, G., and McCullers, J. A. The age distribution of mortality due to influenza: pandemic and peri-pandemic. *BMC Medicine*, 10(1):162, 2012.
- [320] Reluga, T. C., Medlock, J., and Perelson, A. S. Backward bifurcations and multiple equilibria in epidemic models with structured immunity. *Journal of Theoretical Biology*, 252(1):155–165, 2008.

- [321] Renaud, C., Boudreault, A. A., Kuypers, J., Lofy, K. H., Corey, L., Boeckh, M. J., and Englund, J. A. H275Y mutant pandemic (H1N1) 2009 virus in immunocompromised patients. *Emerging Infectious Diseases*, 17(4):653–660, 2011.
- [322] Reperant, L. A., Kuiken, T., Grenfell, B. T., and Osterhaus, A. D. M. E. The immune response and within-host emergence of pandemic influenza virus. *The Lancet*, 384(9959): 2077–2081, 2014.
- [323] Ribeiro, R. M., Lo, A., and Perelson, A. S. Dynamics of hepatitis B virus infection. *Microbes and Infection*, 4(8):829–835, 2002.
- [324] Roda, W. C., Varughese, M. B., Han, D., and Li, M. Y. Why is it difficult to accurately predict the COVID-19 epidemic? *Infectious Disease Modelling*, 5:271–281, 2020.
- [325] Roosa, K., Lee, Y., Luo, R., Kirpich, A., Rothenberg, R., Hyman, J., et al. Real-time forecasts of the COVID-19 epidemic in China from February 5th to February 24th, 2020. *Infectious Disease Modelling*, 5:256–263, 2020.
- [326] Rossman, H., Shilo, S., Meir, T., Gorfine, M., Shalit, U., and Segal, E. COVID-19 dynamics after a national immunization program in Israel. *Nature Medicine*, 27(6): 1055–1061, 2021.
- [327] Rouse, B. T. and Sehrawat, S. Immunity and immunopathology to viruses: what decides the outcome? *Nature Reviews Immunology*, 10(7):514–526, 2010.
- [328] Russell, T. W., Wu, J. T., Clifford, S., Edmunds, W. J., Kucharski, A. J., and Jit, M. Effect of internationally imported cases on internal spread of COVID-19: a mathematical modelling study. *The Lancet Public Health*, 6(1):e12–e20, 2021.
- [329] Sachak-Patwa, R., Byrne, H. M., and Thompson, R. N. Accounting for cross-immunity can improve forecast accuracy during influenza epidemics. *Epidemics*, 34:100432, 2021.
- [330] Saenz, R. A., Quinlivan, M., Elton, D., MacRae, S., Blunden, A. S., Mumford, J. A., et al. Dynamics of influenza virus infection and pathology. *Journal of Virology*, 84(8): 3974–3983, 2010.
- [331] Saéz, A. M., Weiss, S., Nowak, K., Lapeyre, V., Zimmermann, F., Dux, A., et al. Investigating the zoonotic origin of the West African Ebola epidemic. *EMBO Molecular Medicine*, 7(1):17–23, 2014.
- [332] Sallam, M. COVID-19 vaccine hesitancy worldwide: A concise systematic review of vaccine acceptance rates. *Vaccines*, 9(2):160, 2021.
- [333] Saltelli, A., Tarantola, S., and Chan, K. P.-S. A quantitative model-independent method for global sensitivity analysis of model output. *Technometrics*, 41(1):39–56, 1999.

- [334] Saltelli, A., Annoni, P., Azzini, I., Campolongo, F., Ratto, M., and Tarantola, S. Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index. *Computer Physics Communications*, 181(2):259–270, 2010.
- [335] Schulman, J. L., Khakpour, M., and Kilbourne, E. D. Protective effects of specific immunity to viral neuraminidase on influenza virus infection of mice. *Journal of Virology*, 2(8):778–786, 1968.
- [336] Shah, A. S. V., Gribben, C., Bishop, J., Hanlon, P., Caldwell, D., Wood, R., et al. Effect of vaccination on transmission of COVID-19: an observational study in healthcare workers and their households. *medRxiv*, 2021.
- [337] Shaman, J., Karspeck, A., Yang, W., Tamerius, J., and Lipsitch, M. Real-time influenza forecasts during the 2012–2013 season. *Nature Communications*, 4(1), 2013.
- [338] Siekmann, I., Sneyd, J., and Crampin, E. J. MCMC can detect nonidentifiable models. *Biophysical Journal*, 103(11):2275–2286, 2012.
- [339] Simon, P. F., de La Vega, M.-A., Paradis, É., Mendoza, E., Coombs, K. M., Kobasa, D., and Beauchemin, C. A. A. Avian influenza viruses that cause highly virulent infections in humans exhibit distinct replicative properties in contrast to human H1N1 viruses. *Scientific Reports*, 6(1), 2016.
- [340] Singer, M. *Introduction to syndemics: A critical systems approach to public and community health*. John Wiley & Sons, 2009.
- [341] Skountzou, I., Koutsouanos, D. G., Kim, J. H., Powers, R., Satyabham, L., Maseoud, F., et al. Immunity to pre-1950 H1N1 influenza viruses confers cross-protection against the pandemic swine-origin 2009 A (H1N1) influenza virus. *The Journal of Immunology*, 185(3):1642–1649, 2010.
- [342] Smith, A. P., Moquin, D. J., Bernhauerova, V., and Smith, A. M. Influenza virus infection model with density dependence supports biphasic viral decay. *Frontiers in Microbiology*, 9:1554, 2018.
- [343] Smith, A. M. Host-pathogen kinetics during influenza infection and coinfection: insights from predictive modeling. *Immunological Reviews*, 285(1):97–112, 2018.
- [344] Smith, A. M. Validated models of immune response to virus infection. *Current Opinion in Systems Biology*, 12:46–52, 2018.
- [345] Smith, A. M. and Ribeiro, R. M. Modeling the viral dynamics of influenza A virus infection. *Critical ReviewsTM in Immunology*, 30(3):291–298, 2010.

- [346] Smith, A. M., Adler, F. R., McAuley, J. L., Gutenkunst, R. N., Ribeiro, R. M., McCullers, J. A., and Perelson, A. S. Effect of 1918 PB1-F2 expression on influenza A virus infection kinetics. *PLoS Computational Biology*, 7(2):e1001081, 2011.
- [347] Smith, A. M., Adler, F. R., Ribeiro, R. M., Gutenkunst, R. N., McAuley, J. L., McCullers, J. A., and Perelson, A. S. Kinetics of coinfection with influenza A virus and *Streptococcus pneumoniae*. *PLoS Pathogens*, 9(3):e1003238, 2013.
- [348] Sobol', I. Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Mathematics and Computers in Simulation*, 55(1-3):271–280, 2001.
- [349] Sompayrac, L. M. *How the immune system works*. John Wiley & Sons, 2019.
- [350] Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and van der Linde, A. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(4):583–639, 2002.
- [351] Stark, G. R., Kerr, I. M., Williams, B. R. G., Silverman, R. H., and Schreiber, R. D. How cells respond to interferons. *Annual Review of Biochemistry*, 67(1):227–264, 1998.
- [352] Stevens, P. Diseases of poverty and the 10/90 gap. <https://www.who.int/intellectualproperty/submissions/InternationalPolicyNetwork.pdf>, 2004. Accessed on 01/06/21.
- [353] Streftaris, G. and Gibson, G. J. Bayesian inference for stochastic epidemics in closed populations. *Statistical Modelling*, 4(1):63–75, 2004.
- [354] Sudre, C. H., Lee, K. A., Lochlainn, M. N., Varsavsky, T., Murray, B., Graham, M. S., et al. Symptom clusters in COVID-19: A potential clinical prediction tool from the COVID symptom study app. *Science Advances*, 7(12):eabd4177, 2021.
- [355] Sym, D., Patel, P. N., and Ei-Chaar, G. M. Seasonal, avian, and novel H1N1 influenza: Prevention and treatment modalities. *Annals of Pharmacotherapy*, 43(12):2001–2011, 2009.
- [356] Takeuchi, O. and Akira, S. Innate immunity to virus infection. *Immunological Reviews*, 227(1):75–86, 2009.
- [357] Tang, X., Wu, C., Li, X., Song, Y., Yao, X., Wu, X., et al. On the origin and continuing evolution of SARS-CoV-2. *National Science Review*, 7(6):1012–1023, 2020.
- [358] Taubenberger, J. K. and Morens, D. M. 1918 influenza: the mother of all pandemics. *Emerging Infectious Diseases*, 12(1):15–22, 2006.

- [359] Tea, F., Stella, A. O., Aggarwal, A., Darley, D. R., Pilli, D., Vitale, D., et al. SARS-CoV-2 neutralizing antibodies longevity, breadth, and evasion by emerging viral variants. *medRxiv*, 2020.
- [360] Texier, G., Jackson, M. L., Siwe, L., Meynard, J.-B., Deparis, X., and Chaudet, H. Building test data from real outbreaks for evaluating detection algorithms. *PLoS One*, 12(9):e0183992, 2017.
- [361] Thompson, R. N., Thompson, C. P., Pelerman, O., Gupta, S., and Obolski, U. Increased frequency of travel in the presence of cross-immunity may act to decrease the chance of a global pandemic. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 374(1775):20180274, 2019.
- [362] Thompson, R. N., Gilligan, C. A., and Cunliffe, N. J. Will an outbreak exceed available resources for control? Estimating the risk from invading pathogens using practical definitions of a severe epidemic. *Journal of The Royal Society Interface*, 17(172):20200690, 2020.
- [363] Thompson, R., Stockwin, J., van Gaalen, R. D., Polonsky, J., Kamvar, Z., Demarsh, P., et al. Improved inference of time-varying reproduction numbers during infectious disease outbreaks. *Epidemics*, 29:100356, 2019.
- [364] Thompson, R. N. Epidemiological models are important tools for guiding COVID-19 interventions. *BMC Medicine*, 18(1), 2020.
- [365] Thompson, R. N. Novel coronavirus outbreak in Wuhan, China, 2020: Intense surveillance is vital for preventing sustained transmission in new locations. *Journal of Clinical Medicine*, 9(2):498, 2020.
- [366] Thompson, R. N., Gilligan, C. A., and Cunliffe, N. J. Detecting presymptomatic infection is necessary to forecast major epidemics in the earliest stages of infectious disease outbreaks. *PLoS Computational Biology*, 12(4):e1004836, 2016.
- [367] Thompson, R. N., Hollingsworth, T. D., Isham, V., Arribas-Bel, D., Ashby, B., Britton, T., et al. Key questions for modelling COVID-19 exit strategies. *Proceedings of the Royal Society B: Biological Sciences*, 287(1932):20201405, 2020.
- [368] Tisoncik, J. R., Korth, M. J., Simmons, C. P., Farrar, J., Martin, T. R., and Katze, M. G. Into the eye of the cytokine storm. *Microbiology and Molecular Biology Reviews*, 76(1):16–32, 2012.
- [369] Tizzoni, M., Bajardi, P., Poletto, C., Ramasco, J. J., Balcan, D., Gonçalves, B., et al. Real-time numerical forecast of global epidemic spreading: case study of 2009 A/H1N1pdm. *BMC Medicine*, 10(1), 2012.

- [370] Torjesen, I. Covid-19 will become endemic but with decreased potency over time, scientists believe. *BMJ*, page n494, 2021.
- [371] Tridane, A. and Kuang, Y. Modeling the interaction of cytotoxic T lymphocytes and influenza virus infected epithelial cells. *Mathematical Biosciences and Engineering*, 7(1):171–185, 2010.
- [372] Trifonov, V., Khiabani, H., and Rabadan, R. Geographic dependence, surveillance, and origins of the 2009 influenza A (H1N1) virus. *New England Journal of Medicine*, 361(2):115–119, 2009.
- [373] Tuite, A. R., Greer, A. L., Whelan, M., Winter, A.-L., Lee, B., Yan, P., et al. Estimated epidemiologic parameters and morbidity associated with pandemic H1N1 influenza. *Canadian Medical Association Journal*, 182(2):131–136, 2009.
- [374] van den Driessche, P. and Watmough, J. Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission. *Mathematical Biosciences*, 180(1-2):29–48, 2002.
- [375] van der Vries, E., Stittelaar, K. J., van Amerongen, G., Kroeze, E. J. B. V., de Waal, L., Fraaij, P. L. A., et al. Prolonged influenza virus shedding and emergence of antiviral resistance in immunocompromised patients and ferrets. *PLoS Pathogens*, 9(5):e1003343, 2013.
- [376] Van Kerkhove, M. D., Asikainen, T., Becker, N. G., Bjorge, S., Desenclos, J.-C., Dos Santos, T., et al. Studies needed to address public health challenges of the 2009 H1N1 influenza pandemic: Insights from modeling. *PLoS Medicine*, 7(6):e1000275, 2010.
- [377] Volz, E. and Meyers, L. A. Susceptible–infected–recovered epidemics in dynamic contact networks. *Proceedings of the Royal Society B: Biological Sciences*, 274(1628):2925–2934, 2007.
- [378] Volz, E. M., Miller, J. C., Galvani, A., and Meyers, L. A. Effects of heterogeneous and clustered contact patterns on infectious disease dynamics. *PLoS Computational Biology*, 7(6):e1002042, 2011.
- [379] Voysey, M., Clemens, S. A. C., Madhi, S. A., Weckx, L. Y., Folegatti, P. M., Aley, P. K., et al. Safety and efficacy of the ChAdOx1 nCoV-19 vaccine (AZD1222) against SARS-CoV-2: an interim analysis of four randomised controlled trials in Brazil, South Africa, and the UK. *The Lancet*, 397(10269):99–111, 2021.
- [380] Voysey, M., Clemens, S. A. C., Madhi, S. A., Weckx, L. Y., Folegatti, P. M., Aley, P. K., et al. Single-dose administration and the influence of the timing of the booster dose

on immunogenicity and efficacy of ChAdOx1 nCoV-19 (AZD1222) vaccine: a pooled analysis of four randomised trials. *The Lancet*, 397(10277):881–891, 2021.

- [381] Wang, X., Sontag, D., and Wang, F. Unsupervised learning of disease progression models. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '14*. ACM Press, 2014.
- [382] Wang, Z., Schmidt, F., Weisblum, Y., Muecksch, F., Barnes, C. O., Finkin, S., et al. mRNA vaccine-elicited antibodies to SARS-CoV-2 and circulating variants. *Nature*, 2021.
- [383] Webster, R., Reay, P., and Laver, W. Protection against lethal influenza with neuraminidase. *Virology*, 164(1):230–237, 1988.
- [384] Weiss, H. H. The SIR model and the foundations of public health. *MATerials MATemàtics*, (3):1–17, 2013.
- [385] White, L. F., Moser, C. B., Thompson, R. N., and Pagano, M. Statistical estimation of the reproductive number from case notification data. *American Journal of Epidemiology*, 190(4):611–620, 2020.
- [386] Widdowson, M.-A., Bresee, J. S., and Jernigan, D. B. The global threat of animal influenza viruses of zoonotic concern: Then and now. *The Journal of Infectious Diseases*, 216(suppl_4):S493–S498, 2017.
- [387] Wise, J. Covid-19: Pfizer BioNTech vaccine reduced cases by 94% in Israel, shows peer reviewed study. *BMJ*, page n567, 2021.
- [388] World Health Organization. Influenza (Avian and other zoonotic). [https://www.who.int/news-room/fact-sheets/detail/influenza-\(avian-and-other-zoonotic\)](https://www.who.int/news-room/fact-sheets/detail/influenza-(avian-and-other-zoonotic)), 2018. Accessed on 03/04/21.
- [389] World Health Organization. Influenza (Seasonal). [https://www.who.int/en/news-room/fact-sheets/detail/influenza-\(seasonal\)](https://www.who.int/en/news-room/fact-sheets/detail/influenza-(seasonal)), 2020. Accessed on 04/11/20.
- [390] World Health Organization. WHO COVID-19 Dashboard. <https://covid19.who.int/>, 2021. Accessed on 08/06/21.
- [391] Worldometer. Current world population. <https://www.worldometers.info/world-population/>, 2021. Accessed on 21/04/21.
- [392] Xiang, Y., Jia, Y., Chen, L., Guo, L., Shu, B., and Long, E. COVID-19 epidemic prediction and the impact of public health interventions: A review of COVID-19 epidemic models. *Infectious Disease Modelling*, 6:324–342.

- [393] Xing, Z. and Cardona, C. J. Preexisting immunity to pandemic (H1N1) 2009. *Emerging Infectious Diseases*, 15(11):1847–1849, 2009.
- [394] Xu, R., Ekiert, D. C., Krause, J. C., Hai, R., Crowe, J. E., and Wilson, I. A. Structural basis of preexisting immunity to the 2009 H1N1 pandemic influenza virus. *Science*, 328(5976):357–360, 2010.
- [395] Xu, Y., Allen, L. J., and Perelson, A. S. Stochastic model of an influenza epidemic with drug resistance. *Journal of Theoretical Biology*, 248(1):179–193, 2007.
- [396] Yan, A. W. C., Cao, P., and McCaw, J. M. On the extinction probability in models of within-host infection: the role of latency and immunity. *Journal of Mathematical Biology*, 73(4):787–813, 2016.
- [397] Yan, A. W. *Influenza viral dynamics models to explore the roles of innate and adaptive immunity*. PhD thesis, Imperial College London, 2017.
- [398] Yang, W., Lipsitch, M., and Shaman, J. Inference of seasonal and pandemic influenza transmission dynamics. *Proceedings of the National Academy of Sciences*, 112(9):2723–2728, 2015.
- [399] Yu, X., Tsibane, T., McGraw, P. A., House, F. S., Keefer, C. J., Hicar, M. D., et al. Neutralizing antibodies derived from the B cells of 1918 influenza pandemic survivors. *Nature*, 455(7212):532–536, 2008.
- [400] Yu, X., Zhang, X., Zhao, B., Wang, J., Zhu, Z., Teng, Z., et al. Intensive cytokine induction in pandemic H1N1 influenza virus infection accompanied by robust production of IL-10 and IL-6. *PLoS One*, 6(12):e28680, 2011.
- [401] Zelner, J., Riou, J., Etzioni, R., and Gelman, A. Accounting for uncertainty during a pandemic. *Patterns*, 2(8):100310, 2021.
- [402] Zhang, X.-Y., Trame, M., Lesko, L., and Schmidt, S. Sobol sensitivity analysis: A tool to guide the development and evaluation of systems pharmacology models. *CPT: Pharmacometrics & Systems Pharmacology*, 4(2):69–79, 2015.
- [403] Zhou, P., Yang, X.-L., Wang, X.-G., Hu, B., Zhang, L., Zhang, W., et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*, 579(7798):270–273, 2020.
- [404] Zhuang, Z., Zhao, S., Lin, Q., Cao, P., Lou, Y., Yang, L., et al. Preliminary estimates of the reproduction number of the coronavirus disease (COVID-19) outbreak in Republic of Korea and Italy by 5 March 2020. *International Journal of Infectious Diseases*, 95:308–310, 2020.

- [405] Zou, D., Wang, L., Xu, P., Chen, J., Zhang, W., and Gu, Q. Epidemic model guided machine learning for COVID-19 forecasts in the United States. *medRxiv*, 2020.
- [406] Zou, L., Ruan, F., Huang, M., Liang, L., Huang, H., Hong, Z., et al. SARS-CoV-2 viral load in upper respiratory specimens of infected patients. *New England Journal of Medicine*, 382(12):1177–1179, 2020.
- [407] Zumla, A., Hui, D. S., and Perlman, S. Middle East respiratory syndrome. *The Lancet*, 386(9997):995–1007, 2015.
- [408] Zuo, J., Dowell, A. C., Pearce, H., Verma, K., Long, H. M., Begum, J., et al. Robust SARS-CoV-2-specific T cell immunity is maintained at 6 months following primary infection. *Nature Immunology*, 22(5):620–626, 2021.