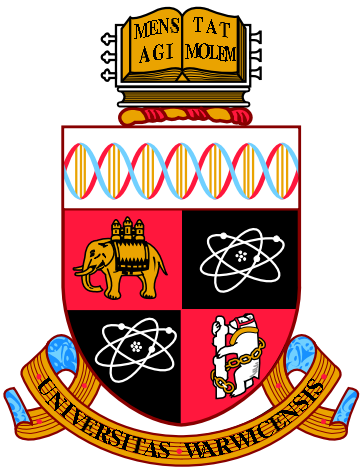

Mathematical modelling approaches for spreading processes: zoonotic influenza and social contagion

by

Edward Hill

Thesis

Submitted to the University of Warwick

for the degree of

Doctor of Philosophy

Centre for Complexity Science

July 2017

THE UNIVERSITY OF WARWICK

Contents

List of Tables

List of Figures

Acknowledgments

I would firstly like to thank my supervisors, Mike Tildesley and Thomas House, for their guidance and invaluable support over the past few years. I have enjoyed working with them both immensely and this would not have been possible without them! I also thank the staff of the Complexity Science Centre for their help in all matters, and the WIDER/SBIDER group at Warwick for their ideas and feedback.

Finally, I am especially grateful to my family and friends for their constant support in all my endeavours, and for the encouragement received throughout the writing of this thesis.

This work was supported by the Engineering and Physical Sciences Research Council as part of the University of Warwick Complexity Science Doctoral Training Centre. The Bangladesh poultry premises data was made available to us through the Bangladesh Department of Livestock Services. Colleagues at FAO-ECTAD (Emergency Centre for Transboundary Animal Diseases) office in Bangladesh are thanked for their contribution. This research also uses data from Add Health, a programme project directed by Kathleen Mullan Harris, and designed by J. Richard Udry, Peter S. Bearman and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill. Special acknowledgement is due to Ronald R. Rindfuss and Barbara Entwisle for assistance in the original design.

Declarations

The work presented here is my own, except where stated otherwise. This thesis has been composed by myself and has not been submitted for any other degree or professional qualification.

Chapter 3 and its accompanying appendix (appendix A) have been published as

- E. M. Hill, T. House, M. S. Dhingra, W. Kalpravidh, S. Morzaria, M. G. Osmani, M. Yamage, X. Xiao, M. Gilbert and M. J. Tildesley. 2017.
  Modelling H5N1 in Bangladesh across spatial scales: Model complexity and zoonotic transmission risk.
  Epidemics. doi:10.1016/j.epidem.2017.02.007.

Chapter 5 has been published as

- E. M. Hill, M. J. Tildesley and T. House. 2017.
  Evidence for history-dependence of influenza pandemic emergence.
  Scientific Reports **7**:43623. doi:10.1038/srep43623.

Chapter 6 and its accompanying appendix (appendix C) have been published as

- E. M. Hill, F. E. Griffiths and T. House. 2015.
  Spreading of healthy mood in adolescent social networks.
  Proceedings of the Royal Society B **282**(1813):20151180.
  doi: 10.1098/rspb.2015.1180.

Abstract

Mathematical models are a fundamental component of many epidemiological studies. While models of infectious disease are well established, there are evident methodological gaps when attempting to provide realistic descriptions of particular biological systems. In this thesis we probe questions related to two global public health problems, zoonotic influenza and depression, requiring innovative modelling approaches to be developed, analysed and fitted to data. We give particular consideration to parameter inference schemes to gain insights into the dynamics of these illnesses, and model simulation for validation and prediction purposes, including assessing intervention impact.

First, we investigate zoonotic influenza transmission at a local scale, our example being H5N1 in Bangladesh. It is vital to devise new models incorporating zoonotic transmission, and establish the factors enabling both continued transmission within poultry and spillover across the poultry-human divide. We outline a set of candidate transmission models, with a zoonotic transmission component, parameterised with a Bayesian inference scheme using data from two H5N1 outbreaks in the Dhaka region. Applied at two distinct spatial scales, we elucidate the model considerations that best capture the size and spatial distribution of reported cases. Simulations then illustrate the predicted impact of interventions designed to reduce H5N1 transmission.

Second, the emergence of influenza strains with pandemic potential is considered from a global viewpoint. Using a Bayesian model selection approach we compare plausible model hypotheses regarding the mechanisms driving influenza pandemic occurrences. Analysing the time periods between putative influenza pandemics since 1700, it is shown the weight of evidence favours influenza pandemic emergence being history-dependent, rather than a memoryless process. Predictive distributions are then presented for the expected number of pandemic events from 2010 to 2110.

Third, spread of behaviour-linked health problems are amenable to being represented with methodological approaches typically used to model infectious diseases. We explore this with regards to depression, using a longitudinal dataset comprising information on both the in-school friendships and mood status of US adolescents. A novel model is described that exploits the dynamical behaviour of mood over time to ascertain which mood states spread on social networks, via a contagion-like mechanism, and which do not.

Chapter 1

Introduction

Diseases have been a great affliction on animals and humans alike throughout the course of history. These can be classified as to whether they are infectious, with the disease-causing pathogen spreading from host to host, or non-infectious, where the medical condition or disease is not caused by infectious agents.

While there are notorious examples of the grave impact of infectious disease epidemics, such as the Black Death in Europe (1347-50 and subsequent recurrences) amongst humans and the UK Foot-and-mouth disease (FMD) outbreak amongst commercial livestock (2001), there are significant on-going control costs attributed to combating infectious diseases that are endemic, such as bovine tuberculosis in the UK.

Non-infectious diseases encompass conditions such as cardiovascular diseases (e.g. heart attacks, strokes), cancers, chronic respiratory diseases (such as asthma) and mental disorders. They tend to be of long duration and are the result of a combination of genetic, physiological, environmental and behavioural factors. There is burgeoning interest in whether conditions in humans that are typically viewed as non-communicable, but whose risk of occurrence are influenced by behaviours that are modifiable, exhibit spread from person to person in a manner analogous to infectious pathogens.

In light of these challenges, and the hope of averting the prospect of future epidemic and illness outbreaks where possible, the field of epidemiology endeavours to study and analyse the distribution and determinants of health and disease conditions, with the application of findings acquired from such investigative work to the control of

disease and other health problems.

A significant component of epidemiological work is analytical studies involving the use of mathematical models. This arose in the early twentieth century through the seminal works of Ross [1], and Kermack and McKendrick [2], who outlined a mechanistic interpretation for infectious pathogen dynamics and how they progress through time. The influence of opinion and/or behavioural factors on infectious disease dynamics has since been approached in such a manner [3–6], while the possible spread of behavioural-linked health problems is amenable to being represented and analysed with this methodological approach [7, 8].

In each of these cases a real-world epidemiological problem is being observed and studied via the application of the mathematical modelling cycle (figure 1.1). Following the recognition of the issue, the next step is to make the problem as precise as possible. One important aspect of this step is to identify and select those concepts to be considered as basic in the study and to define them carefully. This involves simplifying the information as much as possible and eliminating what is deemed unnecessary, typically through making certain idealisations and approximations. Having established a qualitative description of the problem of interest, the formulation of a mathematical model can then take place whose solution, or analysis in the case of theory construction, enables us to better understand the original situation. To complete the cycle, as a form of verification we must check whether the model can capably produce outputs that have reasonable correspondence with what has been observed. As an additional option, if the models are considered satisfactory then they may be used to make outbreak projections and assess the impact of proposed control initiatives. The process usually proceeds through several iterations, with multiple models and/or scenarios being evaluated (and undergoing refinement) to determine which provides the best correspondence with the empirical data available.

This thesis concerns the application of the above mathematical modelling cycle procedure to analyse two prominent public health issues, namely zoonotic influenza and depression. The following chapters present our extensions to pre-existing mathematical models, complemented by the development of novel model frameworks, that are subsequently fitted to data and utilised to yield biological insights into these diseases/medical conditions that are of great epidemiological concern.

In **chapter 2** we provide background information relevant to this thesis. This

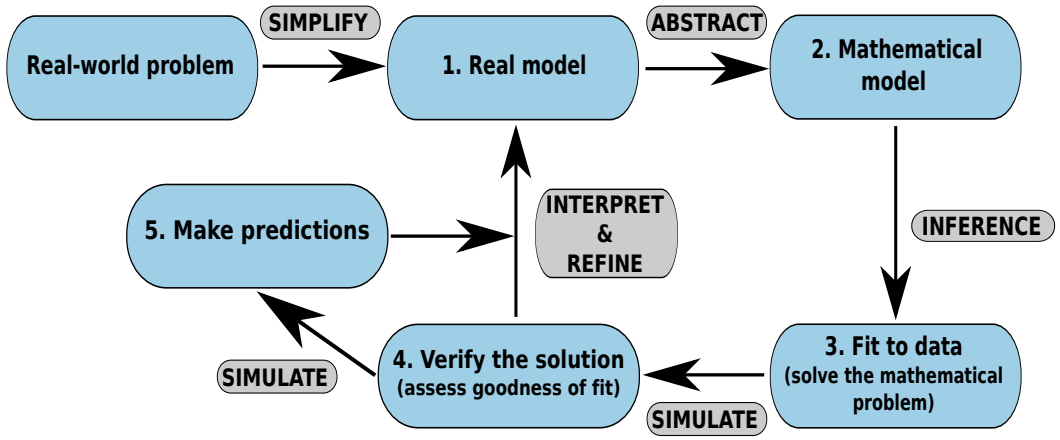

Figure 1.1: Schematic diagram of the mathematical modelling cycle used throughout the thesis. The procedure used to analyse the real-world epidemiological problems we are concerned with can be broken down into five constituent parts. Stage one - Following recognition of the issue, a precise, qualitative description of the problem is made through simplifying information and eliminating what is deemed unnecessary. Stage two - Formulation of a mathematical model to describe the simplified problem. Stage three - Obtain solutions to the mathematical problem by fitting the model to applicable data and inferring model parameter values. Stage four - Verify through computer simulation that the model can capably produce outputs that have reasonable correspondence with what has been observed. Stage five - If the models are considered satisfactory, then a second simulation application is to use the models to make outbreak predictions and assess the impact of proposed control initiatives. Following the completion of one cycle outputs can be interpreted, with proposed models undergoing refinement if required.

covers influenza A and depression in a global context, and introduces relevant technical background and methodological approaches. We show how theoretical and data-driven studies alike have so far been used to study the emergence of novel pathogens and social contagion. We then describe the particular statistical methods used in fitting our models to data through parameter inference, and the variety of simulation techniques employed to generate stochastic realisations of these models for verification and prediction purposes, these both being pertinent features of the mathematical modelling cycle (figure 1.1) and the work described in the ensuring chapters.

Few dynamic models of zoonotic disease currently incorporate the transmission from the animal reservoir to humans. It is of the utmost importance that a new generation of modelling approaches is devised for zoonotic pathogens that incorporate transmission at the human-animal interface. These can subsequently be connected

to real-life epidemics through data fitting. One particular case where the inclusion of this aspect should be encouraged is within model frameworks for H5N1 avian influenza, as mathematical modelling thus far has generally only quantified poultry transmission parameters. In **chapter 3** we present a modelling framework that incorporates zoonotic transmission of avian influenza at the human-poultry interface, in addition to within-poultry disease dynamics. We fit our set of candidate models to two H5N1 epidemic outbreaks in the Dhaka region of Bangladesh (in 2008 and 2011 respectively), focusing on two distinct administration levels to elucidate the crucial modelling considerations that are necessary to best capture the size and spatial distribution of reported cases at different spatial scales. We find that identifying a suitable within-poultry transmission model of minimal complexity is dependent on the administration level being analysed. For the zoonotic transmission component, we reveal that the main contributor to spillover transmission of H5N1 from poultry to humans in Bangladesh differed from one poultry epidemic to another. Across spatial scales, a consistent outcome of non-optimal reporting of infected premises indicates we should seek procedural improvements that will reduce the notification time. These results promote further development of mathematical models for zoonotic transmission to potentially yield important insights into influenza transmission dynamics at the human-animal interface.

Identifying the optimal control policy in response to an emerging disease outbreak is a key challenge for policy-makers. A common approach is to perform simulation studies comparing plausible strategies, while accounting for known capacity restrictions. In **chapter 4**, we utilise the model framework developed in the previous chapter to assess the predicted impact of a variety of interventions in limiting the spread of H5N1 between poultry farms in the Dhaka division of Bangladesh and curbing the risk of zoonotic transmission. We explore how the targeting and implementation of these interventions alters if it is believed transmission is predominately premises-to-premises, versus the scenario where external factors should also be taken into account. While we find that reactive culling and vaccination control policies should pay close attention to this factor to ensure intervention targeting is optimised, targeted proactive active surveillance schemes appear to significantly outperform reactive surveillance procedures in both cases. Our findings may advise the type of control measure, plus its severity, that should be applied in the event of a re-emergent outbreak of H5N1 amongst poultry in Bangladesh.

A perpetual risk to global public health is the emergence of influenza strains with

pandemic potential. Our understanding of the mechanisms driving the appearance of such strains is typically informed by indirect measurements from expensive and sometime dangerous laboratory and field surveillance efforts. In **chapter 5**, we present a mathematical modelling approach analysing the waiting times between proposed historic influenza pandemics, which could potentially provide safe and direct validation of proposed assumptions regarding pandemic emergence. We show there is strong support for influenza pandemic emergence being history-dependent, rather than a memoryless process. Our findings may subsequently help inform the type of interventions that would have the greatest impact in reducing the risk of pandemic emergence.

While standard epidemic models are well established in the mathematical modelling of infectious disease, the application of these techniques into other complex spreading processes that occur among society (e.g. opinions, behaviours) is becoming more routine. Improving our understanding of the social processes that drive the epidemiology of mood disorders, such as depression, has the potential to bring about highly significant public health benefits. However, to date, studies investigating whether such disorders spread from person-to-person via a kind of social contagion have been hampered by being unable to distinguish this mechanism from other possible phenomena that could confound any positive findings. In **chapter 6**, we analyse a longitudinal dataset comprising information on both the in-school friendships and mood status of adolescents in the USA to ascertain whether there is evidence for healthy mood and/or depressive symptoms spreading via a contagion-like mechanism. By using the dynamical behaviour of mood over time, our approach allows us to distinguish directly whether transmission of this kind occurs. We reveal statistically significant evidence for spreading of healthy mood, but not for spreading of depressive symptoms. This outcome suggests the hypothesis of enabling networks of friendship between adolescents has the potential to reduce both incidence and prevalence of depression.

Chapter 2

Background

Within this chapter we describe the significant global health burden created by influenza A and depressive disorders (sections 2.1 and 2.2), as well as reviewing the modelling methodological challenges faced thus far in analysing these problems (section 2.3). To permit the completion of each aspect of the mathematical modelling cycle (figure 1.1), we go on to detail the parameter inference schemes used for fitting the models we develop to data (see section 2.4), and the variety of simulation techniques drawn on throughout this work to subsequently generate stochastic realisations of these models for verification and prediction purposes (see section 2.5).

2.1 Influenza A - A persistent pandemic threat

Influenza is a respiratory infection of mammals and birds caused by an RNA virus in the family of Orthomyxoviridae [9]. It incorporates four virus types: types A, B, C, and D. It should be noted that the Influenza D virus is a recently identified genus within the family Orthomyxoviridae [10]. Type A influenza is a zoonotic disease with the ability to inhabit many host species, in addition to having multiple strains. In contrast, types B and C normally infect only humans, thus limiting the opportunity for cross-species transfer [11]. Finally, type D has been confirmed to affect only bovine hosts thus far [10], meaning it is not currently a prominent human epidemiological issue.

The zoonotic capability of influenza A makes it the most significant of the four types in an epidemiological and public health context, associated with most of the widespread seasonal influenza epidemics and the type capable of causing occasional

global pandemics. The natural host of influenza A viruses are an assortment of aquatic bird species. These viruses occasionally spillover into other animal hosts, including domestic poultry, pigs, horses, a variety of carnivores and marine mammals [12]. Sporadically, the viruses adapt to their new animal hosts, leading to enzootic virus circulation for sustained periods. However, apart from a few cases of reputed direct zoonotic transmission of influenza A viruses to humans from wild birds, due to close contact and de-feathering activities [13, 14], humans have been primarily infected with zoonotic influenza viruses via intermediate species to which human exposure is more frequent. Domestic livestock such as pigs and poultry have a key role in this regard (figure 2.1). Influenza A is therefore not considered an eradicable disease, with prevention and control the only realistic goals [15].

For classification purposes, influenza A is further divided into subtypes based on differences in the two most abundant surface proteins, hemagglutinin (HA) and neuraminidase (NA). The ability of influenza A to adapt to a disparate range of species is a consequence of these two surface proteins exhibiting a great extent of variability, with the virus as a whole undergoing two types of immunologically significant evolution: antigenic drift and antigenic shift.

Antigenic drift is a rapid minor genetic variation in currently circulating subtypes, with the viral genes undergoing subtle changes in structure almost annually. These small genetic changes can accumulate over time and result in viruses that are antigenically different. As a result, antibodies created by the body's immune system against older viruses may no longer recognise the 'newer' variant, meaning those individuals are susceptible to infection. This process drives the occurrence of recurrent seasonal influenza epidemics, with the additional outcome of there being a large number of strains in circulation within a population at any given time [9, 11].

Very occasionally, humans become infected with a virus bearing HA and/or NA antigens derived from non-human sources that are essentially novel to humans (i.e. the HA and/or NA surface antigens of such viruses are not initially recognised by the specific human host defence mechanisms). These abrupt, major immunological changes are known as antigenic shift [9, 11, 12]. The usual cause of an antigenic shift event is two or more strains infecting a host at the same time, with reassortment generating a new strain that is a mixture of genetic material from the contributing strains. Due to the viruses meeting with little or no established resistance, they can, following mutation and adaptation to their new host, spread relatively easily in the

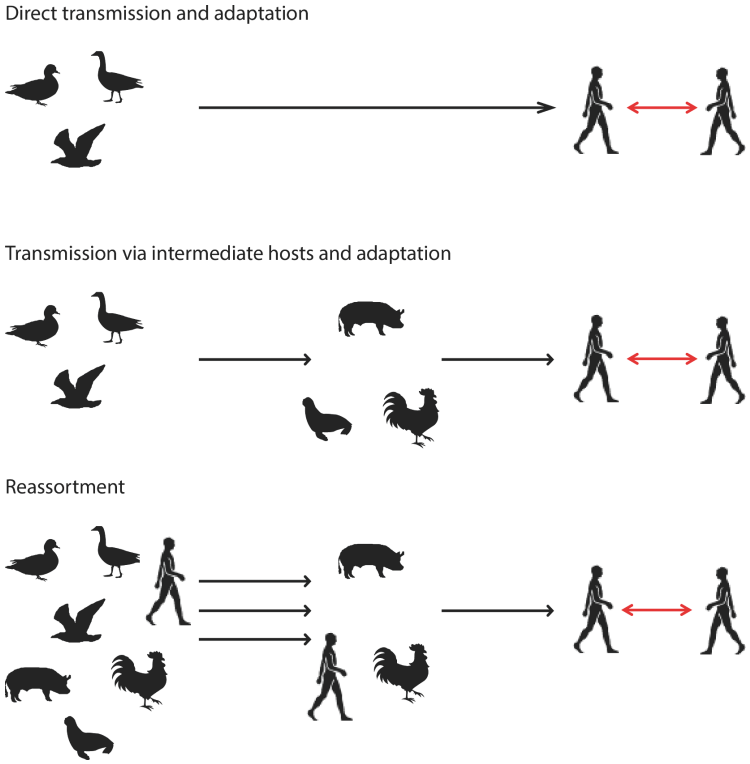

Figure 2.1: Pictorial representation of potential routes of transmission for emergence of pandemic influenza A strains in humans from the animal reservoir. Many scenarios could support the transmission of influenza viruses from their original reservoir to humans and subsequent adaption to transmit via the airborne route. **(Top)** Spillover transmission occurs from the avian reservoir directly to humans. **(Middle)** Crossover from the aquatic bird reservoir to intermediate hosts (such as domestic livestock), followed by adaption in these hosts and subsequent transmission to humans. **(Bottom)** Reassortment in intermediate hosts of influenza viruses originating from diverse animals species and transmission to humans. Two or more strains can infect a host at the same time, with reassortment generating a new strain that is a mixture of genetic material from the contributing strains. Such a process results in abrupt, major immunological changes in the virus. Reproduced from [16].

human species. This can give rise to a localised outbreak that may develop into a worldwide influenza pandemic [12].

Previous influenza pandemics have had a grave impact on humanity. The most recent pandemic (as of writing), caused by H1N1 and commonly referred to as 'swine flu', began in April 2009. Shortly before the end of the pandemic (as of 1st August 2010), more than 214 countries and overseas territories worldwide had reported laboratory confirmed cases of H1N1, including over 18,449 deaths [17] (figure 2.2).

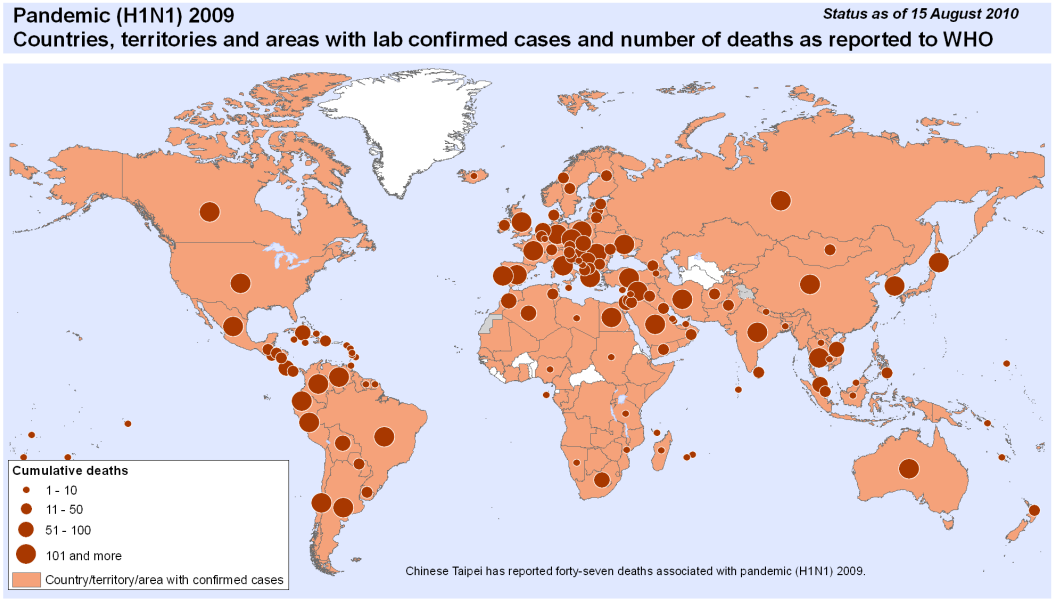

Figure 2.2: Global map of 2009 H1N1 pandemic laboratory confirmed cases and number of deaths as reported to WHO. Shaded regions correspond to countries/territories with confirmed cases. Filled circles signify confirmed deaths in that country/territory, with larger filled circles corresponding to higher cumulative deaths. More than 214 countries and overseas territories or communities had reported laboratory confirmed cases of H1N1, with cumulative deaths of over 18,449. Figure reproduced from the World Health Organisation [18].

It remains of critical importance to understand the pandemic potential of strains widely circulating in the animal reservoir that would be novel to humans, or for which there is already evidence of transmission from animal hosts to humans taking place.

2.1.1 Candidate subtypes with possible pandemic potential

A key cross-species transmission pathway for influenza A viruses is at the poultry-human interface. The threat of a strain with pandemic potential emerging via this route is greater due to the broad diversity and global coverage of influenza viruses within the avian population. In the period January 2014 to November 2016, avian influenza was identified in 77 countries and a total of 13 strains were detected [19].

H5N1

Particularly noteworthy is the H5N1 subtype, which was first observed in southern China during 1996 [20]. This causes a highly infectious, severe respiratory disease

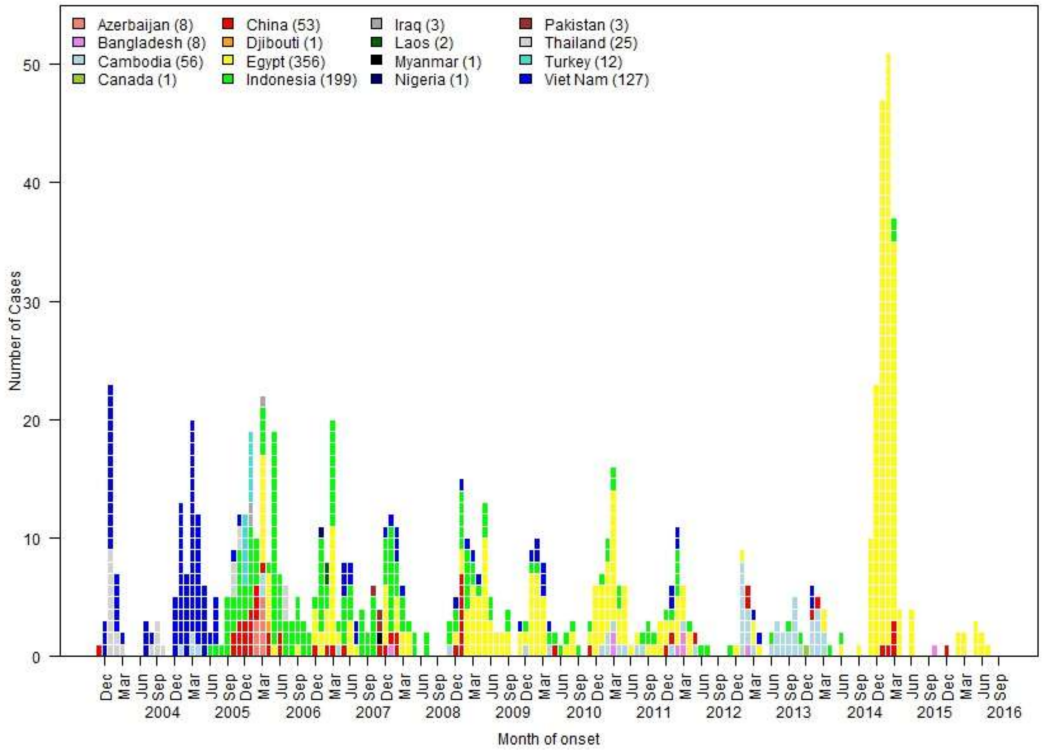

Figure 2.3: Epidemiological curve of avian influenza A(H5N1) cases in humans by month of onset, up to 21 September 2016. Bar heights correspond to case count per month, with the bar segment colour denoting the country reporting the case. In total, there have been over 850 laboratory-confirmed cases reported to WHO. Figure reproduced from the World Health Organisation [25].

in birds (avian influenza). Additionally, if an outbreak occurs, many healthy birds risk being culled to prevent spread of the disease. This has had devastating consequences globally. From 2003 to 2012, the H5N1 virus has killed or forced the culling of more than 400 million domestic poultry and caused an estimated US$20 billion in economic damage, with 63 countries infected at its peak in 2006 [21]. Zoonotic transmission events of H5N1 avian influenza to humans have occurred sporadically (over 850 laboratory-confirmed cases, figure 2.3), with almost all cases to date associated with close contact with infected live or dead birds, or H5N1-contaminated environments [22]. For now, spread from person to person is unusual. The concern, however, is that the overall case-fatality risk is high, at approximately 50-60% [23, 24]. With influenza viruses evolving frequently, the H5N1 virus may become more easily transmissible from human-to-human, posing a serious risk to public health.

As a consequence, H5N1 has been scrutinised in numerous experimental and an-

alytical studies. This encompasses evolutionary analysis [26, 27], shedding experiments [27, 28], national and international spatio-temporal analysis [29–33], and determining risk factors through exploration of data from past epidemics [34–39]. A prominent example of work in the latter category was carried out by Gilbert and Pfeiffer [39], who identified three types of variables with similar statistical association with H5N1 presence across studies and regions: domestic waterfowl, several anthropogenic variables (human population density, distance to roads) and indicators of water presence. Modelling studies to infer transmission rates have also been carried out, using both past epidemic data [40] and data obtained from experimental set ups [41]. We specifically discuss the H5N1 subtype in more depth in chapters 3 and 4, where we focus on developing a mathematical framework for modelling historical H5N1 outbreaks and optimal control policy in Bangladesh.

H7N9

An example of a newly emerging viral strain at the poultry-human interface is H7N9, which arose in China in 2013 through the antigenic shift evolutionary process (figure 2.4). Typically circulating among birds, although some H7 viruses (H7N2, H7N3 and H7N7) have occasionally been found to infect humans, no human infections with H7N9 viruses had been reported previously. Similar to H5N1, this influenza virus subtype is also of considerable concern because most patients have been severely ill, with many suffering from severe pneumonia [42]. Although all cases were sporadic and there was no evidence of an epidemiological link between them [43], it is suggested the pandemic potential of novel H7N9 viruses may be greater than that of highly pathogenic H5N1 viruses, with a further worry being the emergence of H7N9 viruses with resistance against anti-viral treatments [44].

Since being first notified of cases in China on 31st March 2013, over 1300 laboratory-confirmed cases of human infection with avian influenza A(H7N9) viruses have been reported to WHO, with more than 500 deaths (figure 2.5) [45]. According to reports received by the Food and Agriculture Organization (FAO) on surveillance activities for avian influenza A(H7N9) viruses in China, positives among virological samples continue to be detected mainly from live bird markets (LBMs), vendors and some commercial or breeding farms [46]. This ties in with recent exposure to LBMs being linked to a number of cases, while their closure has been found to be highly effective in reducing human risk for H7N9 infection [47].

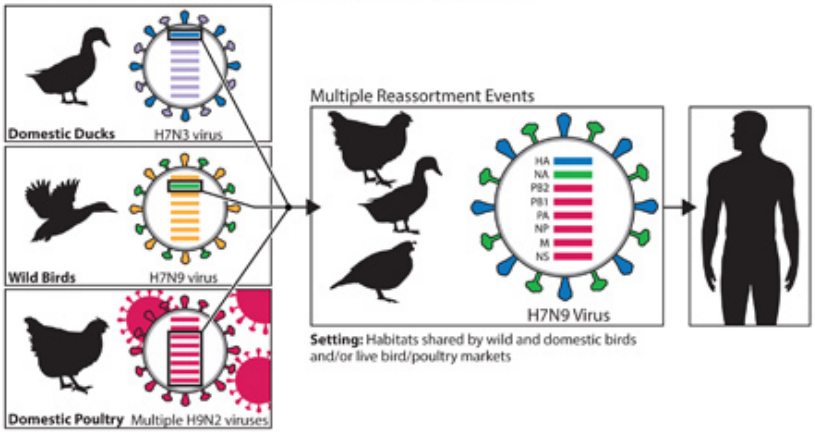

Figure 2.4: Schematic of the genetic evolution of H7N9 virus in China, 2013. The eight genes of the H7N9 virus are closely related to avian influenza viruses found in domestic ducks, wild birds and domestic poultry in Asia. Experts think multiple reassortment events led to the creation of the H7N9 virus, likely obtaining its HA gene from domestic ducks, its NA gene from wild birds, and its six remaining genes from multiple related H9N2 influenza viruses in domestic poultry. Figure reproduced from the Centers for Disease Control and Prevention (CDC) [48].

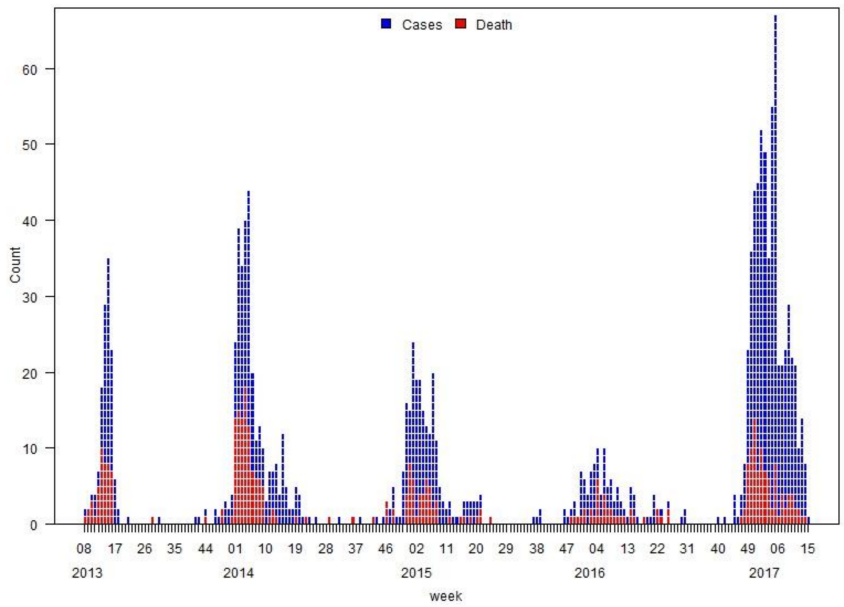

Figure 2.5: Epidemiological curve of avian influenza A(H7N9) cases and deaths in humans by week of onset, up to 20 April 2017. Bar heights correspond to case count per week, with red blocks corresponding to a confirmed human death due to H7N9. There have been over 1300 laboratory-confirmed cases of human infection with avian influenza A(H7N9) virus reported, causing in excess of 500 deaths. Figure reproduced from the World Health Organisation [45].

Other candidate subtypes

A prominent area of ongoing research is determining the risk of spillover transmission into humans of influenza subtypes that have so far remained confined to the animal reservoir. For instance, H9N2 subtype strains have been isolated worldwide from wild and domestic avian species for several decades, but their low pathogenic nature to poultry has made them a low priority for animal disease control [49]. This has allowed them to continue to evolve and spread, with potential to pose a risk to public health as these viruses have been shown to preferentially bind to the human-type receptor, plus be capable of causing disease and transmit between ferrets by respiratory droplet [49, 50]. Gathering interest for similar experimental analysis are newly emerging HPAI H5 viruses (such as H5N8), following widespread incursions in poultry and wild birds. Up to January 2017, over 30 countries had confirmed cases of H5N8, with the majority of European nations affected [51].

2.1.2 Influenza preparedness - planning for the next pandemic

As discussed above, influenza A has garnered substantial attention from the international community due to its ability to cause considerable morbidity and mortality in both humans and livestock. This, in tandem with other zoonotic disease threats emerging in the last few years like Ebola, has motivated a growing literature emphasising the need for improved collaborations between the human and veterinary sciences to address these pandemic threats [52–56].

Specifically, McCloskey *et al.* [52] discuss how action against pandemics is focused on detection and response, not prevention. The authors believe the focus of pandemic preparedness should include upstream prevention through better collaboration between human and animal health sciences to enhance capacity to identify potential pathogens before they become serious human threats, and to prevent their emergence where possible. This is complemented by Al-Tawfiq *et al.* [53] stating that "more effective national, regional, and international surveillance systems are required to enable rapid identification of emerging respiratory epidemics, diseases with epidemic potential, their specific microbial cause, origin, mode of acquisition, and transmission dynamics." This push towards cross-sectoral collaborations to combat infectious disease encourages the adoption of a One Health approach, which aims to improve health and well-being through the prevention of risks and the mitigation of effects of crises that originate at the interface between humans, animals and their various environments [55].

Utilising a One Health approach may help inform strategies to mitigate the impact of the next pandemic. The objective would be to use genetic sequence and/or biological assays of viral traits to identify those non-human influenza viruses with the greatest risk of evolving into pandemic threats, and/or to understand drivers of such evolution, to prioritize pandemic prevention or response measures. Yet, Russell *et al.* [57] summarise four areas requiring improvement: (i) experimental approaches, (ii) computational predictions; (iii) evolutionary theory and modelling, (iv) surveillance methodology. Integrating these approaches would increase the power of tools for more objectively assessing pandemic risk and decrease the time required for assessing the pandemic threat posed by extant non-human influenza A viruses. We make our own contribution to elucidating the mechanism behind historic pandemic emergence in chapter 5.

As well this, there is an urgent need for new tools to assess the pandemic risk posed by a detected virus. A current example of a risk assessment framework is the Influenza Risk Assessment Tool (IRAT), giving a systematic approach for assessing and comparing threats posed primarily by avian and swine influenza viruses [58]. Although the IRAT was not developed to predict the next pandemic influenza virus, it can be used to determine those viruses that are deemed, on the basis of current knowledge, to have the greatest potential to cause a serious pandemic. Once identified, pandemic preparedness resources can then be directed towards those particular strains.

Although it is imperative that preventative measures are worked on and put in place, ensuring they are effective at all times is a considerable challenge. Emergent outbreak events may therefore still occur (albeit less frequently), and other polices must be ready to be enacted when they arise. These actions aim to restrict circulation of viruses in livestock populations, and/or minimise the risk of a cross-species transmission event. See chapter 4, our analysis of H5N1 avian influenza control policy effectiveness in Bangladesh, for further discussion on this topic.

2.2 Social contagion - Global burden of behaviourally-linked conditions

A major contributor to global mortality and morbidity are non-communicable diseases that are linked to modifiable behaviours. The four main types of the aforemen-

tioned afflictions, accounting for over 30 million deaths a year worldwide, are cardiovascular diseases (like heart attacks and stroke), cancers, chronic respiratory diseases (such as chronic obstructive pulmonary disease and asthma), and diabetes [59]. Two of the leading causes of these conditions are attributed to tobacco use and obesity, which arise through smoking and dietary and physical activity behaviours respectively [60]. Specifically, tobacco accounts for around 6 million deaths every year (including from the effects of exposure to second-hand smoke), and is projected to increase to 8 million by 2030 [59]. In the case of obesity, a particular striking fact is that global prevalence has more than doubled between 1980 and 2014, with worldwide estimates from 2014 stating 13% of adults aged 18 years and over were obese [61].

2.2.1 Depressive disorders - A universal affliction

Another class of conditions that may be influenced by modifiable behaviours are mental disorders such as depression, which are a growing affliction upon the health of modern society [62]. Estimates from 2015 put the proportion of the global population living with depression at 4.4% (figure 2.6), corresponding to 322 million people, while the total number of people living with the condition is believed to have increased by approximately 18.4% between 2005 and 2015 (reflecting the overall growth of the global population) [63, 64]. This leads to considerable loses in health and functioning. Depressive disorders are ranked as the single largest contributor to non-fatal health loss worldwide, with a predicted total of over 50 million years lived with disability in 2015 [63, 64].

Adding to these economic costs and human health issues depressive disorders can, in the worst cases, result in attempted suicide. As a result, depression is the major contributor towards suicide deaths, which number close to 800,000 people every year, with further consequences of many millions of people being affected or experiencing suicide bereavement. Of particular concern is that suicide is the second leading cause of death among individuals aged 15 to 29 globally [65].

Therefore, there is a particular focus on the burden of depression among adolescents. In England, the Office of National Statistics Child and Adolescent Mental Health Survey, carried out in 2004, found 1.4% of 11-16 year-olds (approximately 62,000 people) were seriously depressed [66]. In certain instances the prevalence of depressive disorders among young people has been found to exceed the global average. For example, the National Survey on Drug Use and Health in the United States

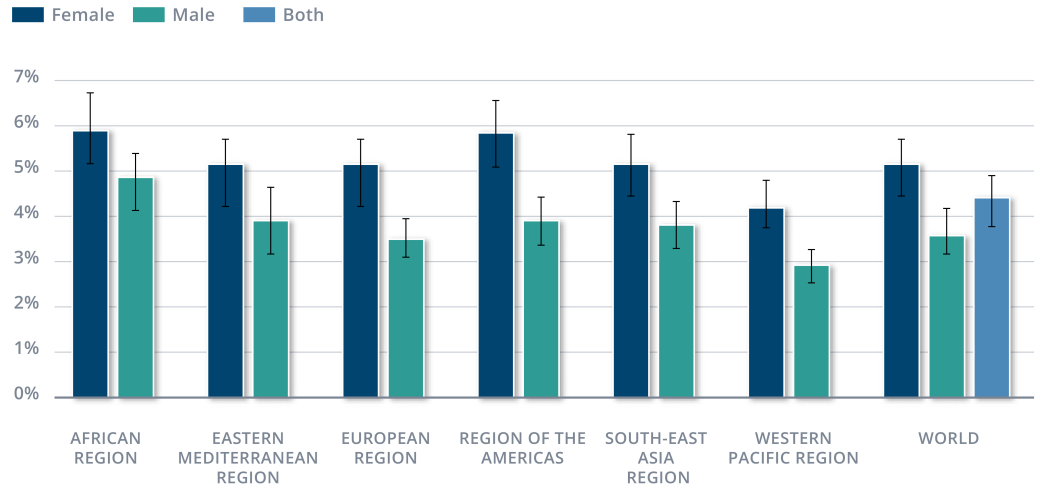

Figure 2.6: Bar plots displaying prevalence of depressive disorders stratified by World Health Organisation regions and gender. Bar heights correspond to the percentage of the population estimated to be living with a depressive disorder. Uncertainty intervals are represented by the error bars. Figure reproduced from the World Health Organisation [64].

estimated three million adolescents aged 12 to 17 had at least one major depressive episode in 2015, representing 12.5% of the population within that age group [67].

Health systems have not yet adequately responded to this global, medical issue, with there being a stark gap between the need for treatment and its provision. In low- and middle-income countries, where more than 80% of the depressive disorder burden occurs [63, 64], between 76% and 85% of people with mental disorders receive no treatment. In high-income countries, between 35% and 50% of individuals living with mental disorders are in the same situation [68].

Attempts to elucidate the mechanisms that generate observed behaviours in social networks that are associated with the diseases, illnesses and disorders described above may provide insights into what would be effective prevention and intervention schemes. This would support efforts to reduce the overall strain of these medical conditions upon public health.

2.3 Mathematical modelling of spreading processes

A major component of epidemiological research is to garner information from mathematical models, which is the approach we also adopt here to probe epidemiological

questions related to zoonotic influenza and depressive disorders. Recalling the description of the mathematical modelling cycle (figure 1.1), such models are a mathematical description of the system of interest, using a set of underlying assumptions to allow us to infer parameters. Following any model refinements being implemented, if required, and verification of model outputs, we can go on to use those findings to project how an outbreak may spread and calculate the effects of possible interventions, thus helping inform public health policy. Details of the approaches used to fit and verify our models are summarised in sections 2.4 and 2.5 respectively.

In the context of generating mathematical descriptions of the spread of disease, pioneers of the initial couple of stages of the modelling cycle included Kermack and McKendrick [2], who introduced a compartmental model for epidemics and the corresponding system of ordinary differential equations (ODEs). Now known as the susceptible-infected-recovered (SIR) model of disease, this model and its variants have been central in the mathematical study of infectious disease [2, 69–71], with compartmental models ubiquitously used. For the interested reader, a numbers of examples of host and pathogen heterogeneities, with appropriate model extensions, are summarised in Keeling and Rohani [72].

In detail, individuals are placed in a compartment corresponding to their disease status and they progress through the compartments according to some rates. The numbers in each class are then treated as continuously varying quantities. In the specific case of the SIR model without the demographic processes of birth and death, each individual can be in one of three states: susceptible to the disease (with individuals in this class labelled S), infectious (labelled I) when they can spread the disease to susceptibles, and recovered (labelled R) when they have been infectious but can no longer spread or catch the disease. The system of ODEs can then be written as:

$$\begin{aligned} \frac{d}{dt}S(t) &= -\frac{\beta}{N}S(t)I(t) \ , \\ \frac{d}{dt}I(t) &= \frac{\beta}{N}S(t)I(t) - \gamma I(t) \ , \\ \frac{d}{dt}R(t) &= \gamma I(t) \ . \end{aligned} \tag{2.1}$$

N denotes the population size, while β and γ are the transmission rate and recovery rate respectively. Due to the absence of demographic processes, we effectively conserve the total number of individuals in the system, corresponding to $S+I+R=N$. Such an assumption is valid if the time-scale of the epidemic is much less than that

of the natural lifespan of the host. Under other conditions, in the case of modelling endemic diseases for example, population size may alter over time meaning demography should be incorporated.

Inspecting equation (2.1) we observe the number of infectious individuals is increasing at time t (i.e. $\frac{dI}{dt} > 0$) if, and only if,

$$\frac{S(t=0)}{N} > \frac{\gamma}{\beta} = \frac{1}{R_0}. \tag{2.2}$$

This gives a condition for the epidemic to be growing in terms of R_0, a fundamental quantity in epidemiology known as the basic reproductive ratio [73, 74]. Defined as the average number of secondary cases produced by an infectious individual in a totally susceptible population, it is a measure for both the biological infectiousness and persistence of the infectious disease and the underlying mixing between infectious and susceptible individuals. If $R_0 < 1$, the infectious disease cannot invade a naive population (one in which every individual, aside from those individuals initially infected, is initially susceptible to the infectious disease under consideration). In contrast, for $R_0 > 1$ the number infected will increase until the susceptibles are sufficiently depleted, leading to a large epidemic. We revisit this concept in chapter 3.

A further result of equation (2.2) is that an infectious disease can only spread if its reproductive ratio is greater than the inverse density of susceptibles in the population. This simple observation is crucial since it gives the threshold proportion of the population that must be removed in order for the entire population to be effectively protected, $V^* = 1 - \frac{1}{R_0}$. Vaccination programmes are one possibility to achieve such a goal. In short, the entire population can be effectively protected from an infectious disease by directly protecting just a fraction of the population. This phenomenon is known as herd immunity, with a review of its standing in the literature provided by Fine [75].

The classical SIR model (equation (2.1)) has several limitations. Among them are assuming all host individuals are identical and mix homogeneously. By relaxing these assumptions, in conjunction with assumptions and arguments on the mixing pattern of the population, the hope is the spatial spread of infectious diseases and behaviours can be accurately captured. Infectious disease transmission between individuals often occurs at close proximity, with long range spreading in a population

due to individual movement. Realistically representing individual movement in a model can be challenging. Transmission kernels are one method of capturing the spatial spread of infection, specifying the relative risk of infection with distance, being prominently used in the spatial modelling of disease epidemics in livestock (the 2001 UK foot-and-mouth disease outbreak for example [76]). Alternatively, it is now common to model infectious diseases as spreading processes on networks [77]. This approach is increasingly applied to behaviours, both related to infectious risk [3]) and non-infectious diseases that are linked to behaviours that can spread socially [78, 79]. In this context, networks depict the social contact structure; whom each individual interacts with and thus the possible routes over which behaviours could be transmitted. This approach is, however, reliant on sufficiently detailed data being available to construct the network.

Furthermore, the ODE model is deterministic; given the same starting conditions, exactly the same trajectory is always observed. Naturally, we would not expect this to apply to the dynamics of contagion spreading processes, especially in a spatial and/or heterogeneous mixing context. An alternative, stochastic, formulation implicitly recognizes that the population is made up of individuals, and that transitions between classes are random events [69, 80, 81]. The rates at which various transitions occur within a deterministic framework can be reinterpreted to calculate the probabilities of each event occurring in an infinitesimal time interval. An added key distinguishing feature of the stochastic formulation versus their deterministic counterparts is that they can exhibit behaviours such as disease fadeout. In other words, while in a stochastic model you may go from $I(t) > 0$ to $I(t + \delta t) = 0$ for finite t, this can not occur in a deterministic model.

Studying real-life epidemic examples highlights the necessity of modifying epidemic models in order to explain and predict observed outcomes. Understanding the impact of spatial structure and contact structure is also of practical importance. Of particular concern is the effect it has on the expected impact of control measures and preventing epidemic outbreaks [82, 83], an outcome of interest in the ensuring chapters. The models presented throughout this work will therefore, in general, be stochastic and incorporate elements of spatial heterogeneity where appropriate. For specific modelling challenges related to the epidemiological problems of concern in this thesis, we refer the reader to sections 2.3.1 and 2.3.2.

2.3.1 Challenges in modelling zoonotic diseases

Although wildlife species are a major source of new human pathogens (such as zoonotic influenza strains) [84], Lloyd-Smith *et al.* [85] highlighted a dearth of mathematical models incorporating spillover transmission from animals to humans. At that time, only six dynamical studies were found that included a mechanistic model of animal-to-human spillover for directly transmitted zoonoses (which includes a number of diseases, not just influenza A).

There are, however, exceptions to the trend of excluding the zoonotic transmission process from modelling frameworks. Hsieh *et al.* [86] presented a novel deterministic compartmental modelling framework for transmissions of H7N9 in China among (wild and domestic) birds as well as from birds to human, to infer epidemiological quantifiers (such as bird-to-human infection rate). This claimed to be the first study where both the avian and human components of an avian influenza epidemic could be quantified using only the human case data. In a similar vain, a subsequent study of H7N9 influenza in southern and eastern regions of China by Li *et al.* [87] used only human case data and a combined framework of mathematical model and Bayesian inference to demonstrate that transmission dynamics within an avian reservoir can be estimated, and that the real-time forecast of spillover avian influenza in humans is possible. However, a limitation of both studies is the assumption of homogeneous mixing of chicken and human populations. This corresponds to spillover transmission from poultry to humans being constant through time. In addition, the risk of poultry infection having no dependency on environmental and/or other external factors means the human case data is effectively a proxy for infection among poultry, which may be an overly strict presumption.

The array of questions in this area warranting further scrutiny is so expansive that an outline of challenges in modelling the emergence of novel pathogens has been produced, spanning reservoir dynamics, cross-species spillover, and outbreak dynamics [88]. The work contained within chapters 3 and 4 of this thesis aims to contribute towards addressing a couple of these challenges, namely: (i) better capturing the disease dynamics in proximal non-human species, and (ii) expanding models for cross-species spillover transmission from general principles to specific, mechanistic frameworks integrating all relevant data types.

An additional modelling consideration is the limitation of lack of sufficient data to parametrise multi-host models. Studies focusing on zoonoses in their animal reser-

voir have relied largely on individual-level parameters, owing to the relative rarity of collated population data for animal diseases [85, 88]. Further, there is little known about the intensity and type of contact patterns between livestock and humans that result in micro-organism transmission, with a better understanding needed to provide options for prevention [89]. Until such collated population data sets can be obtained, potentially at great cost and personal risk via field surveillance efforts, modelling studies on zoonotic diseases will continue to be reliant on individual-level parameters. Motivated by this, chapter 5 describes our attempt to provide direct validation of proposed assumptions regarding pandemic emergence by using mathematical models to analyse the spacing between historic influenza pandemics.

2.3.2 Challenges in modelling social contagion

There is a growing literature on modelling behaviours and non-infectious diseases that are linked to behaviours as spreading processes on networks [3, 7, 90]. Outcomes from recent studies applying such an approach suggest a variety of behaviours and behaviour-based illnesses may spread socially from person to person, akin to transmission of infectious diseases. Examples include obesity [78, 91], smoking [79], loneliness [92], fertility behaviour [93] and happiness [94].

However, such work has come under criticism for being unable to distinguish causal effects from other possible phenomena that produce non-causal associations, namely homophily and shared context [95–98]. Homophily is the tendency for individuals to associate with one another due to sharing the same behaviour. While evidence of assortative mixing and temporal clustering of behaviours among linked nodes is used to support claims of social contagion in networks, homophily may also explain such evidence [96, 97].

Shared context is where individuals tend towards the same behaviour due to external factors. Cohen-Cole and Fletcher [95] demonstrated how the standard methods by Christakis and Fowler [78, 79] may fail to differentiate a causal spreading mechanism from shared context when applied to three health outcomes that are unlikely to be subject to network phenomena: acne, headaches, and height. The authors observed 'network effects' in all three cases. However, this finding no longer held after controlling for environmental confounders.

Consequently, we should be cautious in attributing correlations in health outcomes of social associates to social network effects. This highlights that novel modelling

approaches are required that are capable of distinguishing between transmission and non-transmission mechanisms. To see the first steps we took towards realising this target we refer the reader to chapter 6.

2.4 Parameter inference

Following the development of a dynamic mathematical model to represent the system of interest, the next stage of the mathematical modelling cycle is obtaining a solution to the mathematical problem (stage three of figure 1.1). Throughout the thesis, this corresponds to fitting our models to the available data in order to infer parameter values. These may then be interpreted in a biological context, providing insights on the dynamical behaviour of the system. To ensure that the inferred parameters and model outputs can be interpreted robustly, it is vital to account for parameter uncertainty, as well as stochasticity, arising from the model dynamics. The approaches used here to achieve this are likelihood-based, namely maximum-likelihood estimation (MLE) and Markov chain Monte Carlo (MCMC), with the outputs from these fitting methods having subtle differences.

2.4.1 Maximum-likelihood estimation

If the primary interest is in obtaining point estimates for the parameters of a statistical model given observations, then a suitable statistical approach is the MLE method, which was first introduced by R.A. Fisher in 1922 [99, 100].

For an independent and identically distributed sample of observations, $x_1, x_2, \ldots, x_n$, and underlying statistical model with parameters θ, the joint density function is given by

$$f(x_1, x_2, \ldots, x_n|\theta) = \prod_{i=1}^{n} f(x_i|\theta)$$

Considering the observed values to be fixed 'parameters' of the function, whereas θ is now a variable of the function, we obtain the likelihood L for the sample data, the probability of obtaining that particular set of data given the proposed probability distribution model:

$$L(\theta|x_1, x_2, \ldots, x_n) = f(x_1, x_2, \ldots, x_n|\theta) = \prod_{i=1}^{n} f(x_i|\theta)$$

The values of θ that maximize the sample likelihood are known as the maximum-

likelihood estimators, $\hat{\theta}(D)$. Mathematically,

$$\hat{\theta}(D) = \arg\max_{\theta} L(\theta|x_1, x_2, \ldots, x_n)$$

When no closed-form solution to the maximization problem is known or available, an MLE is instead found numerically using optimization methods.

To quantify parameter uncertainty, asymptotic confidence intervals can be ascertained through calculation of the Hessian matrix at the MLE parameters and use of standard asymptotic formulae. More specifically, by using a local approximation of the likelihood surface to estimate the covariance matrix, and by assuming errors are normally distributed, the covariance matrix can be used to compute a confidence interval around each parameter. We implement this method in chapter 5 to compare candidate models fitted to historic influenza pandemic waiting time data, and in chapter 6 to infer parameter values for a dynamic model of mood applied to adolescent school friendship networks.

2.4.2 Bayesian inference

If instead the required outputs are predictions of model parameter distributions, a Bayesian approach to parameter inference can provide a continuous estimation of this without specific assumptions and is therefore particularly suitable for fitting nonlinear epidemiological models to data [101–105]. Specifically, adopting a Bayesian framework means we treat the parameters, θ, as random variables that come from some probabilistic distribution given the observed data, D, known as the posterior distribution $f(\theta|D)$.

Central to the Bayesian inference concept is the use of Bayes' theorem to establish a relationship between the prior $\pi(\theta)$, the probability distribution that expresses one's beliefs about the parameters before taking into account any evidence, and posterior,

$$f(\theta|D) \propto L(D|\theta)\pi(\theta). \tag{2.3}$$

Hence, an estimate of the posterior distribution is attainable using only the prior $\pi(\theta)$ and likelihood function $L(D|\theta)$. The prior can incorporate previous findings and knowledge provided by biological experts, though if there is no informed prior belief an objective prior is typically used, corresponding to 'neutral' knowledge. When it is difficult to compute the posterior distribution analytically, meaning direct sampling is not possible, we resort to numerical estimation methods such as MCMC [106].

2.4.3 Markov chain Monte Carlo

MCMC methods are rooted in Bayesian statistics. Assuming the likelihood function for the particular problem is tractable, the objective of MCMC is to use equation (2.3) to construct a Markov chain with $f(\theta|D)$ as its stationary distribution. This chain can then be sampled from, with the sequence of random samples subsequently generated being used to estimate the target posterior distribution $f(\theta|D)$.

A number of algorithms have been developed to construct Markov chains that will eventually converge on the target posterior. One of the first was the Metropolis-Hastings algorithm [107, 108] outlined below (algorithm 1). It is still widely used at the time of writing and we make use of it ourselves in chapter 5. This method generates a random walk using a proposal distribution q and a method for rejecting some of the proposed moves. Specifically, in each iteration a move from the current parameter state θ to a new parameter state θ' is proposed, with the move accepted with probability

$$\alpha = \min\left\{1, \frac{\mathcal{L}(D|\theta')q(\theta|\theta')\pi(\theta')}{\mathcal{L}(D|\theta)q(\theta'|\theta)\pi(\theta)}\right\}.$$

The possibility of accepting the move even if there is a decrease in the likelihood of the proposed parameters given the data (relative to the previous parameter state) allows the chain to explore parameter space and is necessary to sample effectively from multi-modal distributions.

Algorithm 1 Single iteration of the Metropolis-Hastings algorithm

Input: θ_{t-1}, q, π, D
▷ Last state, proposal and prior distributions, observed data
Output: θ_t ▷ Next state
$\theta' \sim q(\theta'|\theta_{t-1})$ ▷ Propose new parameters
$\alpha = \min\left\{1, \frac{\mathcal{L}(D|\theta')q(\theta_{t-1}|\theta')\pi(\theta')}{\mathcal{L}(D|\theta_{t-1})q(\theta'|\theta_{t-1})\pi(\theta_{t-1})}\right\}$ ▷ Calculate the acceptance probability
$u \sim \text{Unif}[0, 1]$
if u $< \alpha$ **then** ▷ Accept θ' with probability given by α
    $\theta_t = \theta'$
else
    $\theta_t = \theta_{t-1}$
end if

The computational time for this inference procedure may be lengthened due to the following issues. First, initial samples may be very far from the target distribution, particularly if a poor choice of prior is used, meaning a number of the initial sam-

ples must be discarded (commonly referred to as a 'burn-in' period). Second, nearby samples will be highly correlated, so if there is slow decorrelation of the sampled chain it must be 'thinned' (generally by selecting every nth sample) to obtain (approximately) independent samples. Finally, if the step size in parameter space is too large there will be great inefficiency due to a high rejection rate of proposed samples.

Therefore, more efficient Markov chain methods have since been constructed that adapt to characteristics of the distribution being sampled. One such example is slice sampling [109] (algorithm 2). This is based on the principle that one can sample from a distribution by sampling uniformly from the region under the plot of its density function. In contrast to Metropolis-Hastings, slice sampling automatically adjusts the step size to match the local shape of the density function, enabling simpler and more efficient implementation. A second is the Adaptive Metropolis algorithm [110] (algorithm 3), in which a Gaussian proposal distribution is updated along the process using the full information cumulated so far. Adapting continuously to the target distribution enhances algorithm efficiency by diminishing the number of function evaluations needed. We utilise both algorithms in chapter 3 to fit a series of nested disease transmission models to two H5N1 epidemic outbreaks in the Dhaka region of Bangladesh.

Despite not being utilised in this thesis, another set of sophisticated MCMC methods to reduce the correlation between successive samples (and thus the number of samples required for the chain to converge) is to use a non-random walk MCMC method. One such algorithm is Hybrid Monte Carlo (also commonly known as Hamiltonian Monte Carlo) [111], which tries to avoid random walk behaviour by introducing an auxiliary momentum vector and implementing Hamiltonian dynamics, so the potential energy function is the target density. Consequently, proposals move across the sample space in larger steps, are therefore less correlated and converge to the target distribution more rapidly. However, HMC performance is highly sensitive to two user-specified parameters: a step size ϵ and a desired number of steps L. In particular, if L is too small then the algorithm exhibits undesirable random walk behaviour, while if L is too large the algorithm wastes computation. Recent research efforts to address these limitations include the No-U-Turn Sampler (NUTS) by Homan and Gelman [112]. This is an extension to HMC that eliminates the need to set a number of steps L, while the authors also derive a method for adapting the step size parameter on the fly.

Algorithm 2 Single iteration of the Slice Sampling algorithm

Input: $\theta_{t-1}, D, \pi, w, \text{dim}$
▷ Last state, prior distribution, observed data, step-size parameter, parameter no.
Output: θ_t ▷ Next state
$u \sim \text{Unif}[0, \mathcal{L}(D|\theta_{t-1})\pi(\theta_{t-1})]$ ▷ Set the threshold for the next state
$u' \sim \text{Unif}[0, 1]$
$\theta_{\min} = \theta_{t-1} - u'w$ ▷ Initialise the upper and lower bounds of the slice
$\theta_{\max} = \theta_{\min} + w$
while $u < \mathcal{L}(D|\theta_{\min})\pi(\theta_{\min})$ **do**
    for $i = 1$ to dim **do**
        $\theta_{\min}(i) = \theta_{\min}(i) - w(i)$
    end for
end while
while $u < \mathcal{L}(D|\theta_{\max})\pi(\theta_{\max})$ **do**
    for $i = 1$ to dim **do**
        $\theta_{\max}(i) = \theta_{\max}(i) + w(i)$
    end for
end while
repeat
    for $i = 1$ to dim **do**
        $u' \sim \text{Unif}[0, 1]$
        $\theta'(i) = \theta_{min}(i) + u'(\theta_{max}(i) - \theta_{min}(i))$ ▷ Draw a new state uniformly
    end for
    if $u < \mathcal{L}(D|\theta')\pi(\theta')$ **then** ▷ Check for the threshold
        $\theta_t = \theta'$
        break ▷ Return θ' as next state if $u < \mathcal{L}(D|\theta')\pi(\theta')$
    end if
    for $i = 1$ to dim **do**
        if $\theta'(i) < \theta_{t-1}(i)$ **then** ▷ Adjust either the lower or upper bound
            $\theta_{\min}(i) = \theta'(i)$
        else
            $\theta_{\max}(i) = \theta'(i)$
        end if
    end for
until θ' is accepted

Algorithm 3 Adaptive Metropolis algorithm

Input: π, dim, N, D ▷ Prior distribution, parameter no., required sample number, , observed data
Output: $\mathcal{S}$ ▷ Accepted samples
$\mathcal{S} = \emptyset$
$\theta_0 \sim \pi(\theta)$ ▷ Set θ_0 from the prior
$\sigma = \frac{2.4^2}{dim}$ ▷ Compute covariance matrix scale factor
for $t = 1 : N$ **do**
    $\theta' \sim \mathcal{N}(\theta_{t-1}, \sigma\Sigma_{t-1})$ ▷ Propose new parameters
    $\alpha = \min\left\{1, \frac{\mathcal{L}(D|\theta')\pi(\theta')}{\mathcal{L}(D|\theta_{t-1})\pi(\theta_{t-1})}\right\}$ ▷ Calculate the acceptance probability
    $u \sim \text{Unif}[0, 1]$
    if u $< \alpha$ **then** ▷ Accept θ' with probability given by α
        $\theta_t = \theta'$
    else
        $\theta_t = \theta_{t-1}$
    end if
    $S = S \cup \{\theta_t\}$ ▷ Add updated state to sample set
    $\mu_t = \mu_{t-1} + \frac{1}{t}(\theta_t - \mu_{t-1})$ ▷ Update mean and covariance matrix
    $\Sigma_t = \Sigma_{t-1} + \frac{1}{t}\{(\theta_t - \mu_{t-1})(\theta_t - \mu_{t-1})^T - \Sigma_{t-1}\}$
end for

Standard MCMC methodology can be extended to allow simulation of the posterior distribution on spaces of varying dimensions, specifically using reversible jump Markov chain Monte Carlo (RJMCMC) methods [113]. Thus, simulation is possible even if the number of parameters in the model is not known, and model selection can be performed. Neal and Roberts [102] used this technique to carry out statistical inference and model selection for the 1861 Hagelloch measles epidemic by comparing a 'full' model to a set of nested models, with each nested model being the same as the full model but with one parameter removed. Further details of this technique are given within chapter 5, where we apply this methodology to compare candidate models fitted to historic influenza pandemic waiting time data.

2.5 Simulations for model verification and intervention analysis

Having constructed a model framework and then fitted it to data via parameter inference, we must make use of a verification tool to discern whether the model can capably produce outputs that have reasonable correspondence with the observed data (stage four of the modelling cycle as given in figure 1.1).

To achieve this we use stochastic simulations, which allow us to approximate the chance element in the spreading processes of interest, with multiple simulations determining the expected range of behaviour. In particular, a prevalent theme within this thesis is the simulation of stochastic models (particularly individual-based models) using event-driven approaches, which capture the fluctuations in population processes that arise because of the random nature of events at the level of the individual. The increased availability of computer power has led to modelling approaches that incorporate this demographic stochasticity becoming increasing popular. This is largely due to their highly mechanistic approach in the individual nature of their formulation and including randomness at the level of the individual [80].

If the models are considered satisfactory, a second simulation application is to make predictions and assess impact of proposed control initiatives (stage five of the modelling cycle as given in figure 1.1). This complements previous extensive uses of computer simulations as a key tool in assessing the epidemiology and possible control strategy options for a number of specific diseases, including foot-and-mouth disease [83, 114] and pertussis [115].

Realisations of stochastic models can be generated by computer using standard Monte-Carlo techniques [69, 116, 117]. We now briefly discuss the development of algorithms for carrying out stochastic simulations. An efficient and accurate implementation of the event-driven framework, that is routinely used, is the Direct Gillespie Algorithm [118, 119] (see algorithm 4 for pseudocode). The scheme first estimates the time until the next event, based on the cumulative rates of all possible events. Then, by converting event rates into probabilities, it randomly selects one of these events. The times and host statuses are then updated according to which event is chosen. Repetitions of this process iterate the model through time. To clarify, randomness only affects the probabilities associated with the epidemiological unit of interest, with no assumption made concerning environmental stochasticity [72]. We use a scheme with similarities to the Gillespie algorithm in chapter 5.

A selection of models introduced in the following chapters are individual-based (chapters 3, 4 and 6), with a defined epidemiological unit of interest (such as an individual or premises), so we may account for spatial interactions and heterogeneity in contacts. A serious drawback of the Gillespie algorithm is that as population size increases, the number of interaction terms that must be considered becomes

Algorithm 4 Direct Gillespie algorithm

1: Specify model run end-time $t_{\max}$.
2: Initialise model with initial conditions, time $t = t_0$, state $x(t_0)$.
3: Label all possible events $E_1, \ldots, E_n$.
4: **while** $t < t_{\max}$ **do**
5: Compute rates at which events occur, $R_1(x(t)), \ldots, R_n(x(t))$.
6: $R_{total}(t) = \sum_{m=1}^{n} R_m(x(t))$ ▷ Rate at which any event occurs
7: $\delta t \sim \exp(R_{total}(t))$ ▷ Time to next event
8: $RAND_1 \sim \text{Unif}[0, 1]$
9: $P = RAND_1 \times R_{total}(t)$
10: Event p occurs if $\sum_{m=1}^{p-1} R_m(x(t)) < P \leq \sum_{m=1}^{p} R_m(x(t))$.
11: Update state vector $x(t)$.
12: $t = t + \delta t$ ▷ Update time
13: **end while**

prohibitively large. Additionally, the time interval between events decreases, which in turn increases the number of iterations that must be performed to advance the model by a given time period. For individual-based models, a solution to counteract this and obtain gains in simulation speed is to discretise time, allowing multiple events to occur each time-step. An algorithm to improve computational efficiency in such as a way was proposed by Gillespie [120], known as the τ-leap method. This results in a great improvement in speed, and is justifiable as long as the chance of an epidemiological unit undergoing multiple events in a short time-step is small (meaning all events occurring in one time-step are likely to be independent). For discrete-time individual-based models, the τ-leap procedure requires minor modifications as it is necessary to convert the event rates into probabilities, as outlined in algorithm 5. This discrete-time construct is repeatedly employed throughout this work.

As an alternative choice to the previously mentioned algorithms, the final simulation technique employed within this thesis makes use of the Sellke construction, first introduced as an analytic tool to investigate the distribution of final sizes of a stochastic epidemic [121]. As well as being a probabilistically equivalent to the Gillespie algorithm, a desirable characteristic of this framework is that the inherent randomness of a epidemic realisation can be encoded at the beginning of the simulation, in our case a random vector Z of Exp(1) distributed resistances. Once calculated, the resultant epidemic can be constructed from the deterministic solution of the infection process and removal (i.e. culling) times. Therefore, this method

Algorithm 5 τ-leap algorithm for individual based models

Choose time-step τ
Specify model run end-time $t_{\max}$.
Initialise model with initial conditions, time $t = t_0$, state $x(t_0)$.
Label all possible events $E_1, \ldots, E_n$. ▷ One event for each epidemiological unit
while $t < t_{\max}$ **do**
    Compute rates at which events occur, $R_1(x(t)), \ldots, R_n(x(t))$.
    for $j = 1 : n$ **do**
        $\mathbb{P}(E_j) = 1 - \exp(-\text{Rate}_j \tau)$. ▷ Prob. of event j in time interval $[t, t + \tau]$
        $RAND_1 \sim \text{Unif}[0, 1]$
        Check if $RAND_1 < \mathbb{P}(E_j)$.
    end for
    Update state vector x.
    $t = t + \tau$ ▷ Update time
end while

may be utilised to provide improved comparisons of interventions, with direct comparison of a collection of control measures achieved by matching values of Z at the epidemic outset. Such a scheme is used in chapter 4 for these reasons, with an example application of the procedure for the SIR model (with no demography) shown in algorithm 6.

Algorithm 6 Sellke Construction Algorithm for the SIR model

Initialise model with initial conditions, time $t = t_0$, state $x(t_0)$.
For each initially susceptible unit generate the epidemic resistances $Z_i \sim \text{Exp}(1)$.
For each unit i generate the infectiousness duration $D_i \sim \text{Exp}(\gamma)$.
For each initially infectious unit i store the recovery time $T_i^R = D_i$.
while Infected units present **do**
    For each susceptible unit i calculate the force of infection against it, $\lambda_i(t)$.
    For each susceptible unit i compute possible time until infection, $\delta t_i^I = \frac{Z_i}{\lambda_i(t)}$.
    Construct list of times until next event, $\{\{\delta t_i^I\}_{i:S_i(t)=1}, \{T_i^R - t\}_{i:I_i(t)=1}\}$.
    Set $\delta t^* = \min\{\{\delta t_i^I\}_{i:S_i(t)=1}, \{T_i^R - t\}_{i:I_i(t)=1}\}$.
    Let E denote the event corresponding to δt^*.
    Deplete resistances for remaining susceptible populations, $Z_i = Z_i - \lambda_i(t)\delta t^*$.
    If E is an infection event, generate the recovery time $T_i^R = t + \delta t^* + D_i$.
    Update time, $t = t + \delta t^*$.
    Update state vector $x(t)$ by carrying out event E.
end while

Chapter 3

Modelling H5N1 in Bangladesh across spatial scales

3.1 Introduction

The H5N1 subtype of highly pathogenic avian influenza (HPAI) has caused considerable concern since the initial observation of the virus in southern China during 1996 [20]. From the time of the first large-scale epizootic that took place in the winter of 2003/2004 in East and Southeast Asia [122], H5N1 has killed or forced the culling of more than 400 million domestic poultry and resulted in an estimated US$20 billion in economic damage, with 63 countries infected at its peak in 2006 [21]. Being a zoonotic disease H5N1 HPAI remains a persistent public health threat, capable of causing infection in humans with a high mortality rate. Since 2003 it has caused over 850 laboratory-confirmed human cases across 16 countries, leading to subsequent deaths in 14 of these nations, with the cumulative death total exceeding 450 [23].

With a number of countries in South and Southeast Asia, including China, Vietnam and Bangladesh, being gravely affected, a number of studies have predominately focused on either spatio-temporal analysis of outbreaks [29, 31, 33, 123, 124], or on determining ecological/environmental risk factors for H5N1 avian influenza emergence and spread at region-wide [35], national [30, 34, 38] and sub-national levels [36]. For example, H5N1 poultry epidemics in Thailand have been associated with the following risk factors: rice crop intensity, free grazing ducks and water presence [34, 38]. Across studies and regions three types of variables with similar statistical associa-

tion with H5N1 were identified: domestic waterfowl, human related variables (e.g. human population density) and indicators of water presence [39].

Bangladesh is one of the most densely populated countries in the world, with a human population exceeding 160 million [125]. In combination with an intensifying farming system and substantial poultry population (1194 birds/km^2) [126], these conditions make Bangladesh a prime candidate for being the source of newly emerging influenza strains with pandemic causing potential. Therefore, it is vital to enhance our understanding of the factors in Bangladesh that enable currently circulating influenza subtypes (e.g. H5N1) to be both continually transmitted between poultry and occasionally spillover across the human-animal interface. Bangladesh specific risk analyses have determined a number of biosecurity related risk factors associated with H5N1 infection in commercial poultry [127, 128], while identifying free grazing duck and duck-rice cultivation interacted ecology as not being significant determinants [129, 130]. Risk factors specific to backyard chickens have also been investigated [131]. Osmani *et al.* [132] found the spread of H5N1 in Bangladesh to be characterised by reported long-distance translocation events, with the relative contribution of trade and the market chain versus wild birds in spreading the disease still to be resolved. Loth *et al.* [30] investigated temporal and spatial patterns of H5N1 poultry outbreaks in Bangladesh, occurring between March 2007 and July 2009, and their relationship with several spatial risk factors at a sub-district level. Human population density, commercial poultry population density and number of roads per sub-district were found to be significantly associated with H5N1 virus outbreaks. However, they emphasise that research on the roles of wildlife, migratory birds and ducks in the epidemiology of H5N1 in Bangladesh is urgently needed. How the risk of H5N1 infection varies at different spatial resolutions must also be determined, from local administrative units (in Bangladesh referred to as districts), to province-level (referred to as divisions), up to the country-level. This work focuses on the district-level and division-level.

To date, very few zoonotic disease dynamic models incorporate zoonotic transmission from the animal reservoir to humans [85]. In particular, mathematical modelling of H5N1 thus far has generally only quantified poultry transmission parameters [40, 41]. Whilst recent seroprevalence and seroconversion studies have been undertaken in poultry workers in Thailand [133] and Bangladesh [134], the work outlined has predominately considered H5N1 infection in livestock only. Devising a new generation of approaches to model cross-species spillover transmission is one of

the several challenges related to modelling the emergence of novel pathogens that requires attention [88].

The purpose of this study is to outline a modelling framework that incorporates zoonotic transmission at the human-poultry interface, in addition to within-poultry disease dynamics. The model will be utilised to ascertain whether the size and spatial distribution of commercial poultry H5N1 cases in specified regions of Bangladesh can be predicted accurately at different administration levels and, if so, the crucial modelling considerations that are necessary for this to be achieved. Furthermore, we analyse whether the main contributor to the spillover of H5N1 influenza from poultry to humans in Bangladesh, between H5N1 prevalence in the commercial poultry population or other factors (such as interactions at live bird markets (LBMs)), can be distinguished. The findings that arise motivate further studies examining the effectiveness of intervention measures aiming to minimise the risk of zoonotic transmission of H5N1 influenza, which we explore in more depth in chapter 4.

3.2 Methods

3.2.1 The data

The data utilised were comprised of four main components: (i) a commercial poultry premises census, (ii) poultry case data, (iii) external risk factors (live bird markets, free-grazing ducks, presence of water, rice cropping), (iv) human case data.

Commercial poultry premises census

In 2010, the Bangladesh office of the Food and Agriculture Organisation of the United Nations (FAO/UN) undertook a census of all commercial poultry premises, listing 65,451 premises in total, of which 2,187 were LBMs. Each premises was visited once, with the premises location recorded along with the number of the following types of avian livestock present during the visit: layer chickens, broiler chickens, ducks, others (e.g. turkeys, quails). Within the census data there were instances of multiple premises having the same location (i.e. identical latitude and longitude co-ordinates). For these occurrences the avian livestock populations were amalgamated, giving a single population for each category at each location.

Of the non-market locations, 23,412 premises had blank entries for all avian types. It has been confirmed this did correspond to no poultry being present on these premises

when the census visit occurred, due to the premises either being between poultry stocks or being temporary closed by the farmer due to an ownership transfer taking place, rather than data entry errors (M.G. Osmani, personal communication). We made a simplifying assumption that at any given time an equivalent proportion of premises would not have any avian livestock at the premises. Therefore, we did not make use of these locations in our analysis. While not discussed here the sensitivity of model outputs to this assumption requires further consideration.

Poultry case data

From 2007 to 2012 inclusive there have been 554 poultry premises with reported H5N1 infection in Bangladesh. These were predominately commercial premises (497 cases), with 57 cases reported from backyard flocks. The Bangladesh office of FAO/UN provided a dataset of confirmed infected premises up to June 2011. Cases occurring after June 2011 were obtained from the OIE World Animal Health Information Database (WAHID) Interface [135]. For the case data provided by the latter source we were informed that the Department of Livestock Services reported regularly to WAHID regarding HPAI outbreaks in Bangladesh, with this usually occurring within 24 hours according to the code of the World Organisation for Animal Health (M.G. Osmani, personal communication). We therefore presumed WAHID contained all reported Bangladesh HPAI event information.

For each infected premises the data documented its spatial location, the date that infection was reported, the date of culling, and the total number of poultry infected and culled. We divided the infected premises data into distinct epidemic waves. These were estimated by looking for significant gaps between premises infection dates, with a gap of two months or more used to signify the end of one wave and the start of a new one. The dates and number of cases for each wave are displayed in table 3.1.

There were 52 poultry premises recorded as being infected that were not part of the 2010 premises census. When analysing a specific wave all additional entries that occurred during that wave were considered, including the reported backyard farm cases when applicable. In addition, for premises infected during a specific wave we modified the poultry populations to match the flock sizes reported in the poultry case dataset (rather than using the reported values from the 2010 census).

Table 3.1: Breakdown of H5N1 HPAI poultry epidemic waves in Bangladesh. Start month, end month and number of reported infected premises in each of the H5N1 poultry epidemic waves in Bangladesh.

	Start month	**End month**	**Reported cases**
Wave 1	March 2007	July 2007	55
Wave 2	September 2007	May 2008	232
Wave 3	November 2008	June 2009	37
Wave 4	January 2010	June 2010	31
Wave 5	January 2011	May 2011	161
Wave 6	November 2011	April 2012	26

In addition to the cases within each wave listed above, the following reported cases occurred between waves: one case in September 2008; two cases in August 2009; one case in June 2011; four cases in August-September 2011; four cases in October 2012 - March 2013.

External risk factors

In addition to LBMs [134], presence of free grazing ducks, water and rice paddy fields have been determined as H5N1 avian influenza risk factors for poultry in other areas of south-east Asia [34, 38]. Thus, to investigate the importance of these ecological covariates our most complex models included information on these factors. Duck density at a 1km resolution was obtained from the Gridded Livestock of the World (GLW 2.0) dataset [136]. Presence of water bodies was determined from global land cover maps produced by GlobCover at a 300m spatial resolution [137]. Rice paddy agriculture and cropping intensity in Asia can be routinely mapped and monitored using images from the moderate resolution imaging spectroradiometer (MODIS) sensor onboard the NASA Terra satellite [138–140]. The satellite-based algorithms permit the production of maps and monitoring of cropping intensity and the crop calendar (planting and harvesting dates). This source provided rice paddy coverage in Bangladesh at a 500m spatial resolution for the years 2008 and 2011.

Human case data

There have been eight reported human cases of H5N1 infection, causing one death [23]. Latitude and longitude co-ordinates for these cases were obtained using the FAOs Global Animal Disease Information System (EMPRES-i) database [141]. Seven of the eight human cases occurred within the poultry epidemic waves outlined above. Of these, six were located in the Dhaka division and five within the Dhaka district, with the following distribution of cases across the poultry epidemic waves: one in wave 2, two in wave 5, three in wave 6 (the timing of these human cases in relation

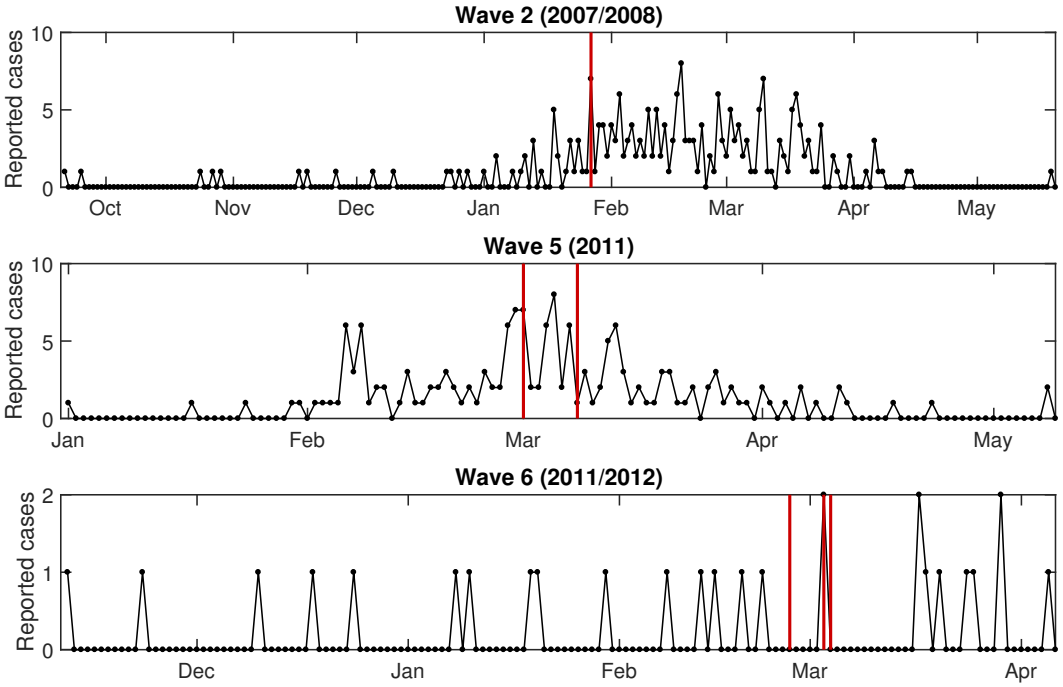

Figure 3.1: Epidemiological curve of reported H5N1 cases in Bangladesh. Black curves correspond to counts of newly reported infected poultry premises per day in Bangladesh during epidemic waves 2, 5 and 6 respectively. Vertical red bars correspond to a day where clinical signs in a confirmed human case (within the Dhaka division) were first observed. The distribution of human cases, within the Dhaka division, across the poultry epidemic waves was as follows: one in wave 2, two in wave 5, three in wave 6.

to the number of premises reporting infection is shown in figure 3.1).

3.2.2 Poultry model

Selection of spatial scales and epidemic waves

With the majority of human cases being located within the Dhaka district (area: 1,464 km^2) and Dhaka division (in 2010, total area 41,761.8 km^2), our model was focused on these two differing administration (spatial) levels. Further, Dhaka district was of notable interest due to being only one of two districts (out of 18 districts in the Dhaka division) that reported presence of H5N1 infection in all six epidemic waves. For applying our poultry model framework (performing parameter inference) we focused on the epidemic waves containing both human cases and over 100 premises reporting H5N1 infection in poultry. These were wave 2 (September 2007–May 2008) and wave 5 (January 2011–May 2011), with reported case epidemiological

curves for these waves presented in figure 3.1.

Specifically for the Dhaka district our analysis of waves 2 and 5 considered 1,271 and 1,270 premises, with H5N1 infection in poultry confirmed at 22 and 25 premises respectively. In particular, four out of the six sub-districts comprising the Dhaka district had presence of infection in both waves (figure 3.2).

For the Dhaka division, our wave 2 dataset contained 13,369 premises while the wave 5 dataset contained 13,359 premises. There were 109 and 75 reported cases in waves 2 and 5 respectively, with specific sub-districts having a notably higher proportion of total infection (see figure 3.3). Overall, 18 sub-districts (out of 113 contained within the Dhaka division) had confirmed cases during both waves of interest, with 41 and 25 individual sub-districts having infection present during waves 2 and 5 respectively. Note that owing to the small number of premises recorded as having ducks or other poultry types present, with only two such premises in the Dhaka district and roughly 20 premises in the Dhaka division, we focused on layer and broiler chickens in our poultry models.

The candidate models described below were fitted at the district and division administration levels. However, it is possible that a poultry epidemic had begun in Bangladesh outside the specified region, and/or continued in another region of Bangladesh after the final case was culled in the specified region. To address this, for each spatial level and wave of interest we considered two different sets of dates. The first was a region-specific epidemic time period. This began on the day poultry cases initially occurred in the region of interest, ending on the day the final infected premises was culled. If required, a second epidemic wave time period took the country-wide dates for that epidemic wave. First reporting and final culling dates for each combination of administration level and wave are provided in table 3.2. Finally, for each reported premises the time delay between notification and culling was recorded in the data.

Model structure

We formulated our candidate models as discrete-time compartmental models. At any given point in time a premises i could be in one of four states, S, I, Rep or C: $i \in S$ implies premises i was susceptible to the disease; $i \in I$ implies premises i was infectious and not yet reported; $i \in Rep$ implies premises i was still infectious, but had been reported; $i \in C$ implies that premises i had been culled.

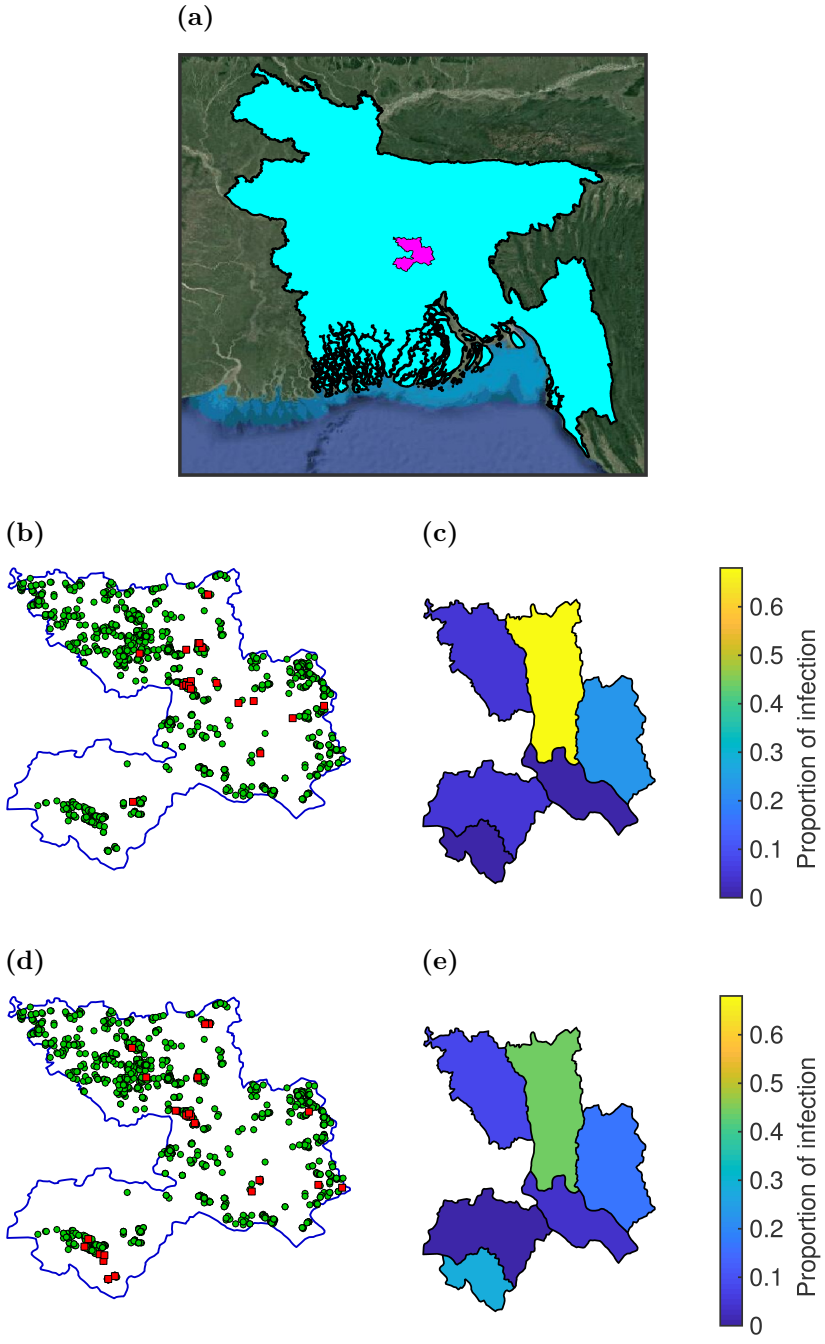

Figure 3.2: Spatial locations of premises infected during the wave 2 and wave 5 poultry epidemic waves, located within the Dhaka district. **(a)** Locator map depicting the location of Dhaka district, shaded in magenta, within Bangladesh, shaded in cyan. **(b-e)** The left column shows infection status of each premises, with red squares depicting premises that were infected and green circles those that remained susceptible. The right column shows the proportion of infection aggregated at a sub-district level. **(b, c)** Wave 2 district; **(d, e)** wave 5 district.

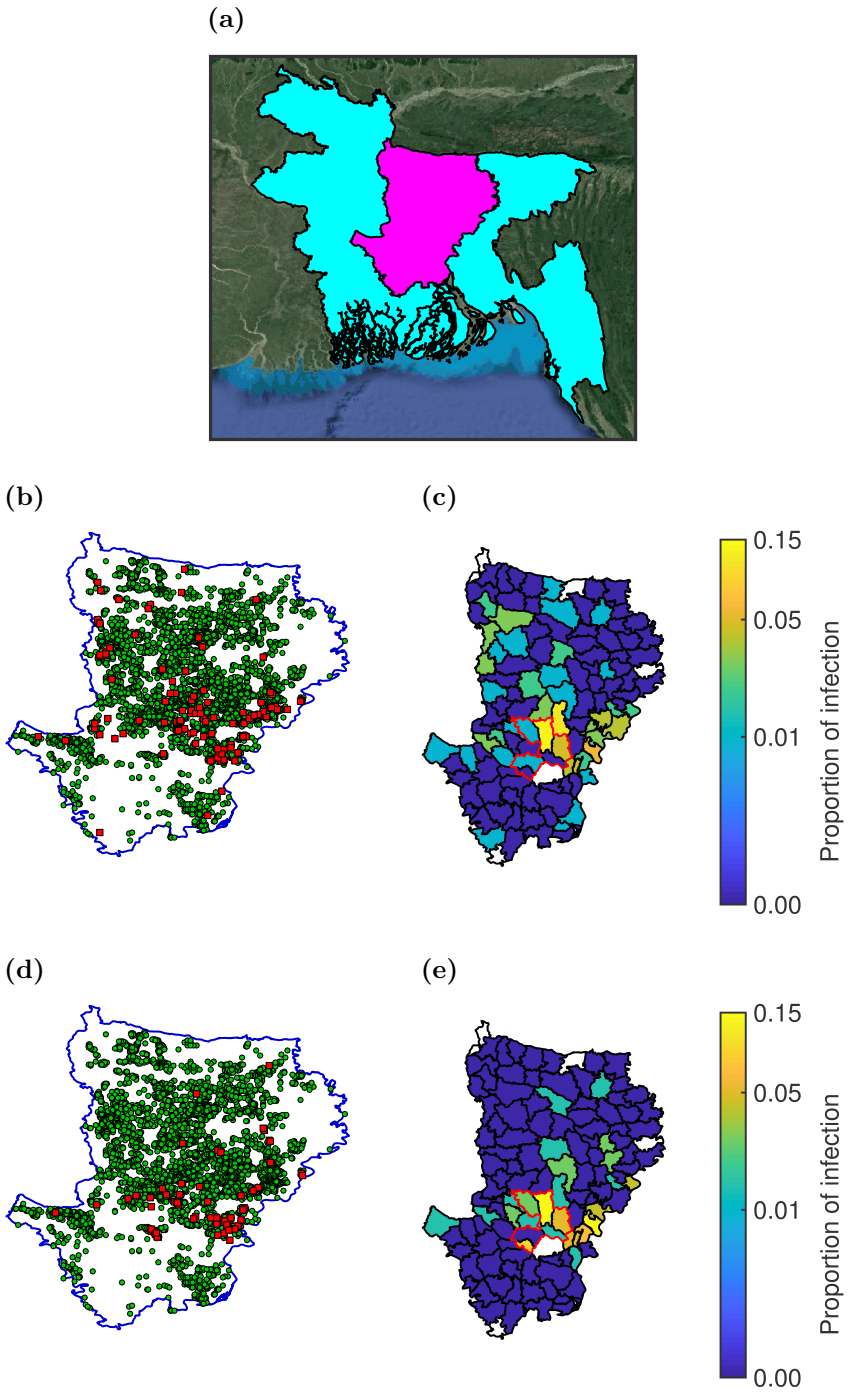

Figure 3.3: Spatial locations of premises infected during the wave 2 and wave 5 poultry epidemic waves, located within the Dhaka division. **(a)** Locator map depicting the location of Dhaka division, shaded in magenta, within Bangladesh, shaded in cyan. **(b-e)** The left column shows infection status of each premises, with red squares depicting premises that were infected and green circles those that remained susceptible. The right column shows the proportion of infection aggregated at a sub-district level. In **(c, e)** Dhaka district is outlined in red to highlight its location within the Dhaka division. **(b, c)** Wave 2 division; **(d, e)** wave 5 division.

Table 3.2: List of region specific and country-wide epidemic wave dates. We use — to denote country-wide epidemic dates that were not required as they matched the region-specific epidemic dates for that particular combination of administration level and wave.

	Wave	Epidemic date type	First report date	Final cull date
District	**2**	Region specific	28 Dec 2007	31 Mar 2008
		Country-wide	21 Sep 2007	19 May 2008
	5	Region specific	1 Jan 2011	15 Mar 2011
		Country-wide	1 Jan 2011	9 May 2011
Division	**2**	Region specific	26 Nov 2007	19 May 2008
		Country-wide	21 Sep 2007	19 May 2008
	5	Region specific	1 Jan 2011	9 May 2011
		Country-wide	—	—

We considered an overall poultry population at each premises (i.e. layer and broiler chickens were not treated as distinct poultry types). This is based on a conceptualisation where the individual poultry premises is the epidemiological unit of interest. In other words, all poultry types within a premises become rapidly infected such that the entire premises can be classified as Susceptible (S), Infected (I), Reported (Rep) or Culled (C). We define a premises i in one of these four states at time t as being in the sets $S(t)$, $I(t)$, $Rep(t)$ or $C(t)$ respectively. While the poultry epidemic was ongoing we assumed a premises was not repopulated once culled.

Notification delays

Our modelling framework incorporated a reporting delay to take into account a premises being infectious before clinical signs of H5N1 infection are observed, which may not be immediate [142], followed by the time taken for premises owners to notify the relevant authorities [143]. We treated the delay time as an integer and found that the distributions of other model parameters were quite sensitive to it. This made it natural to treat different plausible values for the delay as different models, and to select between them (as was done, for example, in the context of selection from discrete outbreak source locations by Hancock *et al.* [144]). We chose three fixed infection notification times of two, four and seven days, corresponding to the 50%, 75% and 90% percentiles of the reporting delay distribution for 2009 H5N1 HPAI reports of domestic poultry infection [145]. We systematically compared model fit and predictions under these different values.

Force of infection

The force of infection towards a susceptible premises could be dependent on a variety of factors. Therefore, we proposed a series of nested models of increasing complexity. The base model used is an adaptation of a foot-and-mouth disease model developed by Keeling *et al.* [76]. The rate at which an infectious premises j infects a susceptible premises i is given by

$$\eta_{ij} = s_c N_{c,i} \times t_c N_{c,j} \times K(d_{ij}), \tag{3.1}$$

where $N_{c,i}$ is the total number of chickens recorded as being on premises i; s_c and t_c measure the individual chicken susceptibility and transmissibility; d_{ij} is the distance between premises i and j in kilometres; and K is the transmission kernel to capture how the relative likelihood of infection varies with distance.

We extended the model by including a 'spark' term parameter ϵ_i to allow for spontaneous, non-distance dependent infections that were unexplained by the susceptibility, transmissibility and kernel components of the model [103]. In combination with the distance-dependent transmission kernel, K, this allows our model framework to capture premises contacts that are both dependent on, and independent of, distance. In the absence of empirical poultry movement data, such an approach (including separate distance-dependent and distance-independent terms) has been found to be preferred for modelling animal movement contact data compared to a solely distance dependent process [146]. Further, despite the absence of explicit data on backyard poultry its contribution to the force of infection could be incorporated into ϵ_i.

Overall, the force of infection against a susceptible premises i on day t ($\text{Rate}_M(i,t)$ for model label M) was comprised of two terms: (i) the force of infection generated by an infectious premises j ($\eta_{ij,M}$), (ii) the spark term ($\epsilon_{i,M}$). As a result, the total force of infection has the following general form for model M:

$$\text{Rate}_M(i,t) = \left(\sum_{j \in I(t) \cup Rep(t)} \eta_{ij,M} \right) + \epsilon_{i,M}.$$

We now outline the key constituents of our proposed nested models for the force of infection, labelled A to E, with the models building upon one another through the inclusion of additional parameters.

Baseline model (A)

For the baseline model (model A) the infected premises contribution to the force of infection matched the Keeling *et al.* [76] model (equation (3.1)),

$$\eta_{ij,A} = s_c N_{c,i} \times t_c N_{c,j} \times K(d_{ij}).$$

In this case K was derived from the Dhaka division poultry case data. For each infected premises we found the nearest premises that was infected within the previous two days before that farm was reported. The distance between the infecting premises and the newly infected premises was calculated, with the process repeated for each infected premises. Kernel density estimation (KDE) was applied, via the Matlab function `kde()`, to the distances obtained from this process. This approximated a smooth functional form for K (figure A.1).

The spark term was the same fixed value for every premises, $\epsilon_{i,A} = \epsilon$. Subsequently, the total rate of infection against a susceptible premises i on day t satisfied

$$\text{Rate}_A(i,t) = \left(\sum_{j \in I(t) \cup Rep(t)} \eta_{ij,A} \right) + \epsilon. \tag{3.2}$$

To make the model identifiable we set $s_c = 1$. This is carried forward in all subsequent models. Note that although this changes the interpretation of the parameters, it does not have any epidemiological implications.

This left two parameters in equation (3.2) requiring estimation, t_c and ϵ.

Parametric kernel model (B)

For the parametric kernel model (model B) we fit a parametric transmission kernel in place of the kernel derived from the poultry case data. Our chosen transmission kernel was pareto distributed:

$$K(d_{ij}) = \begin{cases} 1 & \text{if } 0 \leq d_{ij} < x_{\min}, \\ \left(\frac{x_{\min}}{d_{ij}}\right)^{\alpha+1} & \text{if } x_{\min} \leq d_{ij}, \\ 0 & \text{otherwise.} \end{cases} \tag{3.3}$$

$x_{\min}$ is the minimum possible value of the function (in our case set to 0.1, corresponding to 100 metres, with all between location distances less than 100 metres taking the 100 metre kernel value) and $\alpha \geq -1$. This kernel form could provide

insights into how transmission risk varied with respect to the distance between the infected premises and target susceptible location. Values of α close to -1 would give a relatively constant kernel over all distances, with $\alpha = -1$ corresponding to transmission risk being independent of distance. As α increases away from -1 localised transmission is favoured, with long-range transmission diminished.

The infected premises contribution to the force of infection $\eta_{ij,B}$ has the same form as $\eta_{ij,A}$,

$$\eta_{ij,B} = s_c N_{c,i} \times t_c N_{c,j} \times K(d_{ij}),$$

but with the kernel described above used in place of the data derived kernel.

Thus, for model B the total rate of infection against a susceptible premises i on day t satisfies

$$\text{Rate}_B(i,t) = \left(\sum_{j \in I(t) \cup Rep(t)} \eta_{ij,B} \right) + \epsilon.$$

With this set up the following three parameters were fitted: t_c, α, ϵ.

Nonlinear farm size model (C)

Previous modelling work on foot-and-mouth disease in the UK suggests including parameters that account for a non-linear increase in susceptibility and transmissibility as animal numbers on a premises increase provide a closer fit to historical epidemic data than when these powers are set to unity [103, 147]. We explored whether this behaviour applied to H5N1 avian influenza by adding power law exponents to the susceptible population, p_c, and infected population, q_c.

The updated infected premises contribution to the force of infection was,

$$\eta_{ij,C} = N_{c,i}^{p_c} \times t_c N_{c,j}^{q_c} \times K(d_{ij}).$$

Note that the transmission kernel K used here was the pareto distributed form as described in model B, the parametric kernel model (equation (3.3)). As before, the spark term was the same fixed value for every premises, $\epsilon_{i,C} = \epsilon$. Therefore, the total rate of infection against a susceptible premises i on day t satisfies

$$\text{Rate}_C(i,t) = \left(\sum_{j \in I(t) \cup Rep(t)\, t} \eta_{ij,C} \right) + \epsilon,$$

with five parameters to be estimated $(t_c, \alpha, \epsilon, p_c, q_c)$.

Full ecological model (D)

For each spatial level and epidemic wave, the preferred model out of models A-C was ascertained by comparing deviance information criterion (DIC) values (see model comparison methodology). The preferred model was carried forward with the single premises-independent spark term ϵ replaced by four biologically motivated spark term covariates to form model D. In detail, these were the presence or absence of the following in the neighbourhood of the given premises, with the resolution used for each covariate stated in brackets: (i) water bodies (300m grid); (ii) paddy fields (500m grid); (iii) LBMs (within 5km); (iv) local ducks (1km grid). The rice paddy data was taken from the same year the epidemic wave of interest took place (i.e. from 2008 if considering a wave 2 model and from 2011 if considering a wave 5 model).

Mathematically, the spark term was now premises-dependent,

$$\epsilon_{i,D} = \epsilon_{\text{water}}\text{water}_i + \epsilon_{\text{rice}}\text{rice}_i + \epsilon_{\text{LBM}}\text{LBM}_i + \epsilon_{\text{ducks}}\text{ducks}_i,$$

where water_i is an indicator function for the presence of water in the neighbourhood of premises i (1 if present, 0 if absent), with similar definitions for the remaining factors.

For model D, we obtained the following expression for the total rate of infection against a susceptible premises i on day t,

$$\text{Rate}_D(i,t) = \left(\sum_{j \in I(t) \cup Rep(t)\, t} \eta_{ij,m} \right) + \epsilon_{i,D},$$

where $m \in \{A, B, C\}$ corresponds to the use of one of model A, B or C as required.

Simple ecological model (E)

Despite a number of studies identifying domestic waterfowl and rice crop intensity having a strong association with HPAI H5N1 presence [34, 38, 39], previous work has determined these factors as not being significant within Bangladesh [129, 130]. Due to this, we considered a simplified ecological model that contained only the presence or absence of water bodies and LBMs (in the same manner outlined above for the full ecological model) in the neighbourhood of the given premises as ecological spark

term covariates.

Once more, for each spatial level and epidemic wave we carried forward the preferred model out of models A-C and added the additional spark term covariates to form model E. In this case, we defined

$$\epsilon_{i,E} = \epsilon_{\text{water}}\text{water}_i + \epsilon_{\text{LBM}}\text{LBM}_i.$$

Consequently, the total rate of infection against a susceptible premises i on day t for model E was

$$\text{Rate}_E(i,t) = \left(\sum_{j \in I(t) \cup Rep(t)\, t} \eta_{ij,m} \right) + \epsilon_{i,E},$$

where $m \in \{A, B, C\}$ corresponds to the use of one of model A, B or C as required.

3.2.3 Likelihood function

Parameter estimation was carried out within a Bayesian framework. For constructing the likelihood function we follow the description provided by Deardon *et al.* [103]. Given a record of infection events at discrete time points $t = 0, \ldots, T$ during the epidemic, the likelihood is a product over those time points. We define $p(i,t)$ as the probability of premises i becoming infected on day t. In particular,

$$p(i,t) = 1 - e^{-\text{Rate}(i,t)\hat{t}}$$

where $\hat{t} = 1$. Note the form of $\text{Rate}(i,t)$ is model dependent. The general form for the likelihood of our models is given by

$$L(\mathbf{S}, \mathbf{I}, \mathbf{Rep}, \mathbf{C}|\theta) = \prod_{t=0}^{T} f_t(\mathbf{S}, \mathbf{I}, \mathbf{Rep}, \mathbf{C}|\theta)$$

where

$$f_t(\mathbf{S}, \mathbf{I}, \mathbf{Rep}, \mathbf{C}|\theta) = \prod_{i \in I(t+1) \setminus I(t)} p(i,t) \prod_{i \in S(t+1)} (1 - p(i,t)),$$

$\mathbf{S} = \{S(t)\}_{t=0}^{T}$, $\mathbf{I} = \{I(t)\}_{t=0}^{T}$, $\mathbf{Rep} = \{Rep(t)\}_{t=0}^{T}$, $\mathbf{C} = \{C(t)\}_{t=0}^{T}$. θ is the vector of unknown parameters. In detail, $f_t(\mathbf{S}, \mathbf{I}, \mathbf{Rep}, \mathbf{C}|\theta)$ is the probability of all observed infections in time interval $[t,\ t+1)$ being infected, and all observed non-infected individuals in time interval $[t,\ t+1)$ not being infected.

Our objective was to minimise the negative log posterior, which by equation (2.3) is proportional to - $\log\left(L(\mathbf{S},\mathbf{I},\mathbf{Rep},\mathbf{C}|\theta)\pi(\theta)\right)$, where $\pi(\theta)$ denotes the parameter prior distributions. We define $\mathcal{L}$ as the log likelihood, where

$$\mathcal{L} = \sum_{t=0}^{T} \log(f_t(\mathbf{S},\mathbf{I},\mathbf{Rep},\mathbf{C}|\theta) \tag{3.4}$$

$$= \sum_{t=0}^{T} \left(\left[\sum_{i \in I(t+1)\setminus I(t)} \log(p(i,t)) \right] + \left[\sum_{i \in S(t+1)} \log((1-p(i,t)) \right] \right). \tag{3.5}$$

We used a uniform prior for each parameter that, except for a belief that the posterior distribution resided within the given bounds, corresponded to having no initial preference towards specific values: $t_c \sim \text{Uniform}(0, 0.1)$; $\alpha \sim \text{Uniform}(-1, 10)$; $p_c \sim \text{Uniform}(0, 2)$; $q_c \sim \text{Uniform}(0, 2)$; $\epsilon \sim \text{Uniform}(0, 1)$; $\epsilon_{\text{water}} \sim \text{Uniform}(0, 1)$; $\epsilon_{\text{rice}} \sim \text{Uniform}(0, 1)$; $\epsilon_{\text{LBM}} \sim \text{Uniform}(0, 1)$; $\epsilon_{\text{ducks}} \sim \text{Uniform}(0, 1)$. The posterior distribution was explored via Markov chain Monte Carlo (MCMC) using the slice sampling method [109] (algorithm 2) or adaptive Metropolis parameter updates [110] (algorithm 3), acquiring 10^4 samples for district level models and 10^3 samples for division level models. Across all models we used a minimum burn-in period of 10^4 steps, with a variable thinning factor applied to ensure reasonable decorrelation of the parameter chains. A selection of MCMC diagnostic checks are presented in appendix A.3.

Model comparison methodology

To compare models of the same type, differing only by the value of the reporting delay, we used deviance information criterion (DIC) [148, 149], calculated using the samples generated from our MCMC simulations. We chose to predict the reporting delay based on the DIC in order to account for differences in the effective number of parameters of the fitted models, with the fixed time that gave the lowest DIC being preferred.

Further, after selecting the reporting delay that should be used for each of our models, DIC was used again to compare our set of nested models (at a given spatial level for a specific poultry epidemic wave). This was due to its capability of accounting for additional parameters increasing model complexity. While models with smaller DIC were preferred over models with larger DIC, note that models with a DIC value within two of the model with the lowest DIC value still deserved consideration, while

being at least three greater meant there was considerably less support for that model given the data [148].

Defining the deviance as $D(\theta) = -2\log(p(y|\theta))$, where y are the data and θ are the unknown parameters of the model, the deviance information criterion is calculated as

$$\text{DIC} = p_D + \bar{D}.$$

$\bar{D} = \mathbb{E}[D(\theta)]$ is a measure of how well the model fits the data and p_D is a measure of the effective number of parameters (favouring models with a smaller number of parameters). For the effective number of parameters we use the form suggested by Gelman *et al.* [149], where $p_D = \frac{1}{2}\hat{\text{var}}(D(\theta))$.

3.2.4 Zoonotic transmission model

A simple zoonotic transmission model was constructed to fit to the temporal human case data. We focused solely on the within-region epidemic time period. The rate of spillover transmission on a given day t, $\lambda(t)$, was chosen to have the following dependencies,

$$\lambda(t) = \beta I_b(t) + \epsilon_h, \tag{3.6}$$

where β is the poultry to human transmission rate, $I_b(t)$ is the number of infected poultry within the region of interest and ϵ_h is a constant human spark term.

With previous work finding the Poisson distribution provided an adequate goodness of fit to daily human H5N1 case data in Egypt [150], we assumed the occurrence of human cases followed a Poisson process. As a result, the waiting time until the next human case occurrence followed an exponential distribution. The probability of a human infection event occurring in the next day (i.e. $\delta t = 1$) was given by

$$h_t = 1 - e^{-\lambda(t)}.$$

Over the entire poultry epidemic a likelihood function for human case occurrence, L_h, could be constructed,

$$L_h = \left(\prod_{i \in D_{\text{inf}}} (h_i)\right)\left(\prod_{j \in D_{\text{sus}}} (1 - h_j)\right), \tag{3.7}$$

with D_{inf} the set of days a human case occurred and D_{sus} the set of days there were no

human cases. In detail, the first term corresponds to the probability of a human case occurring on days where the human case data reported at least one person becoming infected, with the second term giving the probability that no human cases occurred on all other days. Subsequently, the log-likelihood $\log(L_h)$ could be derived:

$$\log(L_h) = \left(\sum_{i \in D_{\text{inf}}} \log(h_i) \right) + \left(\sum_{j \in D_{\text{sus}}} \log(1 - h_j) \right), \tag{3.8}$$

With human cases only occurring in epidemic waves 2, 5 and 6, we applied our model to these waves only. Relationships between β and ϵ_h were analysed by producing log-likelihood surfaces using equation (3.8). Parameter summary statistics were inferred using 4×10^4 samples generated via MCMC with adaptive Metropolis updates [110], thinning by a factor of 40 samples with a burn-in period of 10^5 steps. The following uniform prior distributions were used, to again capture our initial belief of the distribution of plausible values residing within the given bounds (but for any value in the range, assigning no extra weight of support for that specific value over any other): $\beta \sim \text{Uniform}(0, 1)$, $\epsilon_h \sim \text{Uniform}(0, 1)$. A selection of MCMC diagnostic checks are presented in appendix A.3.

3.2.5 Model verification

To verify the validity of our model fitting we performed stochastic simulations, checking the correspondence of temporal and spatial summary statistics with the observed data. Our simulated poultry model was a spatial individual-based model at the premises level. It incorporated both the Tau-leap algorithm [120], allowing multiple events to occur each time step, and a grid-based approach outlined by Keeling and Rohani [72]. In addition, we accounted for zoonotic transmission over the entire poultry epidemic in each simulation.

We carried out 1000 simulation runs for each model. Both model components used distinct sampled parameter values, obtained previously via MCMC, in each run. The number of time steps matched the length of the epidemic wave the particular model had been fitted to. For the models fitted using region-specific epidemic dates we initialised the simulation with a single infected premises, corresponding to the location first reporting infection for that respective epidemic wave and spatial level. For the models fitted using country-wide epidemic dates all premises were initialised as susceptible. For infected premises the time between reporting and culling was randomly sampled from the observed reporting to culling time empirical probability

mass function.

For both premises and human cases our first goodness-of-fit check was to compare the distribution of simulated final epidemic sizes to the observed data. For the poultry model we also inspected reported case temporal profiles to ensure our simulations produced similar behaviour.

A separate class of goodness-of-fit tests focused on spatial aspects. First order spatial patterns were compared by computing the difference between a density surface of the observed case locations and a density surface of the predicted case locations averaged over 1% of simulations with the largest aggregate two-dimensional correlation with the data (when aggregated by sub-district). Further, we used Ripley's K function [151, 152] to ascertain whether the measure of clustering in the spatial pattern of observed infected premises could be plausibly generated by our fitted models. The K function has the following theoretical description,

$K(v) = \lambda^{-1}\mathrm{E}[\text{number of extra events within distance } v \text{ of a randomly chosen event}]$

where λ is the density of events. We seeked an estimator $\hat{K}(v)$. As edge effects can arise, biasing $\hat{K(v)}$ (especially at large values of v), we used an edge-corrected estimator proposed by Ripley [151],

$$\hat{K}(v) = \hat{\lambda^{-1}} \sum_i \sum_{j \neq i} w(l_i, l_j)^{-1} \frac{I(d_{ij} < v)}{N}.$$

The density of events is estimated as $\hat{\lambda} = N/A$, where N is the observed number of points and A the area of the study region. d_{ij} is the distance between the ith and jth premises, and $I(x)$ is the indicator function with value 1 if x is true and 0 otherwise. The edge correction weighting function $w(l_i, l_j)$ has value 1 when the circle centred at l_i and with a radius d_{ij} (so it passes through the point l_j) is completely inside the study area. For the case where part of the circle falls outside the study area, the weighting value equals the proportion of the circle that lies inside the study area. Although $\hat{K}(v)$ can be determined for any v, it is common to only consider $v < (A/2)^{1/2}$. Therefore, we used $v < 28$ and $v < 125$ for the district and division models respectively.

To assess whether premises-to-premises transmission could be sustained without the need for importations or infections derived from external sources, one further test

was to compute premises-level basic reproductive ratios. This determined the expected number of other premises the given premises would infect if infected itself. Those premises with the greatest transmission potential would therefore be highlighted. Following Tildesley and Keeling [153], we calculated the premises-level basic reproductive ratio R_i as:

$$R_i = \sum_{j \neq i} \text{Prob}_{ij} = \sum_{j \neq i} 1 - e^{-\text{Rate}_{ij} P}$$

where P is the length of the infectious period and Prob_{ij} is the probability that premises i infects premises j over its entire infectious period. As clarification, spark terms were not considered in this analysis to ensure we were purely dealing with the force of infection generated from the premises. We averaged R_i over 1000 simulation runs for these tests. In each simulation we used distinct sampled parameter values obtained via MCMC, in addition to a reporting to culling time weighted by the corresponding empirical probability mass function.

We also assessed how accurately model parameters could be inferred from the data. To do this we performed simulations using sample outputs from the MCMC model fitting procedure, with each set of simulated data then fitted to the same model it had been generated from using MCMC methods. The estimated posterior parameter densities were then compared to the 'true' value.

All calculations and simulations were performed with MATLAB®.

3.3 Results

3.3.1 Poultry model selection

In our comparison of DIC values for models of the same type, differing only by the value of the reporting delay, a fixed reporting delay of seven days was common across both waves and spatial levels (figures 3.4 and 3.5, see tables A.1 and A.2 for a complete listing of model DIC values).

Consistency in the complexity of the best fit models varied for the two different administration levels of interest (see table 3.3 for a listing of preferred models). At the division level the added complexity in the nonlinear farm size model (model C) was preferred, though adding in additional spark terms that were risk factor specific seemingly brought no additional benefits. On the other hand, across both waves and sets of epidemic dates the chosen models for our district datasets covered a wider array of the possible model options (table 3.3). This implied that differing modelling characteristics were required based on the spatial scale of interest. Fitting to the wave 5 district datasets the parametric kernel model (model B) was preferred. For the wave 2 data, considering only models A-C, the nonlinear farm size model (model C) was chosen. Whilst carrying forward the nonlinear farm size model framework we determined the simple ecological model (model E) as being preferred relative to all candidate models (figure 3.6, see table A.3 for a complete listing of model DIC values).

Though the simple ecological model (model E) was only found to be preferred for the wave 2 district datasets, the majority of the remaining datasets found that this model had a DIC value within two of the DIC value for the best-fit model, meaning such a model was still plausible given the data. In addition, for the majority of our datasets the full ecological model (model D) was found to have considerably less support relative to the simple ecological model (see table A.3).

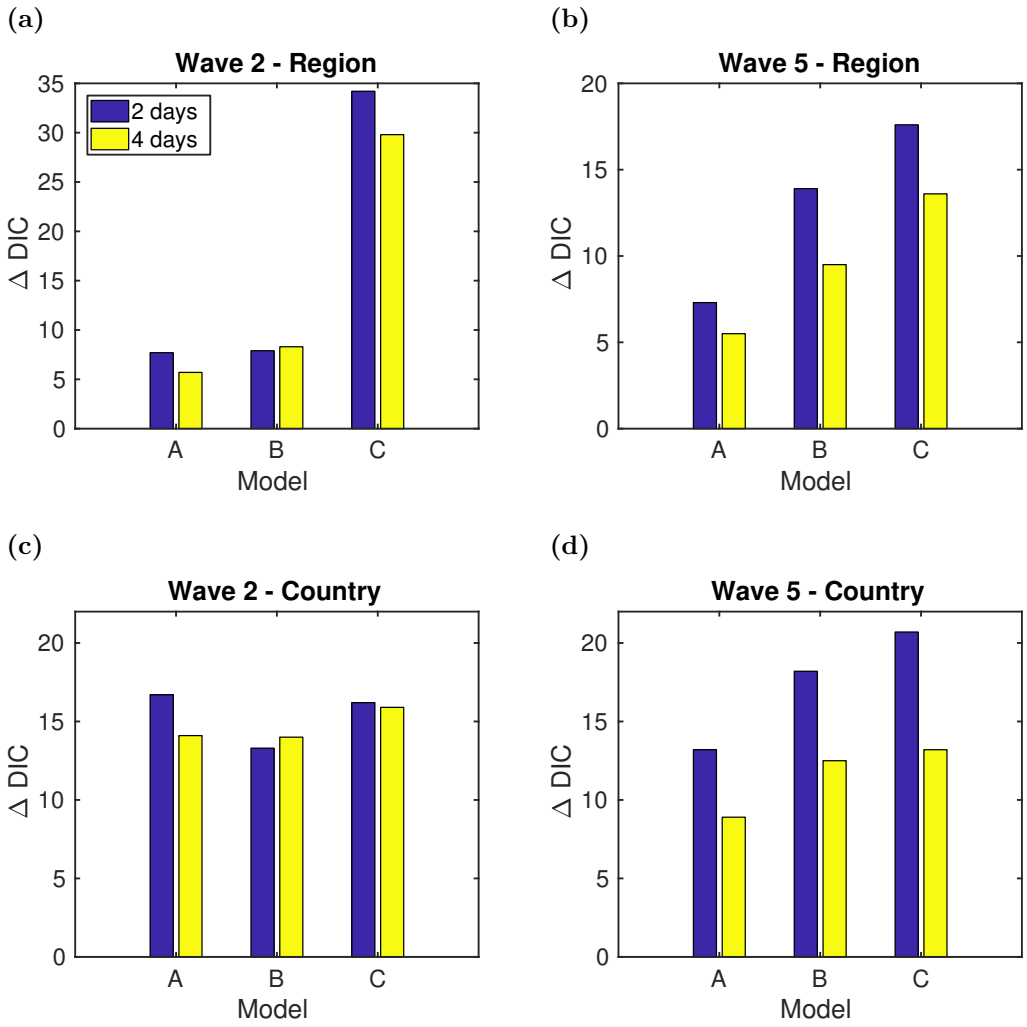

Figure 3.4: Bar plots comparing ΔDIC values for the district datasets fitted to our models with various fixed reporting delay times. For each model a reporting delay of seven days gave the minimum DIC value. The depicted ΔDIC is with respect to the version of the model using a seven day reporting delay. **(a)** Wave 2, region-specific epidemic dates; **(b)** wave 5, region-specific epidemic dates; **(c)** wave 2, country-wide epidemic dates; **(d)** wave 5, country-wide epidemic dates.

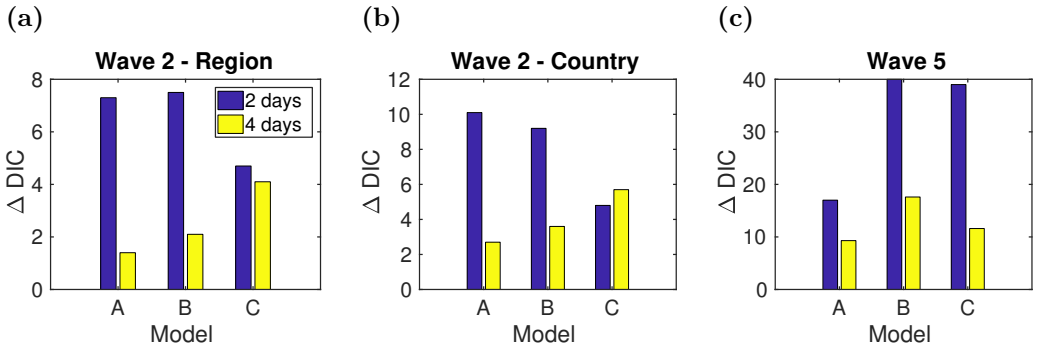

Figure 3.5: Bar plots comparing ΔDIC values for the division datasets fitted to our models with various fixed reporting delay times. For each model a reporting delay of seven days gave the minimum DIC value. The depicted ΔDIC is with respect to the version of the model using a seven day reporting delay. **(a)** Wave 2, region-specific epidemic dates; **(b)** wave 2, country-wide epidemic dates; **(c)** wave 5.

Table 3.3: Model fitting summary Preferred models and fixed reporting delay time for each wave and spatial level.

	Epidemic dates	**District**	**Division**
Wave 2	Region	Model E / 7 days	Model C / 7 days
	Country	Model E / 7 days	Model C / 7 days
Wave 5	Region	Model B / 7 days	Model C / 7 days
	Country	Model B / 7 days	-

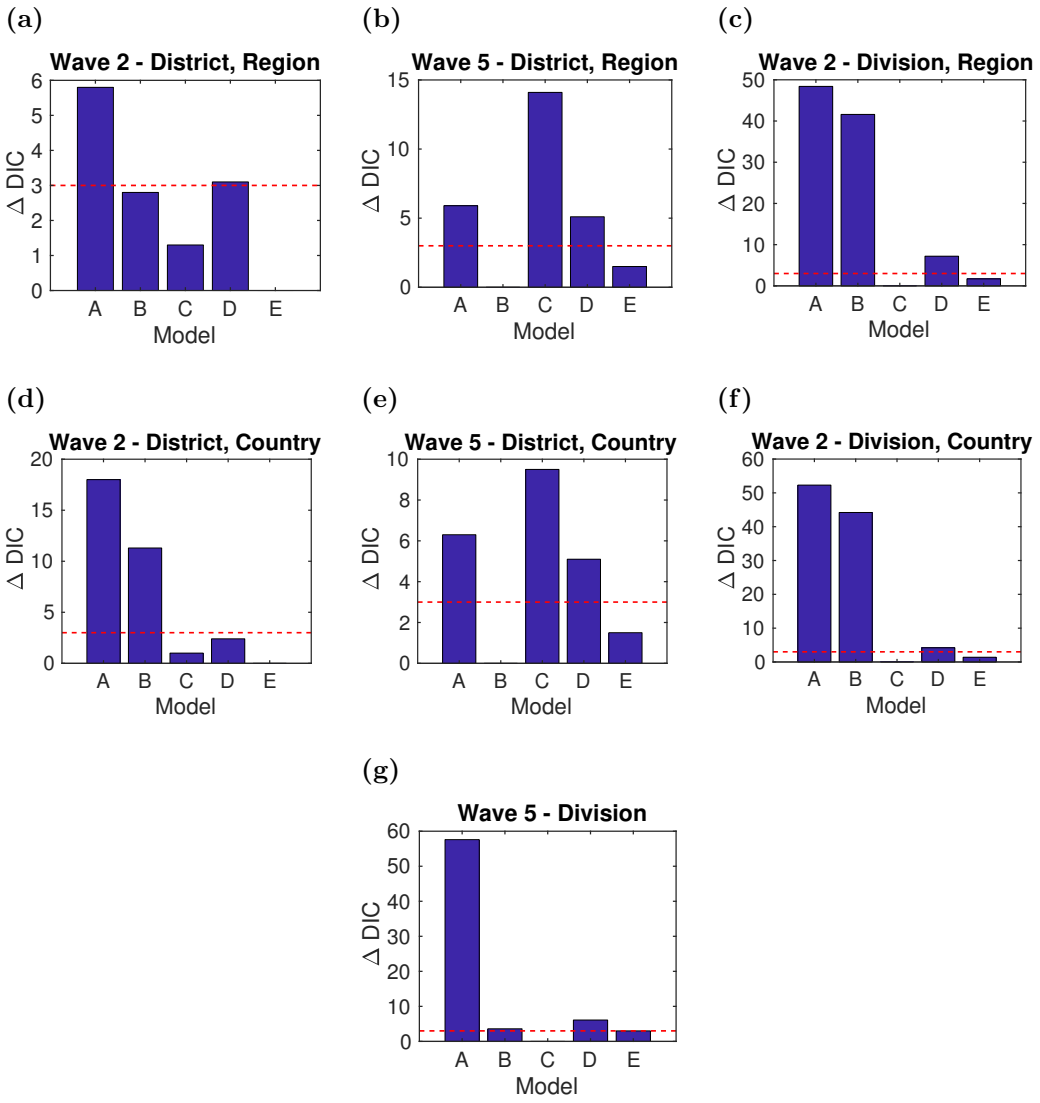

Figure 3.6: Bar plots comparing ΔDIC values for the different datasets fitted to our nested models. For each model the fixed reporting delay time that minimised the DIC was used (see tables A.1 to A.2). The preferred model had a ΔDIC= 0. Models with ΔDIC≥ 3 have considerably less support and lie above the red, dashed line. Full DIC values are given in table A.3. **(a)** Wave 2 district, region-specific epidemic dates; **(b)** wave 5 district, region-specific epidemic dates; **(c)** wave 2 division, region-specific epidemic dates; **(d)** wave 2 district, country-wide epidemic dates; **(e)** wave 5 district, country-wide epidemic dates; **(f)** wave 2 division, country-wide epidemic dates; **(g)** wave 5 division.

3.3.2 Poultry model parameter distributions

When comparing parameter summary statistics for our best fit models at the division level (table 3.6) and district level (see tables 3.4 to 3.6), a spatial level specific feature was the apparent greater contribution of importations and transmission from other sources (characterised by the ϵ parameters) to the force of infection at the district level versus the division level. On the other hand, the relationship between increasing flock size and premises-level susceptibility was approximately linear (i.e. $p \approx 1$) in both model types (tables 3.4 to 3.6). A reasonable level of identifiability was observed for model parameters, giving extra confidence to our results (figures 3.7 and 3.8).

Comparing the estimated parameter distributions for our wave 2 and wave 5 division-level models highlights noticeable differences in the factors driving disease spread across the two waves. Of particular interest were discrepancies in α, the transmission kernel parameter, and q, the infected premises population exponent (see table 3.6). While for wave 2 α was typically below 0, for wave 5 α was approximately zero, giving a stronger preference towards short-range transmission (figure A.3). For q, fitting to the wave 2 epidemic found approximately equal contributions to the force of infection from each infected premises against a susceptible premises, irrespective of the infected premises population size. In stark contrast, fitting to the wave 5 epidemic we inferred the median value of q to be greater than one, implying poultry premises with the largest populations had a significant role in H5N1 transmission.

Table 3.4: Parameter summary statistics for preferred district models using region-specific epidemic dates. Parameter mean, median and 95% credible intervals from 10^4 samples obtained from MCMC. Examples of MCMC diagnostic realisations acquired when fitting to these datasets are presented in figures A.4 and A.5.

		Wave 2	**Wave 5**
	Model	E	B
	Inf. to Rep. time (days)	7	7
$\mathbf{t_c}$	Mean	2.44×10^{-6}	5.29×10^{-10}
	Median	4.65×10^{-7}	4.51×10^{-10}
	(95% CI)	$(1.81 \times 10^{-8}, 1.92 \times 10^{-5})$	$(1.00 \times 10^{-10}, 1.38 \times 10^{-11})$
$\boldsymbol{\epsilon}$	Mean	—	7.53×10^{-5}
	Median	—	7.00×10^{-5}
	(95% CI)	—	$(2.18 \times 10^{-5}, 1.59 \times 10^{-4})$
$\boldsymbol{\alpha}$	Mean	-0.0617	-0.193
	Median	-0.0642	-0.189
	(95% CI)	(-0.384, 0.270)	(-0.528, 0.126)
$\mathbf{p}$	Mean	1.07	—
	Median	1.07	—
	(95% CI)	(0.610, 1.45)	—
$\mathbf{q}$	Mean	0.0926	—
	Median	0.0699	—
	(95% CI)	(0.00244, 0.298)	—
$\boldsymbol{\epsilon}_{\text{water}}$	Mean	2.72×10^{-5}	—
	Median	1.94×10^{-5}	—
	(95% CI)	$(6.20 \times 10^{-7}, 8.95 \times 10^{-5})$	—
$\boldsymbol{\epsilon}_{\text{LBM}}$	Mean	6.88×10^{-5}	—
	Median	6.23×10^{-5}	—
	(95% CI)	$(7.14 \times 10^{-6}, 1.60 \times 10^{-4})$	—

Table 3.5: Parameter summary statistics for preferred district models using country-wide epidemic dates. Parameter mean, median and 95% credible intervals from 10^4 samples obtained from MCMC. Transmission kernels using the median inferred α values are depicted in figure A.2.

		Wave 2	**Wave 5**
	Model	E	B
	Inf. to Rep. time (days)	7	7
$\mathbf{t_c}$	Mean Median (95% CI)	2.67×10^{-6} 4.19×10^{-7} (6.67×10^{-9}, 2.23×10^{-5})	4.96×10^{-10} 4.29×10^{-10} (9.61×10^{-11}, 1.28×10^{-9})
$\boldsymbol{\epsilon}$	Mean Median (95% CI)	— — —	3.37×10^{-5} 3.10×10^{-5} (9.11×10^{-6}, 7.37×10^{-5})
$\boldsymbol{\alpha}$	Mean Median (95% CI)	-0.142 -0.136 (-0.464, 0.144)	-0.236 -0.232 (-0.555, 0.0509)
$\mathbf{p}$	Mean Median (95% CI)	1.06 1.07 (0.617, 1.46)	— — —
$\mathbf{q}$	Mean Median (95% CI)	0.110 0.0856 (0.00390, 0.343)	— — —
$\boldsymbol{\epsilon}_{\text{water}}$	Mean Median (95% CI)	1.02×10^{-5} 8.15×10^{-5} (3.32×10^{-7}, 3.07×10^{-5})	— — —
$\boldsymbol{\epsilon}_{\text{LBM}}$	Mean Median (95% CI)	1.59×10^{-5} 1.38×10^{-5} (8.27×10^{-7}, 4.21×10^{-5})	— — —

Table 3.6: Parameter summary statistics for preferred division models. Parameter mean, median and 95% credible intervals (CI) from 10^3 samples obtained from MCMC. Transmission kernels using the median inferred α values are depicted in figure A.3. Examples of MCMC diagnostic realisations acquired when fitting to these datasets are presented in figures A.6 to A.8.

		Wave 2		**Wave 5**
		Region-specific	**Country-wide**	**Region-specific**
	Model	C	C	C
	Inf. to Rep. time (days)	7	7	7
$\mathbf{t_c}$	Mean Median (95% CI)	1.06×10^{-7} 7.70×10^{-8} (7.29×10^{-9}, 3.78×10^{-7})	9.63×10^{-8} 7.03×10^{-8} (3.71×10^{-9}, 3.42×10^{-7})	1.71×10^{-10} 1.56×10^{-10} (5.86×10^{-11}, 3.63×10^{-10})
$\boldsymbol{\epsilon}$	Mean Median (95% CI)	4.11×10^{-6} 3.98×10^{-6} (1.42×10^{-6}, 7.91×10^{-6})	2.49×10^{-6} 2.35×10^{-6} (8.80×10^{-7}, 4.80×10^{-6})	1.04×10^{-5} 1.02×10^{-5} (5.02×10^{-6}, 1.72×10^{-5})
$\boldsymbol{\alpha}$	Mean Median (95% CI)	-0.358 -0.345 (-0.666, -0.159)	-0.394 -0.377 (-0.713, -0.172)	0.0136 0.0136 (-0.122, 0.143)
$\mathbf{p}$	Mean Median (95% CI)	1.06 1.06 (0.923, 1.19)	1.05 1.06 (0.916, 1.18)	1.05 1.05 (0.826, 1.26)
$\mathbf{q}$	Mean Median (95% CI)	0.0574 0.0427 (0.00175, 0.189)	0.0732 0.0458 (0.00243, 0.300)	1.06 1.06 (0.844, 1.28)

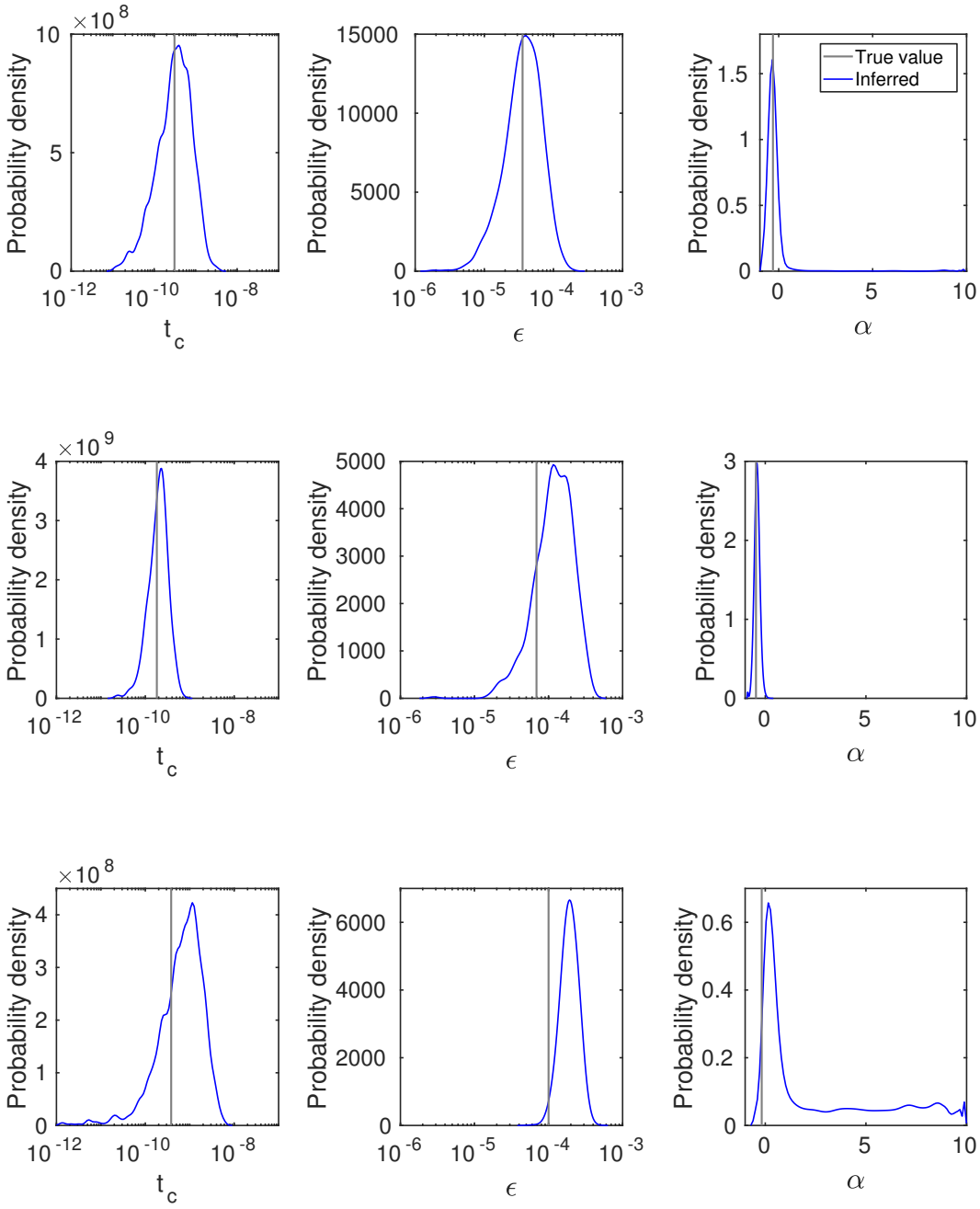

Figure 3.7: Inferred probability densities for poultry transmission model parameters from three simulated samples, using parameter values obtained when fitting to the wave 5 district data with region-specific epidemic dates. True parameter values are depicted by vertical grey bars. Estimated probability densities are given by the solid blue lines. First column: t_c. Second column: ϵ. Third column: α. We found that true values for all parameters could generally be reliably recovered, although there were instances where reasonable uncertainty remained in the posterior distribution for α (as shown in row 3).

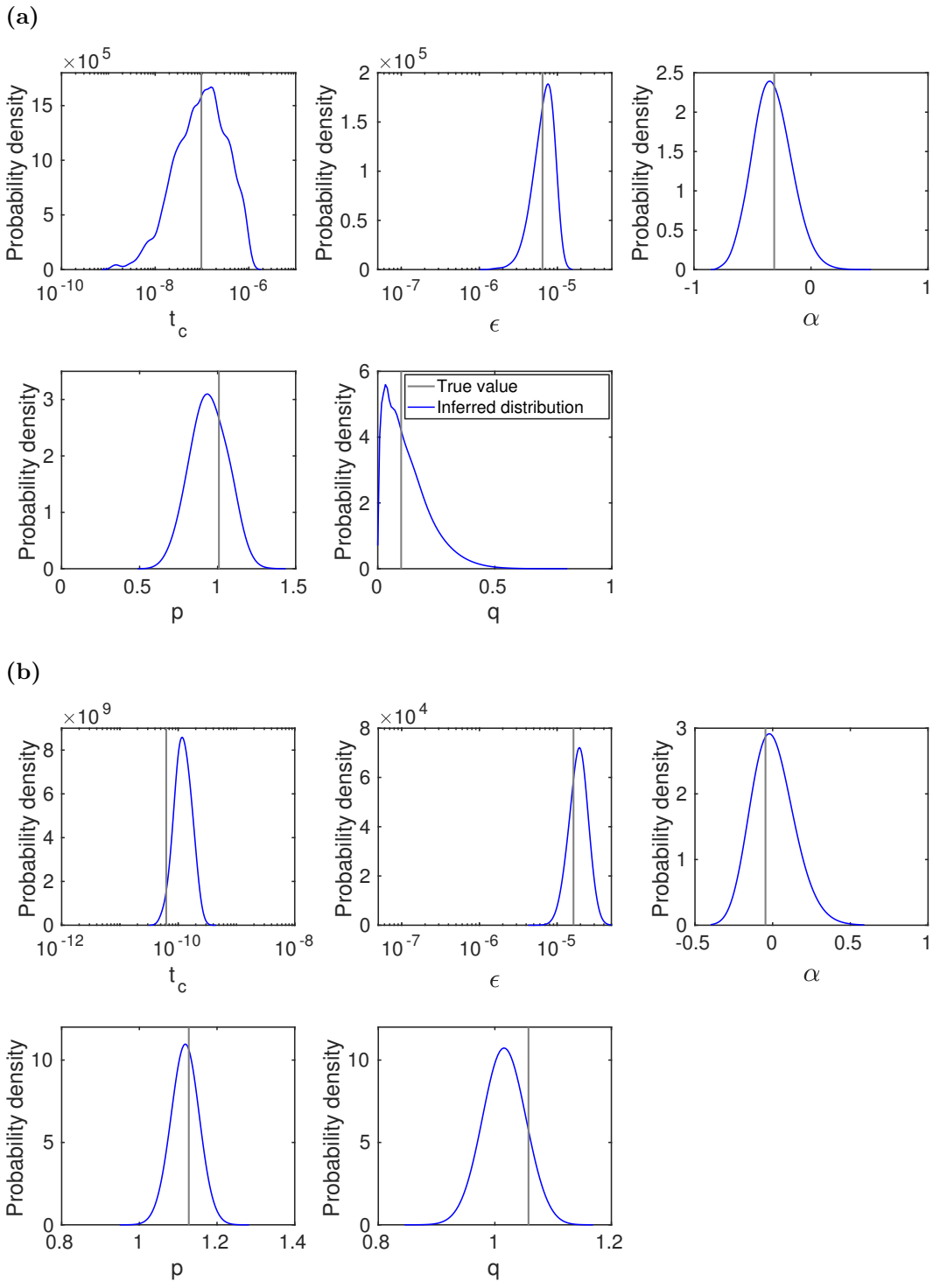

Figure 3.8: Inferred probability densities for poultry transmission model parameters, using parameter values obtained when fitting to division-level data with region-specific epidemic dates. True parameter values are depicted by vertical grey bars. Estimated probability densities are given by the solid blue lines. The true parameter values could be reliably recovered for both **(a)** wave 2 and **(b)** wave 5 datasets.

3.3.3 Zoonotic transmission model

Log-likelihood surfaces were produced for waves 2, 5, and 6 using equation (3.8). Two preferred regions of parameter space were found in general, though there is evidence of parameter dependent threshold values (figures 3.9(a) and 3.9(b)). Below these threshold values the other parameter dominates the dynamics of the system. When fitting to the wave 5 data there was little dependence upon the spark term ϵ_h, with β playing a much more significant role. In other words, the number of infected birds has some significance in the likelihood of zoonotic transmission occurring. The opposite was found to be true for wave 2, with more importance placed on the human case spark term.

Within wave 6 there were three human H5N1 case occurrences. All three were situated inside the Dhaka division, while two of the three were contained within the Dhaka district. At both spatial levels there was very little dependence on the number of infected poultry, with the spark term ϵ_h being dominant (figures 3.9(c) and 3.9(d)).

Parameter summary statistics inferred from 4×10^4 samples generated via MCMC are stated in table A.4. Premises infection dates were computed from the observed reporting dates using the fixed reporting delay time in the preferred model for the respective wave and spatial level. Acceptance rates were between 20 and 25%. As for the poultry transmission component, we were able to recover unbiased estimates of the zoonotic transmission model parameters from simulated data (figures 3.10 and 3.11).

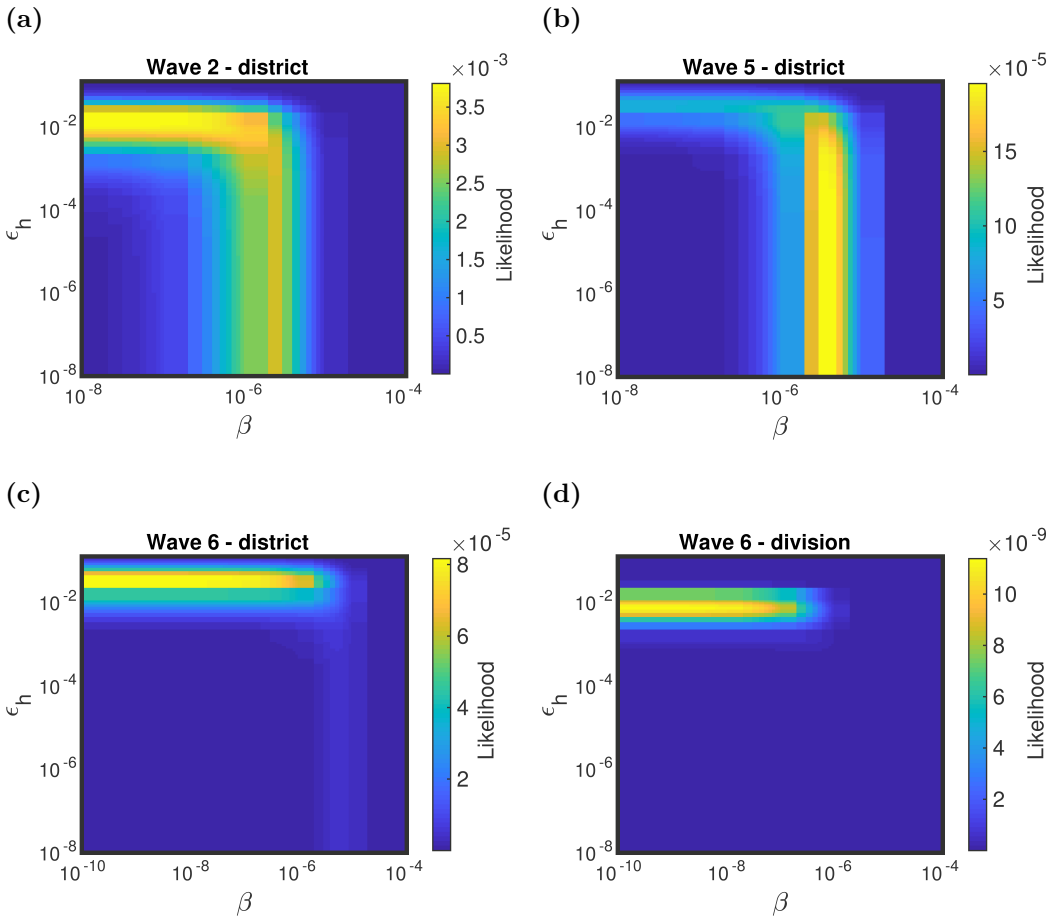

Figure 3.9: 2D likelihood surface of temporal zoonotic transmission model parameters. Lighter colours signify a higher likelihood. The preferred regions of parameter space when fitting to the following datasets were: **(a)** Wave 2 district - the dynamics were dominated by ϵ_h; **(b)** wave 5 district - little dependence upon the spark term ϵ_h, with β playing a much more significant role; **(c,d)** wave 6 district and division - the spark term ϵ_h was dominant at both spatial levels.

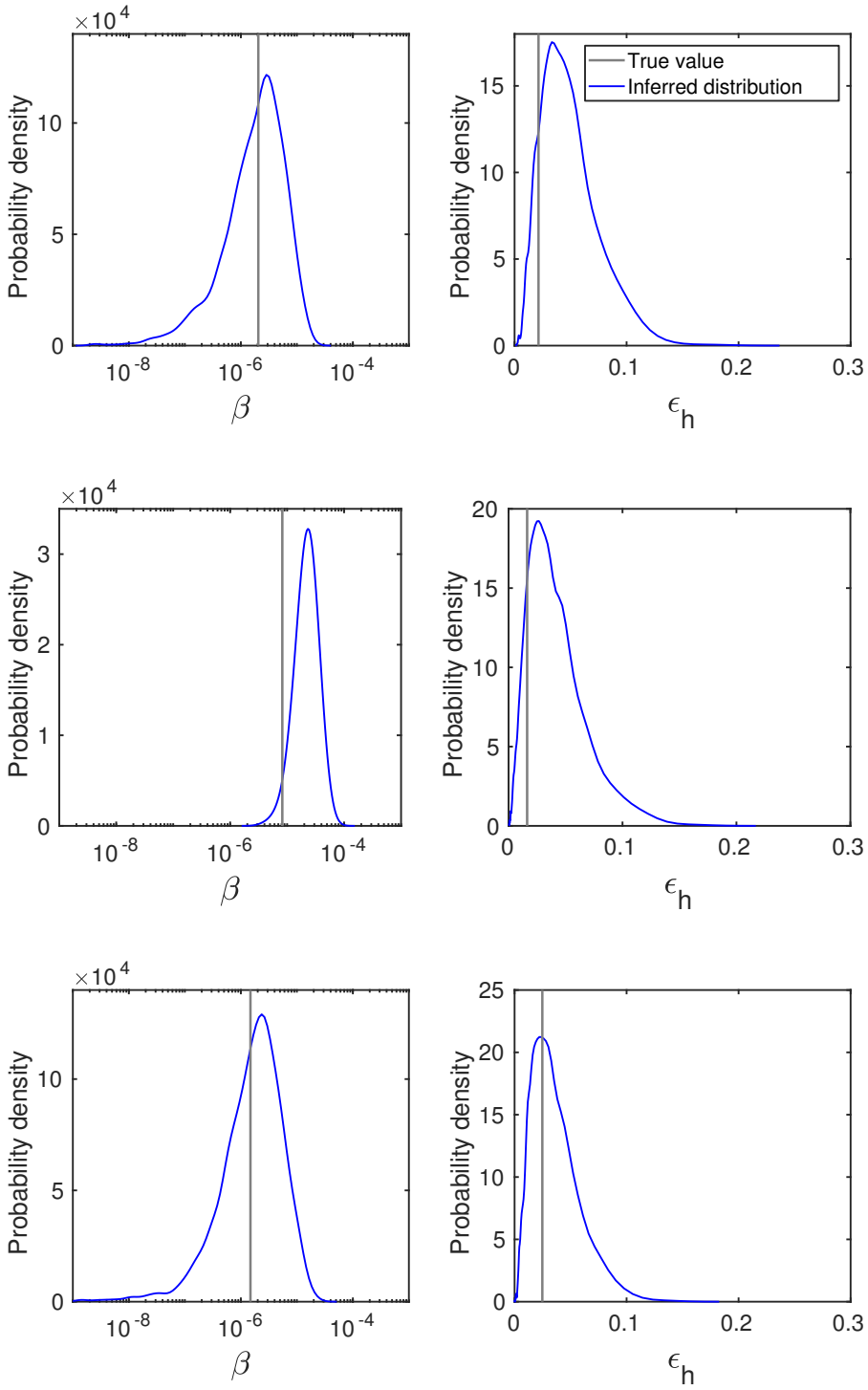

Figure 3.10: Inferred probability densities for zoonotic transmission model parameters from three simulated samples, using parameter values obtained when fitting to the wave 6 district data with region-specific epidemic dates. True parameter values are depicted by vertical grey bars. Estimated probability densities are given by the solid blue lines. We found that true values for both parameters could be reliably recovered. First column: β. Second column: ϵ_h.

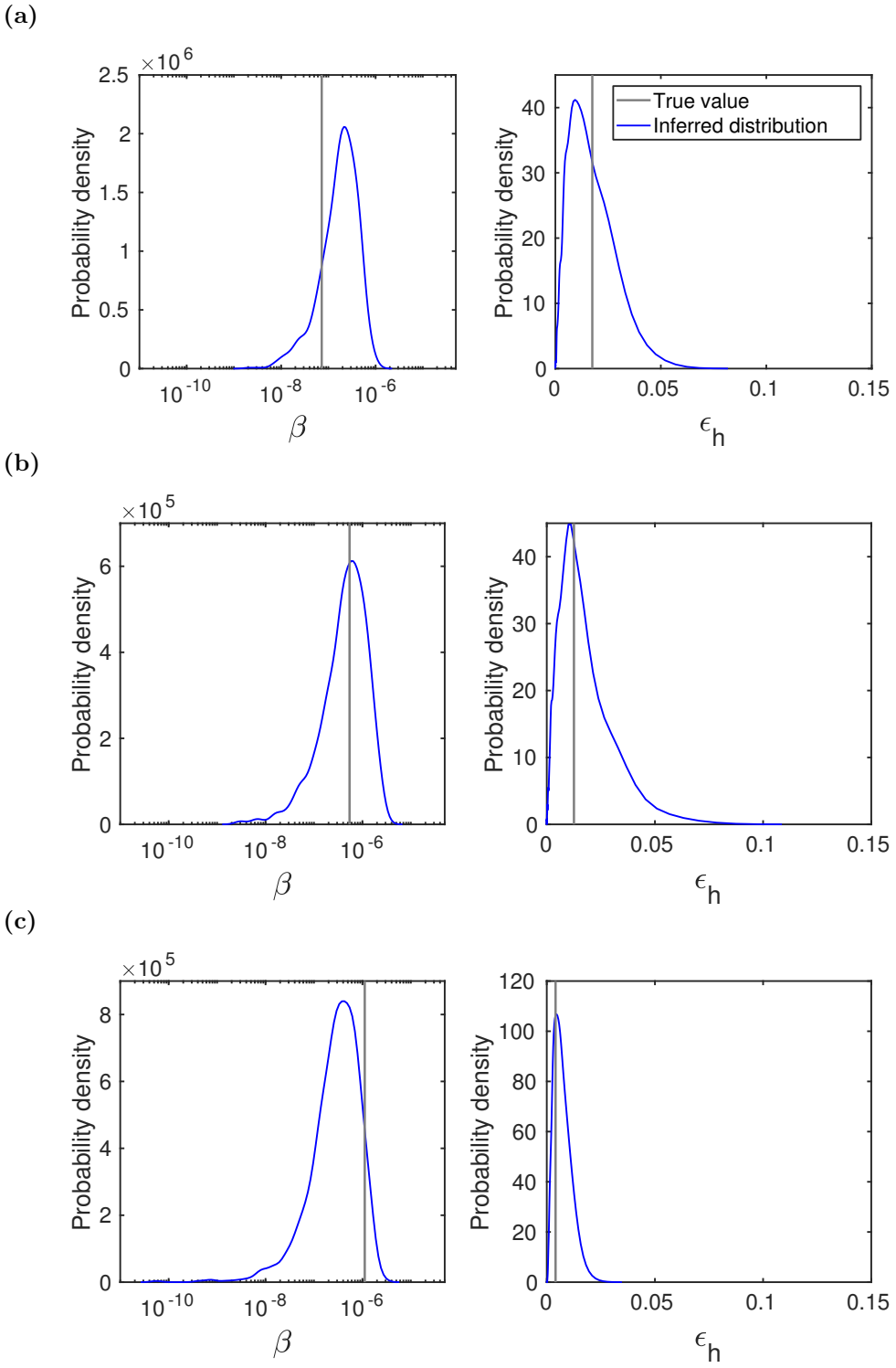

Figure 3.11: Inferred probability densities for zoonotic transmission model parameters, using parameter values obtained when fitting to division-level data with region-specific epidemic dates. True parameter values are depicted by vertical grey bars. Estimated probability densities are given by the solid blue lines. First column gives estimated densities for β; second column gives estimated densities for ϵ_h. We found that true values for both parameters could be reliably recovered. **(a, b)** Wave 2; **(c, d)** wave 5; **(e, f)** wave 6.

3.3.4 District-level model verification

Simulation output from our district-level models were found to agree favourably with the observed data. Using region-specific epidemic dates the simulated final premises and human epidemic sizes captured the observed data for both waves (figure 3.12). This was maintained when considering country-wide epidemic dates, though the distributions for both waves became heavy-tailed (figures 3.12(c) and 3.12(g)). Further, the observed spatial distribution of infected premises could be plausibly generated by our fitted models (figure 3.13). The simulated models exhibited a much broader range of possibilities for spatial structure, but the observed data was predominately within the 95% prediction interval (figure 3.14). Lastly, the fitted district-level models were capable of generating plausible reported case temporal profiles, which were dominated by single daily cases (figures 3.15(a) and 3.15(b)).

Across the majority of the district models the number of premises with premises-level reproductive ratios estimated to be greater than one was limited. Therefore, infections seeded in a random premises by importations or other ecological factors would more than likely fail to spread (figure 3.16). On the other hand, for both waves small clusters of premises, all with $R_i > 1$, were present in the centre third of the district (and in the south-west and north-west for wave 5). Our results indicate that localised outbreaks would be possible here, corresponding well with the true locations of the wave 5 observed cases in particular (see figure 3.2).

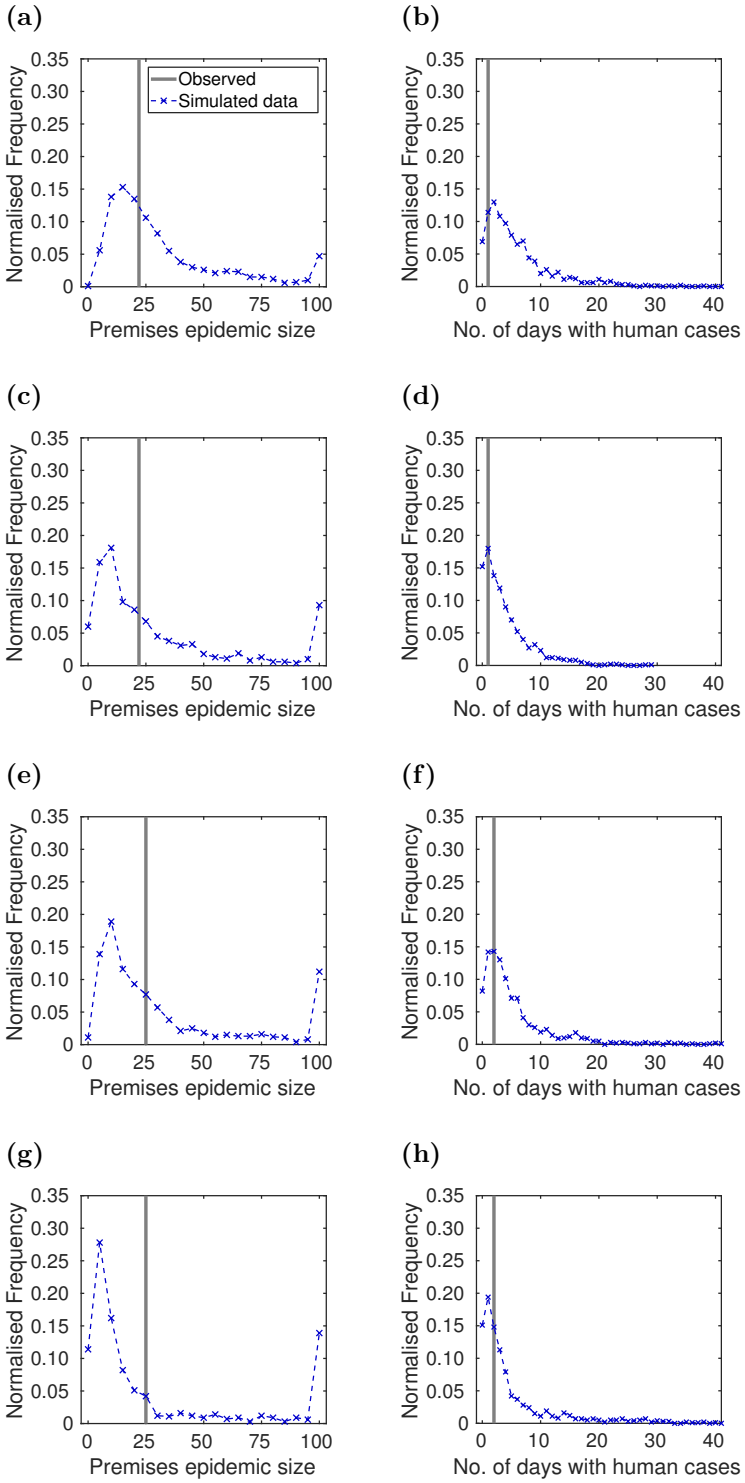

Figure 3.12: Simulated premises epidemic size and human case occurrence versus observed data at the district level. Left column shows the premises epidemic size versus the observed data for each of our district datasets. Similarly, the right column shows simulated human case occurrence verses the observed data. **(a, b)** Wave 2, region-specific epidemic dates; **(e, f)** wave 5, region-specific epidemic dates; **(c, d)** wave 2, country-wide epidemic dates; **(g, h)** wave 5, country-wide epidemic dates. The normalised frequency at 100 also includes all epidemic sizes 100 or greater.

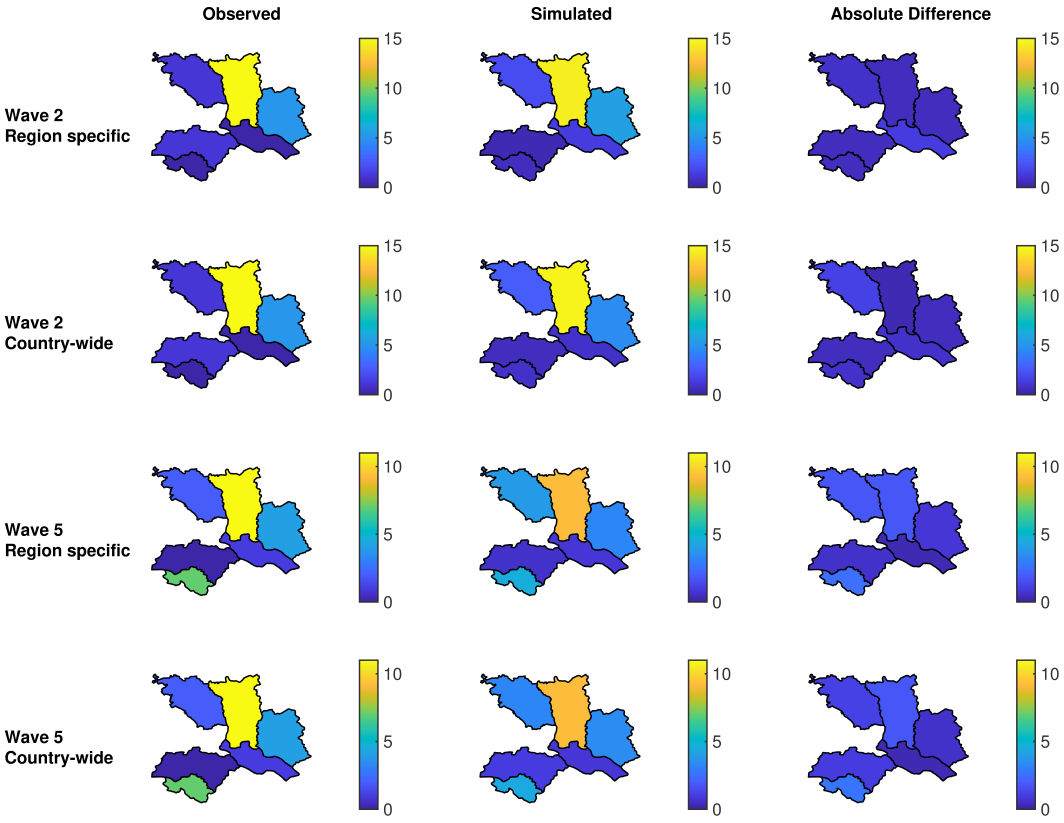

Figure 3.13: Simulated infected premises numbers aggregated at the sub-district level versus observed data for all district model datasets. For each district model dataset the first column shows the empirical data, with the second column showing the mean number of infected premises per sub-district obtained from ten simulations. The third column gives the absolute difference between these two values. In all panels lighter colours correspond to greater values. We see a reasonable spatial fit between the observed and simulated data, with the general spatial pattern captured by the model simulations.

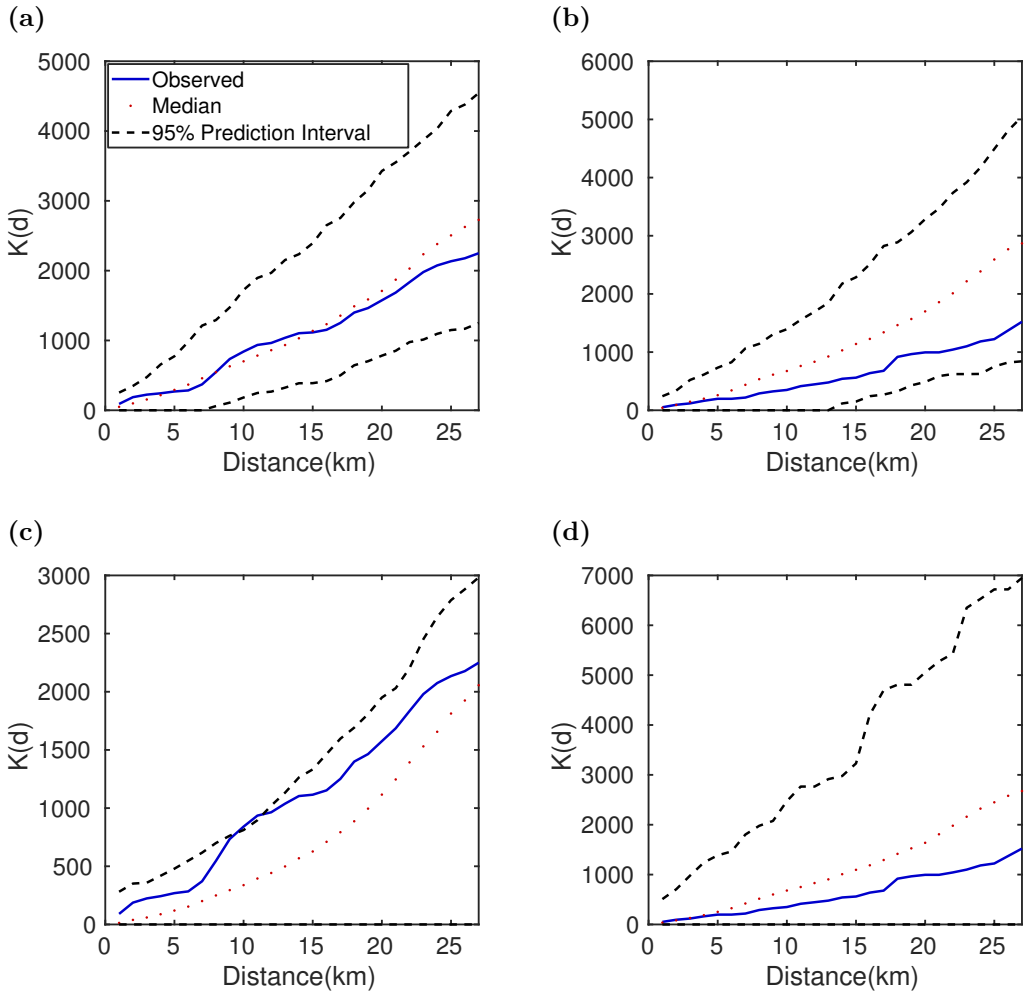

Figure 3.14: Observed Ripley's K function versus simulated Ripley's K function distribution for district-level models. The Ripley's K function for the observed infected premises data is given by the solid blue line. Median Ripley's K function estimated from simulated data is represented by the red dotted line, with the black dashed lines giving the 95% prediction interval bounds. The observed data lay predominately within the 95% prediction interval for all district-level models. **(a)** Wave 2, region-specific epidemic dates; **(b)** wave 5, region-specific epidemic dates; **(c)** wave 2, country-wide epidemic dates; **(d)** wave 5, country-wide epidemic dates.

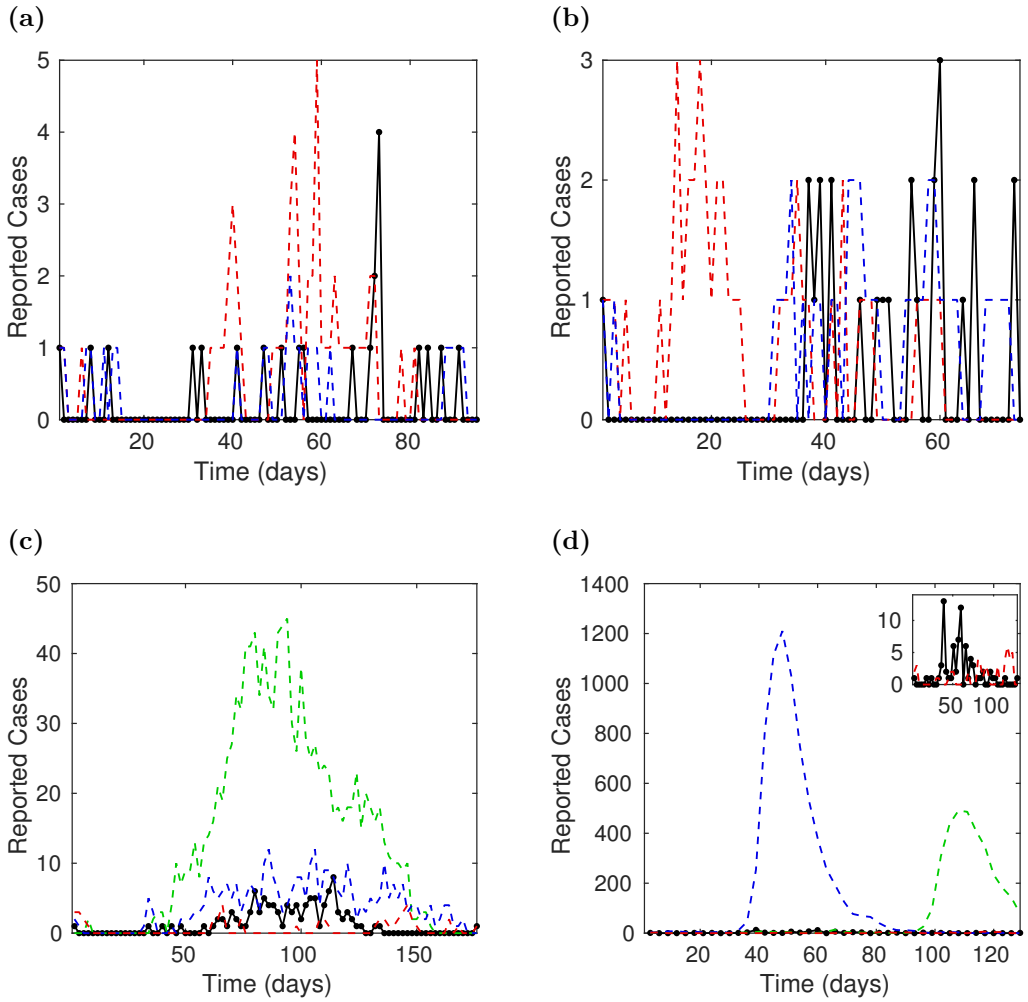

Figure 3.15: Observed and example simulated reported premises temporal profiles. Black dots denote the actual reported cases. Dashed lines are typical examples arising from simulations. Simulation results are from fitted models using region-specific epidemic dates. Note we display the number of reported cases per day unless stated otherwise. **(a)** Wave 2 district; **(b)** wave 5 district; **(c)** wave 2 division, with number of reported cases every two days; **(d)** wave 5 division, with number of reported cases grouped into three day intervals.

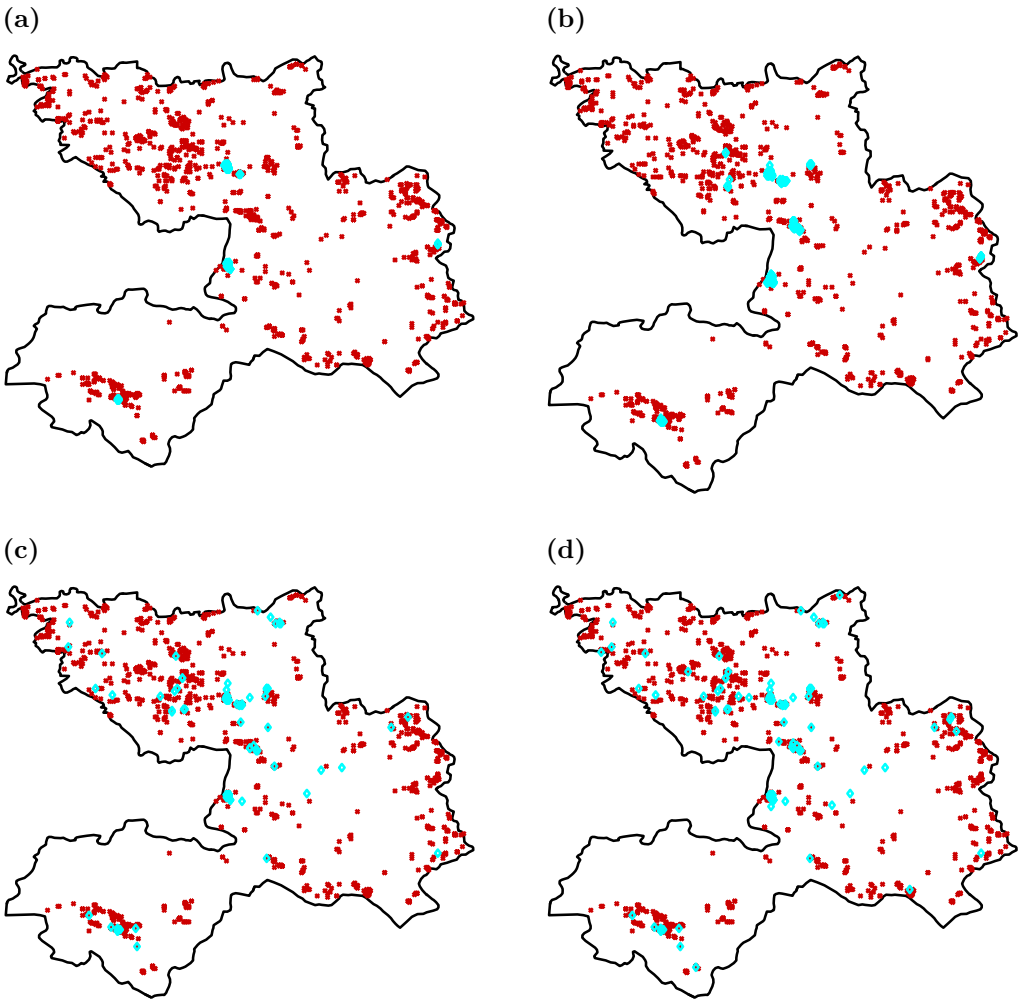

Figure 3.16: Binary indicator of premises-level reproductive ratios R_i being greater than one for the district-level models. Red crosses denote premises with a $R_i < 1$, cyan diamonds premises with $R_i \geq 1$. Across the models only a limited number of premises obtained premises-level reproductive ratios greater than one. However, small clusters of premises with $R_i > 1$ were present in the centre third of the district for both waves, and in the south-west and north-west for wave 5. **(a)** Wave 2, region-specific epidemic dates; **(b)** wave 2, country-wide epidemic dates; **(c)** wave 5, region-specific epidemic dates; **(d)** wave 5, country-wide epidemic dates.

3.3.5 Division-level model verification

For our division-level models the amount of agreement between the simulated output and the observed data was more variable. Although predictions from the models fitted to the wave 2 data generally underestimated the observed premises epidemic size (figures 3.17(a) and 3.17(c)), they did generate infection spatial distributions with comparable measures of spatial homogeneity (figures 3.19(a) and 3.19(b)). In contrast, for wave 5 we obtained a bimodal distribution, with the observed premises epidemic size lying just above the lower valued peak (figure 3.17(e)). The knock-on effect of this is a widely spanning human case distribution (figure 3.17(f)). Although first order spatial correspondence with the case data was not as strong for our fitted division-level models (compared to the district models), a subset of simulation runs could capture the prominent outbreak regions located centrally and on the eastern edge of the division (figure 3.18). Further, measures of spatial homogeneity in the spatial pattern of the observed infected premises data could be plausibly generated (figure 3.19(c)). Temporally, both waves exhibited two typical behaviours. These were either a single large outbreak, or a small outbreak with intermittent spikes in cases that mirrored the true temporal profiles (figures 3.15(c) and 3.15(d)).

Spatially, of particular interest was the wave 5 division model. The observed cases mainly lie in the centre third of the division, spanning the entire width of the region. However, model simulations found the regions infected most often lay further north of this band (figure 3.20). Analysis of premises-level reproductive ratios revealed the extent to which a number of premises are theoretically able to transmit infection. Both wave 2 division models gave the highest R_i values in a similar area, though these were only just above one and smaller in scale when compared to the wave 5 division model (figure 3.21). Furthermore, the wave 2 division-level models predict that the areas capable of continuing a transmission chain are concentrated in a single central region. In contrast, the wave 5 division-level model gave a smaller central region with $R_i > 1$, but indicated extra sporadic areas in the north and south-west with the capability of continuing the chain of transmission if infection arose in those localities (figures 3.21(g) to 3.21(i)).

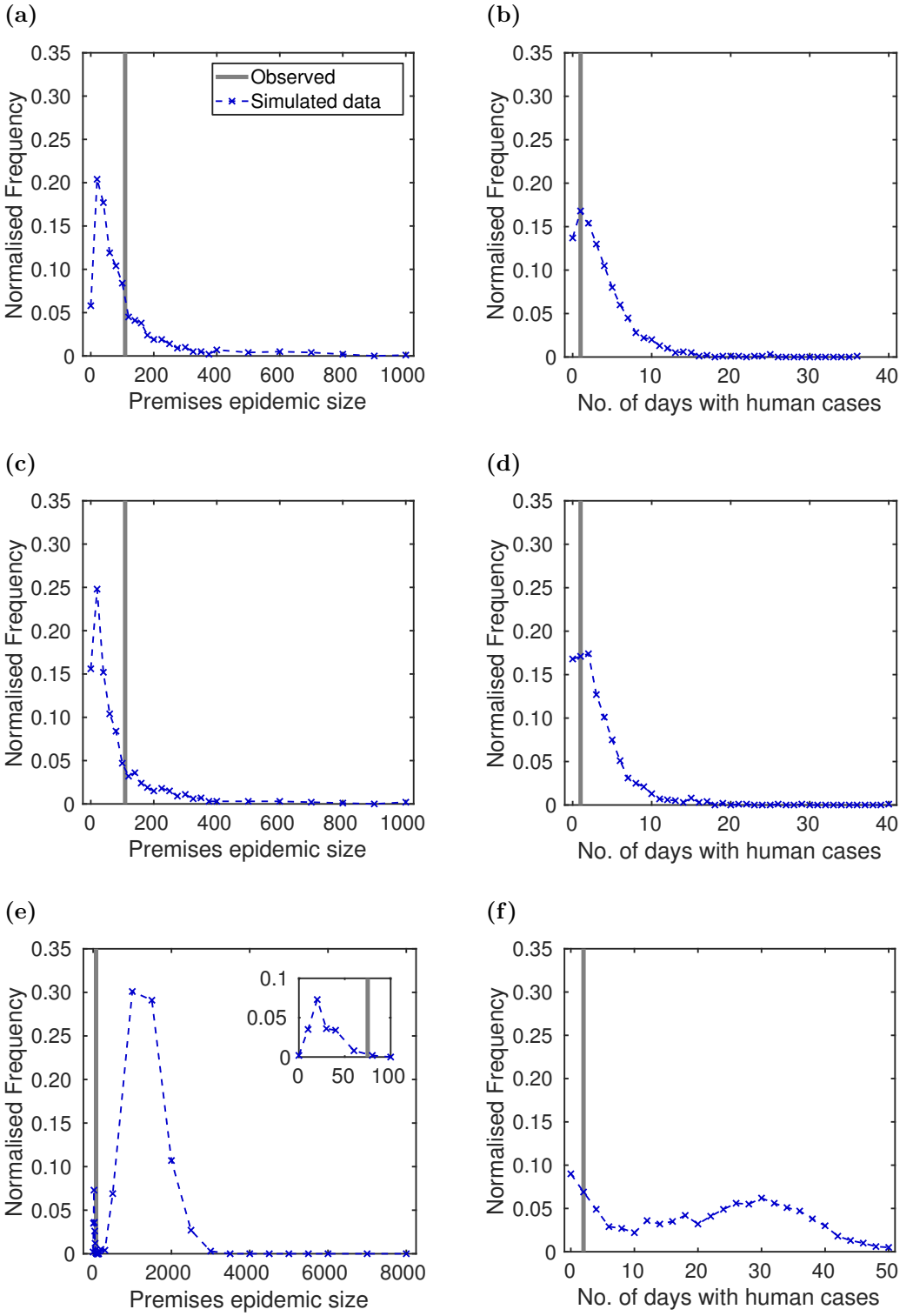

Figure 3.17: Simulated epidemic size versus observed data at the division level. **(a,c,e)** Poultry epidemic size. **(b,d,f)** Human epidemic size. First row **(a,b)** used wave 2 region-specific epidemic dates, second row **(c,d)** used wave 2 country-wide epidemic dates, final row **(e,f)** used wave 5 epidemic dates.

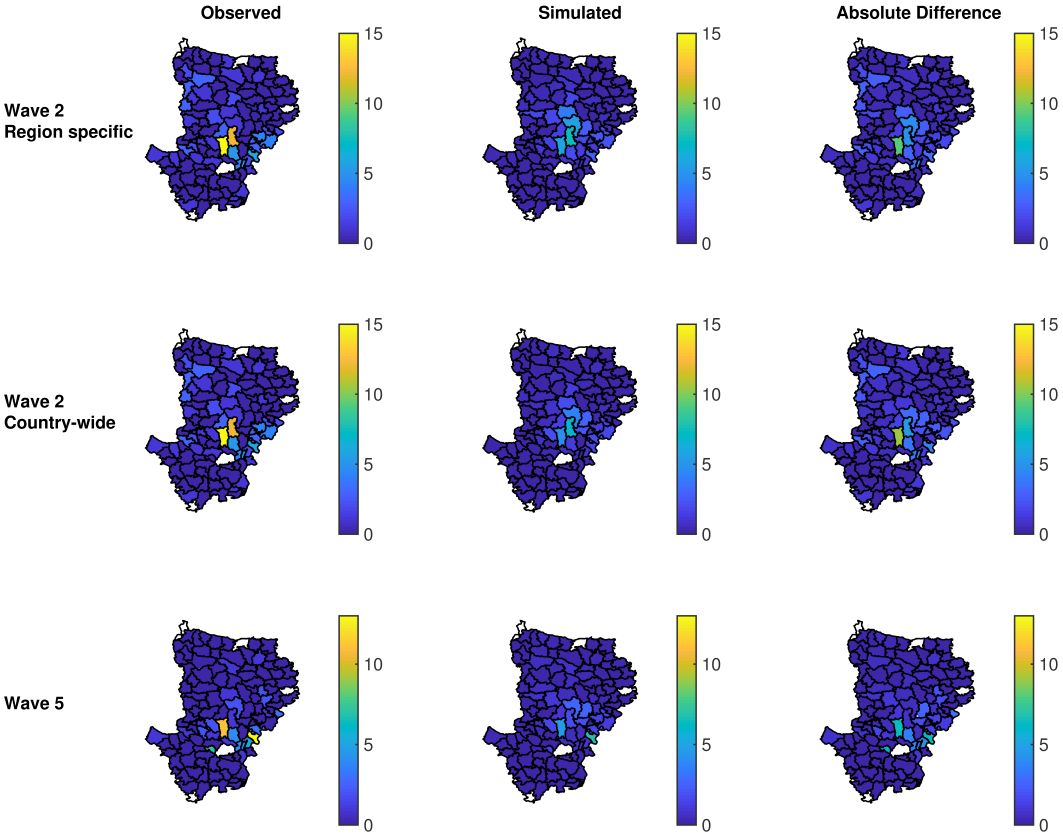

Figure 3.18: Simulated infected premises numbers aggregated at the sub-district level versus observed data for all division model datasets. For each division model dataset, the first column shows the empirical data, with the second column showing the mean number of infected premises per sub-district obtained from ten simulations. The third column gives the absolute difference between these two values. In all panels lighter colours correspond to greater values.

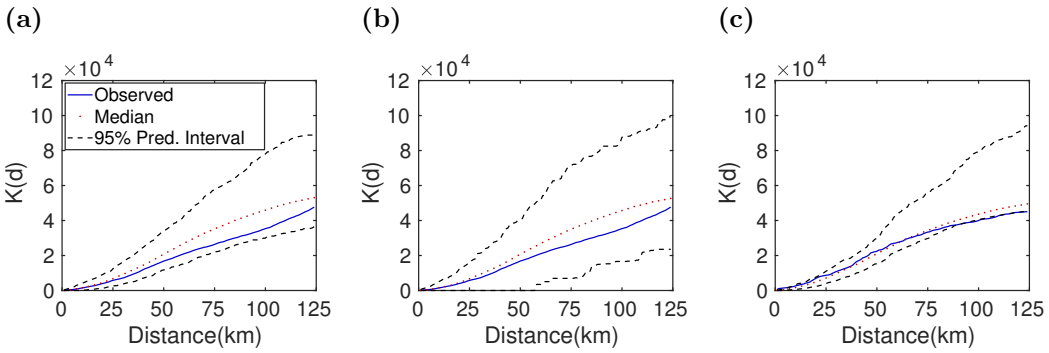

Figure 3.19: Observed Ripley's K function versus simulated Ripley's K function distribution for division-level models. The Ripley's K function for the observed infected premises data is given by the solid blue line. Median Ripley's K function estimated from simulated data is represented by the red dotted line, with the black dashed lines giving the 95% prediction interval bounds. The observed data lay within the 95% prediction interval for all division-level models. **(a)** Wave 2, region-specific epidemic dates; **(b)** wave 2, country-wide epidemic dates; **(c)** wave 5.

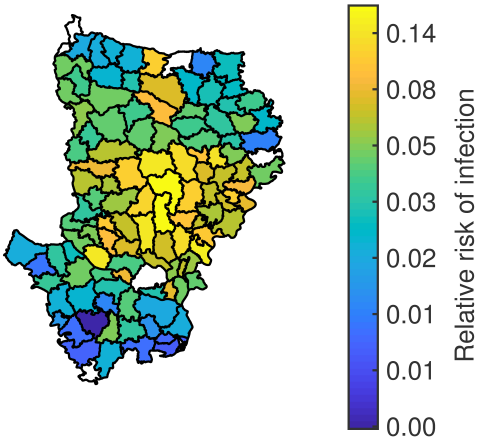

Figure 3.20: Premises relative risk of infection aggregated at the sub-district level for the wave 5 division model. Aggregated probabilities were evaluated by taking the mean value over all individual premises situated in the sub-district. Lighter colours indicate a higher probability of infection.

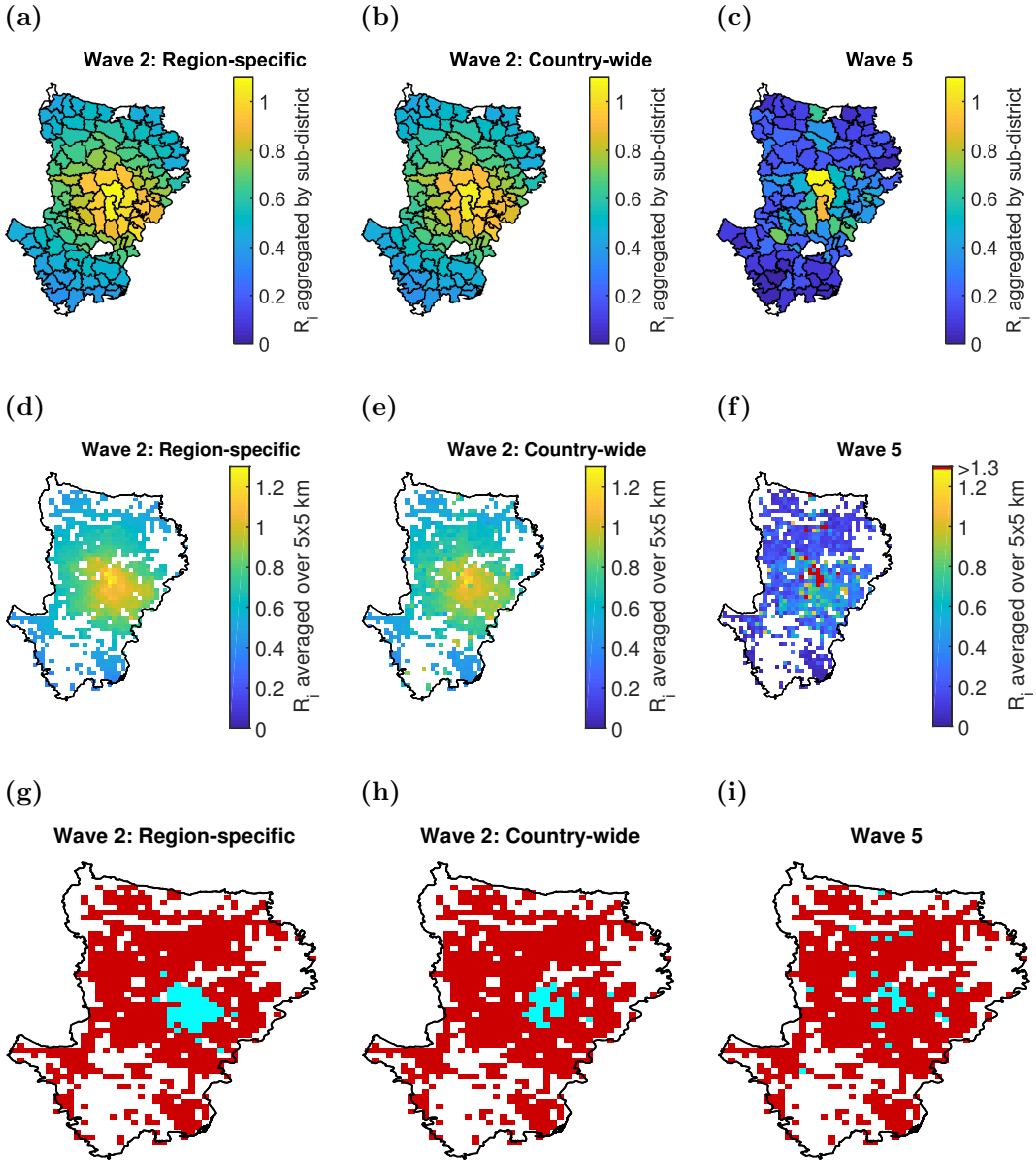

Figure 3.21: Aggregated premises-level reproductive ratios $\mathbf{R_i}$ for the division-level models. Aggregated values were evaluated by taking the mean R_i over all individual premises situated in the region of interest, with lighter colours indicating a higher average premises-level reproductive ratio. First column used the wave 2 model fitted to region-specific epidemic dates; second column used the wave 2 model fitted to country-wide epidemic dates; final column used the wave 5 model. **(a-c)** R_i aggregated by sub-district; **(d-f)** R_i aggregated into 5km $\times$ 5km regions. In **(f)** red denotes regions with an aggregated R_i above 1.3; **(g-i)** binary indicator of 5$\times$5km aggregated premises-level reproductive ratios R_i being greater than one, red denoting regions with $R_i < 1$, cyan regions with $R_i \geq 1$. For wave 2, the areas capable of continuing a transmission chain (where $R_i \geq 1$) were concentrated in a single central region. The wave 5 model had a smaller central region with $R_i \geq 1$, but indicated extra sporadic areas in the north and south-west with the capability of continuing the chain of transmission if infection arose there.

3.4 Discussion

This analysis illustrates how altering the spatial scale of interest can revise the factors meriting inclusion in mathematical models of H5N1 HPAI influenza transmission among poultry. For Bangladesh, a preferred model framework was identifiable at the division level, with the nonlinear farm size model (model C) chosen. This implies that fitting a transmission kernel, rather than using a kernel estimated from case data, and allowing for non-linear dependencies in both infecting and susceptible premises population sizes are important inclusions. In contrast, at the district level a preferred model could not be established, suggesting the data were not sufficient to determine the key aspects of a district-level model for general use.

Finding that the simple ecological model (model E) was strongly preferred to the full ecological model (model D) for the majority of our wave and spatial level combinations corroborates previous studies, which found ducks and rice cropping systems were not strongly associated with H5N1 HPAI infection risk in Bangladesh [129, 130]. However, the inclusion of extra parameters (relative to models A-C) was penalised in the DIC calculation, resulting in the simple ecological model (model E) not being considered as the 'best-fitting' of our candidate models for the majority of our datasets. Nonetheless, other ecological covariate dependencies besides linear could have been chosen. Modified conclusions may be drawn with these alternative choices.

By fitting models at both district and division levels we could uncover model characteristics that were independent of spatial scale. Demonstrated by a reporting delay of seven days being persistently selected across the entire nested model range, this is noteworthy in indicating non-optimal reporting of infected premises during these poultry epidemics. Furthermore, finding an approximately linear relationship between increasing flock size and premises-level susceptibility highlights a potential detection bias (as an alternative to the natural interpretation of larger flock sizes having increased risk of exposure), with outbreaks more likely to be reported by large premises. This is in agreement with Osmani *et al.* [128], who hypothesised poor disease detection and reporting within endemically infected regions of Bangladesh as plausible reasons for genetically identical viruses seemingly causing independent outbreaks over moderate time periods (< 14 days). Such behaviour is conceivable due to the mortality rate within a flock in the early course of HPAI H5N1 infection being low, meaning detection of such clinical events may be delayed [142].

Detection delays are further compounded by many producers being wary following past experiences with government veterinary services, especially those that carried out mass culling or offered poor compensation for poultry destroyed, further extending the time from initial premises infection to reporting [143]. This hinders intervention efforts, with pre-existing strategies to combat H5N1 infection of poultry having a limited impact. For example, a recent H5N1 surveillance study in Bangladesh poultry found no significant difference in anti-H5 seropositivity between vaccinated and unvaccinated chickens, indicating a failure of the vaccination program and a need for updated poultry vaccines [154]. Moreover, a practice of weekly rest days at LBMs that started in April 2012 and the introduction of improved hygienic measures from FAO did not seemingly impact the relative risk of H5N1 circulation in LBMs [155]. To ensure future policy recommendations are well informed, the quantitative evaluation of proposed intervention strategies to reduce the zoonotic transmission risk of influenza warrants further study. This can encompass both traditional methods (culling, vaccination, targeted surveillance) and innovative direct interruption strategies, such as intermittent government purchase plans (so that farms can be poultry-free for a short time and undergo disinfection) or restrictions on species composition (to synchronise flocks to the same birth-to-market schedule and allow for disinfection between flocks). We refer the reader to chapter 4 for our own analysis of the impact of interventions carried out within the Dhaka division on prohibiting widespread outbreaks of H5N1.

Inspecting the inferred parameter distributions for our division-level models revealed an apparent contrast in transmission dynamics across epidemic waves. For wave 2, the fitted transmission kernel exhibited similar values regardless of the distance between premises involved in an infection event (with $\alpha < 0$). On the other hand, the wave 5 data-informed model gave a stronger preference towards short-range transmission ($\alpha \approx 0$). Further, the force of infection was amplified by increasing the infected premises population size ($q > 1$), providing the rationale for our simulations with this fitted model determining that the regions with the greatest infection risk lay to the north of the observed cases (figure 3.20).

These differing transmission characteristics between epidemic waves may have been the result of either a combination of, or solely, a change in disease dynamics and surveillance sensitivity. During the wave 5 poultry epidemic in 2011 a new clade of H5, 2.3.2.1, was identified in Bangladesh [156, 157]. The introduction of this virus could have altered transmission patterns compared to the viruses circulating

during 2008, including how the force of infection scaled with the premises flock size. Additionally, spatial and temporal changes in surveillance between the two poultry epidemics of interest may have altered the proportion of infected cases actually reported. With the surveillance system mainly relying on passive surveillance, substantial under-reporting of poultry cases may have occurred [143]. However, due to experiencing a number of previous epidemics by 2011 there may have been a reduction in this factor since 2008. A greater proportion of subsequent infections occurring in the local neighbourhood of an already infected premises may therefore have been found in the wave 5 epidemic of 2011, giving extra weight to short-range transmission events relative to the wave 2 fitted model.

Our preferred wave 5 division model revealed the presence of premises with 'super-spreader' potential, where premises-level productive ratios R_i were much larger than one. In total, 25 premises obtained an $R_i > 10$. These had large poultry populations (at least 25,000, with only 33 premises in the entire Dhaka division having populations at or above this level) and were situated in areas with a high concentration of poultry farms. We speculate these conditions enhance the ability of a premises to transmit infection, with epidemics of greater magnitude compared to when these premises do not become infected. A bimodal distribution for infected premises epidemic size would subsequently be expected, in agreement with our simulations (figure 3.17(e)). Crucially, these regions were very localised and the observed cases were predominately absent from them, with $R_i < 1$ for the majority of the division (figure 3.21(i)). The small scale of the wave 5 epidemic may have been a direct result of this. Such behaviour may have occurred as a result of heterogeneity in biosecurity compliance. Industrial premises with larger flock sizes implement clearly defined standard operating procedures for biosecurity [158]. In contrast, smaller scale commercial operations may suffer from having less strict measures, such as village farms frequently being built side by side with little separation, which may promote the spread of H5N1 [126]. This emphasises the importance of maintaining compliance of biosecurity regulations, preventing premises with super-spreader potential becoming infected, echoing conclusions drawn by the FAO who stated there was a strong need to improve biosecurity in commercial and government poultry farms in Bangladesh [158].

At the district level, premises-level reproductive ratios suggested that, in principle, chains of transmission from premises-to-premises would not be sustained. Thus, importations and transmission from other sources appeared to be vital contributors

to the poultry outbreak size. Ultimately, this culminates in a low risk of infection across the entire region. This may be a consequence of the poultry value chain, with commercial poultry farms sourcing day-old chicks from a limited set of parent stock farms and grand parent farms [158]. In other words, there is a risk of disease transfer from grand parent or parent farms to the producers, rather than via a chain of transmission occurring between-premises within the district itself. An issue to highlight is the potential for cases caused by premises that lay outside the district (so effectively imported in). While these may in fact have been a short-range transmission event from premises just outside the district, if there were infected premises in the district that were further away the fitting procedure could give a misleading level of support for long-range transmission. This leads to a further knock-on effect with the spark term value, as true infection importation events should be solely captured by that term.

The zoonotic transmission element of our modelling framework discerned differing causal mechanisms for the reported zoonotic spillover events across waves. Infected poultry is no doubt a baseline causal factor, but such event occurrences may also be influenced by LBM specific risk factors like poor biosecurity and slaughter practices [158]. Due to wave 2 only containing a single human case there was a higher chance it was caused by the latter, with ϵ_h encapsulating such determinants. On the other hand, the human cases within wave 5 had a greater association with the number of infected poultry, a consequence of the reduced chances of them both being caused by alternative determinants. Fitting to the wave 6 human case data, ascertaining that the ϵ_h term dominated the zoonotic transmission dynamics is in agreement with the World Health Organisation reporting LBMs as the source of infection for these specific cases [159].

In terms of minimising human risk, the study presented here suggests merely limiting the size of the poultry outbreak may not in isolation reduce the risk of spillover transmission, while reducing contact between humans and poultry would be prudent. Yet, we note that due to the low number of confirmed human cases we cannot attain strong evidence for these conclusions as they are supported by very few events. Furthermore, there are likely to be inherent biases in the reporting of human cases. Intensive community surveillance efforts only happen in a few communities, meaning many cases may have been missed [134]. Additional human cases correlating temporally with peaks in poultry infection would strengthen the models preference for human case occurrence being linked to H5N1 prevalence in poultry. In light of

these conditions, further study is required to verify these findings and ascertain their sensitivity to differing levels of under-reporting. Such analysis is becoming feasible through the development of novel methods for fitting models to an unknown number of infections, including fully Bayesian approaches [160, 161].

The quality of data used had some limitations. Firstly, the following inherent reporting biases could exist and have been discussed above: under-reporting can result in the true extent of both poultry and human cases not being known; the likelihood of an infected premises reporting the outbreak may increase the larger the flock size. Second, in the absence of information on the premises notification time for reporting disease we assumed all premises had the same fixed value reporting delay, treating it as an unknown parameter with a set of different plausible values tested. In reality, there is likely to be variability in this value across premises that may influence the estimated transmission parameters. Third, we chose not to include premises locations that had no poultry populations present in the poultry census database. While this should portray the proportion of premises that are between-flocks at any one time, the impact of alternative sets of poultry farms being populated at a given moment, with the effect on risk of zoonotic H5N1 transmission that follows, requires further study. Further, our assumption of culled premises not being restocked while a poultry epidemic was still in progress may not hold for all our datasets, as in reality the restocking period following a cull is three months [162]. Fourth, due to not having data on movements of poultry we were unable to include highly preferential trading links between premises and LBMs explicitly in our analysis. However, previous work concluded that when proposing models for animal movement contact data between holdings, those that included separate distance-dependent and distance-independent terms were preferred to purely distance dependent models [146]. This therefore motivated our model framework including both a fitted distance-dependent transmission kernel and a spark term to seed infections from other sources independent of distance. Finally, factors that have been previously determined to increase risk of H5N1 infection in poultry, such as the number of roads per sub-district and human population density [30], were not incorporated here. Further work focused specifically on these factors should be able to enhance understanding of public accessibility as a H5N1 poultry infection risk.

To address this the modelling framework outlined can be extended in numerous ways. The first would be to treat layer and broiler chickens as distinct types, rather than considering the total poultry population per premises. It could then be as-

certained whether there are type-specific risk factors or, for a specific risk factor, differing levels of risk against each poultry type. Next, as previously discussed, while we have focused on a linear dependence on the presence/absence of a number of covariates and the resulting contribution to overall risk of infection, many other choices for the spark term dependence could be made. For example, if the necessary data were available, a non-distance dependency on LBMs could be used based on total LBM visits from personnel working on the premises of interest. The impact of varying these dependencies merits further investigation. Thirdly, restocking of previously culled premises can be integrated into the poultry transmission model component, while modifications can be made to the zoonotic transmission component to produce a complete spatially dependent model. Implementing these changes would allow information such as human population density and LBM locations to be explicitly incorporated. Another direction for further work is to relax the assumption of every premises having the same delay-time for reporting disease, and to determine the robustness of the modelling framework by applying it to other regions that have recorded H5N1 cases in both poultry and humans.

Overall, with the data available, our findings suggest the key components that should be incorporated within a general division-level framework for H5N1 poultry infection in Bangladesh were identifiable (despite apparent differences in behaviour for each poultry epidemic of interest), while this was not achievable at the district administration level. Across spatial scales we saw a consistent outcome of non-optimal reporting of infected premises, suggesting we should seek procedural improvements that will reduce the notification time of infected poultry premises. Furthermore, our simple zoonotic transmission model capably identified differing significant contributors to spillover transmission from poultry to humans across epidemics. Yet, for H5N1 influenza the dynamic interplay between animal health, environmental factors and the immune system of the human host must be resolved to ensure policy decisions result in the minimisation of zoonotic transmission occurrence. Given these complexities, it is imperative that further work to enhance understanding of influenza transmission dynamics at the human-animal interface is pursued.

Chapter 4

Optimal control strategies for H5N1 outbreaks in Bangladesh

4.1 Introduction

In the previous chapter we outlined and parameterised a H5N1 highly pathogenic avian influenza (HPAI) modelling framework incorporating zoonotic transmission at the poultry-human interface. Resolving the key transmission-dynamic mechanisms and risk factors behind historical disease epidemics can help inform the actions that should be enacted in a future outbreak, with the successful determination of the optimal control response being a key challenge for policy-makers. With this in mind, we will now utilise our previously fitted H5N1 models to perform a simulation study comparing plausible H5N1 HPAI prevention and control strategies.

For the prevention and control of HPAI, the specific intervention actions to be taken with regards to regulating marketing, imposing movement restrictions or quarantine measures, culling and vaccinating vary according to local circumstances and from country to country. There is no one solution for all situations, and a balance must be established among effective, feasible and socially acceptable control measures that safeguard the short-term and long-term livelihoods of farmers and the health of the population.

In general, however, a number of basic measures are common to all circumstances. One such measure is that infected birds and those in contact with them must be humanely and safely culled to halt spread of the disease. This limits spread by

decreasing the amount of virus released from any one site. However, usually this alone cannot completely prevent further spread because some virus will have been released before culling commences, and often before the disease is detected. As a result, pre-emptive culling (the culling of animals before they are found to be infected) can be used to attempt to make this a more proactive measure. Use of widespread pre-emptive culling based on defined areas around an outbreak has been a standard implementation of this protocol [163]. In Bangladesh, case detection and stamping out remain the key platforms of HPAI control programmes [143].

Disease control programs may also aim to create impediments to spread. An essential part of this is to create an environment in which there are relatively few locations that may become easily infected, with vaccination one of the main methods available for achieving such a goal [163]. Vaccination against HPAI aims to reduce levels of virus shed into the environment and stop infection spreading, as well as preventing clinical disease. It has been implemented and encouraged as part of a control program in poultry in parts of Asia, including China and Vietnam. It was found that H5N1 infected premises in Vietnam reported during an outbreak period when only depopulation-based population control was used had higher within-flock reproductive numbers than infected premises reported during an outbreak period where the control policy in place was depopulation plus nationwide systematic vaccination campaigns [164]. Recent positive developments have seen vaccines against H5N1 and H7N9 prevent birds from shedding the virus through their mouths and droppings, thus stopping transmission from one bird to another [165]. Of particular importance is ensuring the vaccines used have high efficacy. A recent H5N1 surveillance study in Bangladesh poultry found no significant difference in anti-H5 seropositivity between vaccinated and unvaccinated chickens, indicating a failure of the vaccination program and a need for updated poultry vaccines [154].

Naturally, policy effectiveness will depend critically on how swiftly clinical cases are diagnosed and the speed with which the chosen control measure can be administered. By employing active surveillance of premises (i.e. activities which are frequent, intensive and aim at establishing the presence or absence of a specific disease) the notification time for identifying clinical symptoms of infection within a flock may reduce.

Although active surveillance activities can be expensive and time-consuming, there are notable examples of the benefits of strengthening influenza surveillance pro-

grams. Improved influenza virus surveillance in pigs revealed that influenza virus transmission from humans to swine is far more frequent than swine-to-human zoonosis [166]. A follow-up study went on to identify multiple previously uncharacterised influenza viruses of human seasonal H1N2 and H3N2 virus origin that have circulated undetected in swine in Brazil for more than a decade [167]. The public availability of genetic sequence data from databases such as GenBank have allowed pioneering studies to come into fruition, setting out to characterise the cross-species nature and the migration of influenza A viruses on a global scale [168]. On top of that, there are probable long-term advantages to be gained from active surveillance to outweigh the costs. In the first instance there are trade benefits, with eventual proof of disease absence allowing the opening-up of hitherto untapped markets. Secondly, for diseases such as rinderpest beginning active surveillance means vaccination could cease, saving sizeable amounts of money that otherwise would have been spent on blanket vaccination campaigns [169].

In the case of these measures successfully reducing the infection level among poultry, an expected knock-on effect should be a lessened chance of humans becoming infected through spillover transmission. Alongside this, a separate group of control actions may aim to further reduce the threat of zoonotic transmission at the poultry-human interface through other means. This may include reducing dangerous contacts through public awareness campaigns, improving (and enforcing) biosecurity regulations (such as introducing periodic rest days for LBMs [155]) and revising slaughter practices.

In this chapter we assess the predicted impact of these assortment of interventions in mitigating the emergence and spread of H5N1 in poultry premises in the Dhaka division of Bangladesh, and in reducing the risk of zoonotic transmission. This exploratory analysis was done via simulations of our fitted H5N1 influenza transmission models (from chapter 3), optimising specific objectives to control the burden and/or duration of H5N1 outbreaks. Our three primary focuses were as follows: (i) how the targeting and implementation of these interventions alters if it is believed transmission is predominately premises-to-premises, versus the scenario where importations and other external factors are also considered; (ii) how the interventions should be prioritised and implemented when having to account for resource availability; (iii) determining the sensitivity of the management actions under consideration to outbreaks with disparate transmission dynamics.

4.2 Methods

4.2.1 Mathematical model for H5N1 transmission

To assess effectiveness of H5N1 HPAI prevention and control actions in the Dhaka division of Bangladesh, via model simulation, we applied an amended version of our previously developed modelling framework for H5N1 influenza transmission. We carried forward our preferred division-level models fitted to region-specific epidemic dates for the wave 2 and wave 5 Bangladesh poultry epidemics. The selected model in both cases was the nonlinear farm size model (model C), with the contribution by infected premises j to the force of infection against a susceptible premises i given by

$$\eta_{ij} = N_{c,i}^{p_c} \times t_c N_{c,j}^{q_c} \times K(d_{ij}).$$

Following the notation defined in chapter 3, $N_{c,i}$ is the total number of chickens recorded as being on premises i, t_c measures the individual chicken transmissibility, d_{ij} is the distance between premises i and j in kilometres, and K is the transmission kernel to capture how the relative likelihood of infection varies with distance. This model also incorporated power law exponents on the susceptible population, p_c, and infected population, q_c.

The transmission kernel K was the pareto distributed form as described for the parametric kernel model in section 3.2.2 (see equation (3.3)). The spark term was the same fixed value for every premises, ϵ, with the total rate of infection against a susceptible premises i on day t satisfying

$$\text{Rate}(i,t) = \left(\sum_{j \in I(t) \cup Rep(t)\, t} \eta_{ij} \right) + \epsilon,$$

Recall that the wave 5 division-level model, compared to the wave 2 fitted model, had a stronger preference for short-range transmission with infecting premises population size also having a more prominent role (see chapter 3). This allowed us to explore the sensitivity of the management actions under consideration to epidemics with disparate transmission dynamics.

With regards to modelling human case occurrence the rate of spillover transmission was unaltered from equation (3.6), taking the following form,

$$\lambda(t) = \beta I_b(t) + \epsilon_h.$$

As before, β is the poultry to human transmission rate, $I_b(t)$ is the number of infected poultry within the region of interest and ϵ_h is a constant human case spark term.

4.2.2 Poultry control policies of interest

The poultry-targeted policy actions compared were ring culling, ring vaccination and active surveillance. There are typically restrictions on the resources available for enforcing such interventions, limiting the number of poultry and/or premises that can be targeted on any given day. As a consequence, we imposed daily capacities on the maximum number of poultry and/or the maximum number of premises targeted by each control action, with three differing levels of severity related to the availability of resources. In each case a baseline control measure of only culling reported premises was performed, with premises being culled on the same day they were reported if possible (with respect to the resource constraints in place). Note that culling of premises reporting infection was carried out in all subsequent control strategies outlined below.

Ring culling

For this choice of action, in addition to the culling of premises reporting infection, all premises within a specified distance of each location with confirmed infection were marked for culling. The distances evaluated here ranged from 1-10km (in 1km increments). The following three settings were considered for comparing the effect of differing resource constraints:

- **Low**: 20,000 daily bird limit, 20 premises limit.
- **Medium**: 50,000 daily bird limit, 50 premises limit.
- **High**: 100,000 daily bird limit, 100 premises limit.

To clarify, those premises reporting infection would be prioritised above all others for culling, ordered by the date of reporting. For those premises designated for ring culling that had not reported infection, the order of priority was determined using a distance-based approach, with resources allocated from the outer edge and moving inwards to the centre (an "outside-to-centre" approach). In other words, following the determination of premises situated within the ring established around a premises reporting infection, distances between all such premises and the infected premises were computed with the premises then culled in descending distance order.

Note that all premises in the ring established around the initially reported infected premises had to be treated before moving on to locations that were contained within rings established around the next set of subsequently reported infected premises.

Ring vaccination

For this choice of action all those premises within a specified distance of each premises reporting infection were listed for vaccination. As for ring culling, the ring radii sizes evaluated ranged from 1-10km (in 1km increments). Vaccine efficacy was set to 70%, with a seven day effectiveness delay to account for the time required for suitable immune protection to develop after the vaccine is administered. With the epidemiological unit of interest being the individual poultry premises, we assumed poultry within the same flock had the same vaccination status (i.e. post-vaccination the vaccine either successfully induced immunity for all birds within the flock, or the flock remained unprotected).

As the vaccination strategies considered here also involved the culling of reported premises, we had to make an assumption regarding how these two aspects should be factored into the resource limits. We were informed that while culling would be carried out by DLS (Department of Livestock Services) staff, vaccines would be administered by the farms themselves in supervision by DLS staff (personal communication from Madhur S. Dhingra). Therefore, we treated these activities as being independent of each other, assigning separate resource limitations to each control action.

The capacity levels that were considered, with culling and vaccination treated independently, were:

- **Low**: 20,000 daily bird limit, 20 premises limit.
- **Medium**: 50,000 daily bird limit, 50 premises limit.
- **High**: 100,000 daily bird limit, 100 premises limit.

There was no limit on the cumulative number of vaccine doses available. An outside-to-centre resource allocation prioritisation approach was used for vaccination, matching the ring culling prioritisation procedure.

Active Surveillance

A total of four active surveillance strategies were compared. These could be grouped into two distinct types of implementation. For all model simulations of these initiatives those premises designated as undergoing active surveillance had their notification delay time reduced from seven days to two days. A reduction to two days was chosen, and not a larger reduction to a single day or the complete removal of the reporting delay, to account for the fact a flock can be infectious before clinical signs of H5N1 infection are observed, which may not be immediate even when active surveillance procedures are in place [142]. Note that there were no other control actions in place beyond this and the culling of flocks at premises reporting infection (which abided by the previously discussed capacity limitations).

The first two surveillance strategies we consider are reactive in nature. This involved premises within a given distance of premises reporting infection undergoing active surveillance. We imposed a limit on the number of premises that could be monitored. When resource thresholds were exceeded only those premises deemed to be of higher priority underwent active surveillance, with the following two prioritisation strategies studied: (i) "reactive by distance", with premises ordered by distance to the focal premises, nearest first (i.e. inside-to-out approach); (ii) "reactive by population", with premises ordered in descending flock size. For these schemes the ring size for active surveillance was set to be 500m, under the following coverage levels:

- **Low**: 25 premises per outbreak
- **Medium**: 50 premises per outbreak
- **High**: 100 premises per outbreak

The next two surveillance strategies are proactive approaches, with a specified proportion of premises within the Dhaka division selected by some designated criteria to undergo constant active surveillance. The two criteria evaluated here were: (i) "proactive by population", ranking all premises in descending flock size order, (ii) "proactive by density", for each premises we computed the total number of other premises within a distance of 500m, with all premises then ranked in descending order. The coverage levels considered were:

- **Low**: 5% coverage
- **Medium**: 10% coverage
- **High**: 25% coverage

4.2.3 Targeting zoonotic spillover

The actions described in section 4.2.2 intend to minimise poultry infection. While this should cause a reduction in human risk of infection through spillover transmission from the poultry reservoir, a separate group of control actions may aim to reduce the risk of zoonotic transmission at the poultry-human interface through other means (e.g. periodic rest days for LBMs to reduce number of dangerous contacts). With our modelling framework this could be incorporated through modifying the human case spark term ϵ_h. The proportional reductions of the complete ϵ_h values tested were 50%, 75% and 100%, which were considered in the absence of any controls applied directly to poultry. We assessed how these alterations of various magnitudes impacted the expected number of days with human case occurrences.

4.2.4 Simulation outline

As in chapter 3, we utilised commercial poultry premises location and flock size information gathered via a census undertaken by the Bangladesh office of the Food and Agriculture Organisation of the United Nations (FAO/UN) in 2010 (see section 3.2.1 for further details on the dataset). We ran simulations of our transmission models using premises located within the Dhaka division that had chickens (layer or broiler) recorded as being present in the poultry census, giving 13,330 premises in total.

In contrast to the simulation procedure employed in chapter 3, here we used the Sellke construction (algorithm 6). This method allows for improved comparison of interventions, as in this framework the inherent randomness of an epidemic realisation can be encoded at the beginning of the simulation. Once calculated, the resultant epidemic can be constructed from the deterministic solution of the infection process and removal (i.e. culling) times (for additional information see section 2.5).

Dependency of optimal control policy on outbreak origin

For this series of simulations we were interested in elucidating the severity of control actions necessary to minimise epidemic severity based on the district an outbreak originated in, plus how this differed between the two fitted models with their contrasting poultry-to-poultry transmission dynamics. To be able to ascertain the true impact of outbreak origin on the epidemic outcomes of interest we assumed premises infection was predominately driven by premises-to-premises transmission, with no cases arising due to external factors. As a consequence, in all runs the background spark term ϵ was set to zero, while an initial cluster of three infected premises was

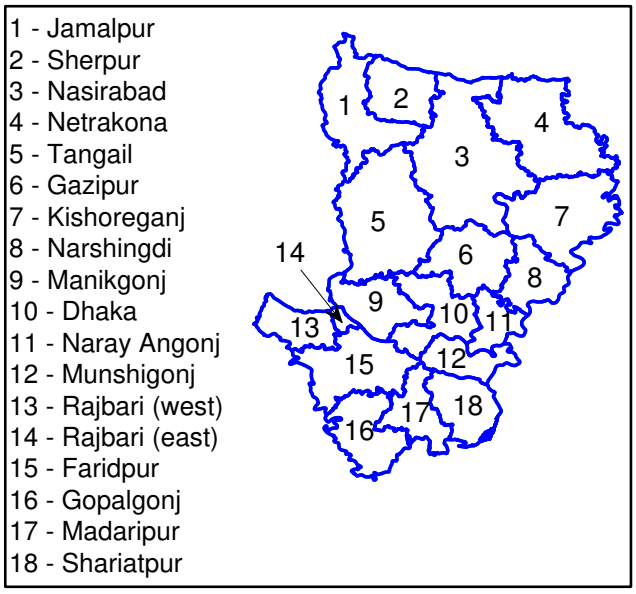

Figure 4.1: Locator map naming each district that is contained within the Dhaka division.

seeded in one of the 18 districts situated within the division (figure 4.1).

For each culling, vaccination and active surveillance management action we performed 1,000 simulation runs using the wave 2 fitted transmission model and at least 300 simulation runs with the wave 5 fitted transmission model. A consistent set of distinct sampled parameter values (obtained previously via MCMC) and initial seed infection locations were used across these runs to aid intervention comparisons. The particular control objectives of interest here were focused on either reducing the expected length of an outbreak, or minimising the likelihood of an outbreak becoming widespread. To this end, the summary outputs analysed for this scenario were as follows: (i) mean outbreak duration, (ii) probability of an epidemic (where we subjectively define an outbreak as an epidemic if there are cases in five or more districts, with the total number of cases exceeding 15).

Optimal control policy in presence of external factors

Our second scenario of interest was determining the optimal control strategy when an outbreak is ongoing and infection may arise anywhere within the division, in addition to premises-to-premises transmission dynamics. With regards to this objective these simulations did incorporate the background spark term ϵ, with a single initial infected premises placed anywhere in the division.

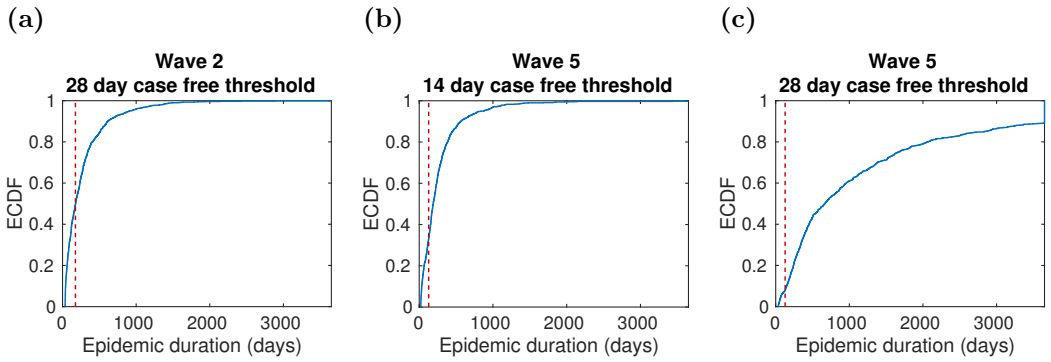

Figure 4.2: ECDF for epidemic duration from simulations of the specified transmission model, with the given number of consecutive infection-free days required for an outbreak to be deemed as completed. All simulations used infected premises culling only (no additional controls were in place), with reporting to culling times weighted by the empirical probability mass function. The following ECDFs were constructed using 1,000 simulated realisations: **(a)** Wave 2, 28 day threshold value; **(b)** wave 5, 14 day threshold value; **(c)** wave 5, 28 day threshold value. The threshold values for number of infection-free days signifying the end of an outbreak were subsequently set to 28 days and 14 days for runs with the wave 2 and 5 fitted models respectively.

We stipulated a simulated outbreak to be complete once a specified number of consecutive infection-free days had occurred. For the wave 2 fitted model, 28 days gave a simulated median epidemic length (using infected premises culling only, with culling times weighted by the empirical probability mass function) that corresponded well with the data (figure 4.2(a)). On the other hand, a 14 day case-free period was more suitable for the wave 5 fitted model (figure 4.2(b)), with runs using the 28 day infection-free condition giving, in general, longer outbreak periods than the observed data (figure 4.2(c)). As a consequence, the infection-free condition values were set to 28 days and 14 days for runs with the wave 2 and 5 fitted models respectively.

For each poultry-targeted management action we performed 1,000 simulation runs with the wave 2 fitted transmission model and at least 300 simulation runs with the wave 5 fitted transmission model. For the zoonotic spillover tests 1,000 simulation runs were performed with both transmission models. To aid intervention comparisons across the runs we again used a consistent set of sampled parameter values and initial seed infection locations. The control objectives of interest in this scenario were again focused on outbreak length and size. In particular, either increasing the chance of an outbreak being short, maximising the likelihood of an outbreak re-

maining below a specified size, or minimising the number of poultry destroyed as a result of culling. The particular summary statistics that we therefore chose for these control objectives were as follows: (i) outbreak duration t being 90 days or less, (ii) outbreak size I not exceeding 25 premises, (iii) mean number of poultry culled.

4.3 Results

4.3.1 Dependency of optimal control policy on outbreak origin

Culling and vaccination

To tackle outbreaks with wave 2 type transmission dynamics, the broad relationship seen for outbreaks originating in any district was an increase in the suggested size of the ring culling zone as capacity limitations were eased (figure 4.3). This held when the intervention of concern was ring vaccination (figure 4.4), while also having dissimilar outcomes compared to the ring culling findings. For example, when the control objective was to reduce the expected outbreak duration the suggested radius size for vaccination exceeded that for culling (figures 4.3(a) and 4.4(a)). Further, for minimising the probability of an epidemic the most suitable action, under certain capacity conditions, against outbreaks whose source districts were Gazipur (low capacity) or Gopalgonj (low and medium capacities) was to carry out culling of infected premises only, with no vaccination of surrounding premises (figure 4.4(b), for a district locator map see figure 4.1). This opposes conclusions drawn when an additional ring cull around infected premises was investigated, with culling confined to infected premises alone never considered to be the optimum approach in those circumstances (figure 4.3). As a cautionary note, variations in these control metric outputs across the suite of ring sizes tested were relatively minor (see appendix B.2).

These trends continued using the wave 5 transmission model. Large disparities between recommended ring culling and vaccination sizes were seen, even if the control objective and capacity setting were the same. Focusing on a specific example, we compared culling against vaccination in a high capacity setting with the intent of minimising epidemic risk. For ring culling, outbreaks emerging in central and northern districts typically required upper radius values of 7km or 8km, while the western district of Rajbari (east) required the 10km upper limit of the range of values explored here. In the event of an outbreak beginning in one of the remaining districts only localised ring culling of 1km or 2km was suggested, though we observe a ring cull of some form was always found to be preferred to only culling infected

premises (figure 4.5(a)). On the other hand, ring vaccination did not improve on merely culling infected premises for outbreaks beginning in northern and southern districts, while central districts typically only required a coverage radius of 5km or less (figure 4.5(b)). However, sensitivity analysis of the variations in epidemic probability against intervention severity (for outbreaks beginning in a given district) revealed these to be small, especially under vaccination measures (figure 4.5(c)).

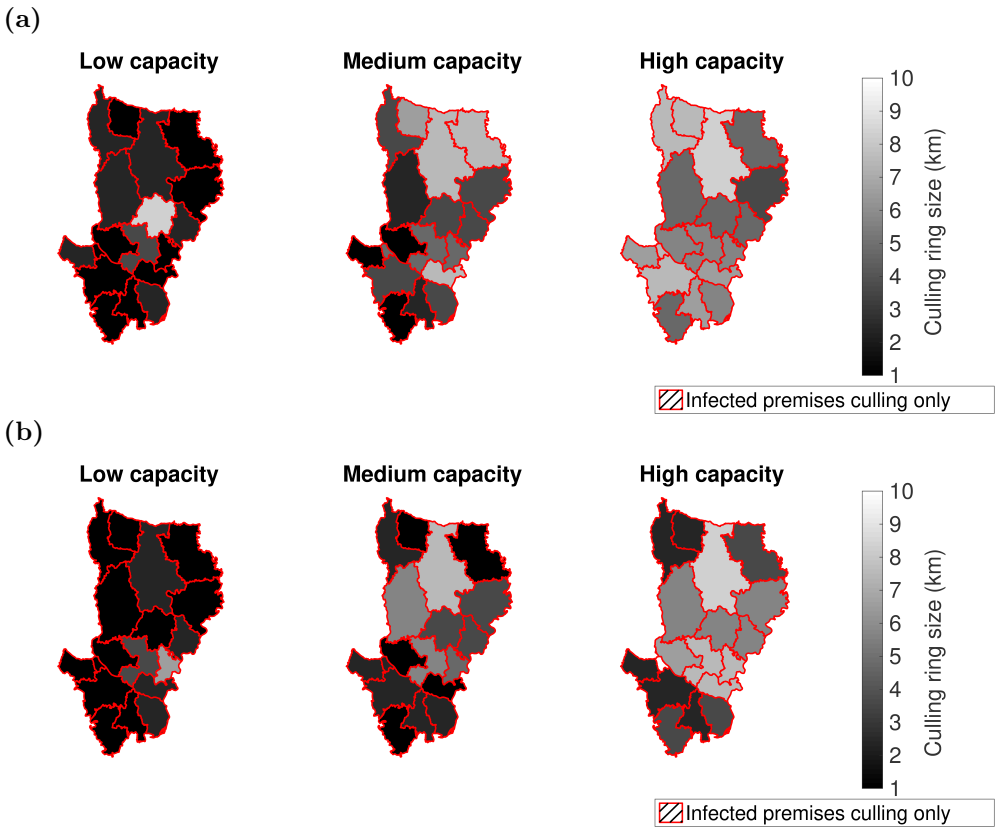

Figure 4.3: Maps displaying the ring culling range to optimise specified control objectives with respect to district of outbreak origin and control capacity level. All runs were with the wave 2 transmission model. For each combination of control capacity level, district of outbreak origin and control objective 1,000 simulation runs were performed. Hatching of a district indicates the preferred strategy was culling infected premises only, while solid shading corresponds to the ring culling size determined as the optimal response against outbreaks that originally emerged in that district. Lighter shading corresponds to a larger ring culling region. The control objectives were **(a)** minimising average outbreak duration, and **(b)** minimising the probability of an epidemic. For full results see table B.1, with sensitivity of these control metrics to intervention severity for outbreaks beginning within particular districts presented in figures B.1 and B.2.

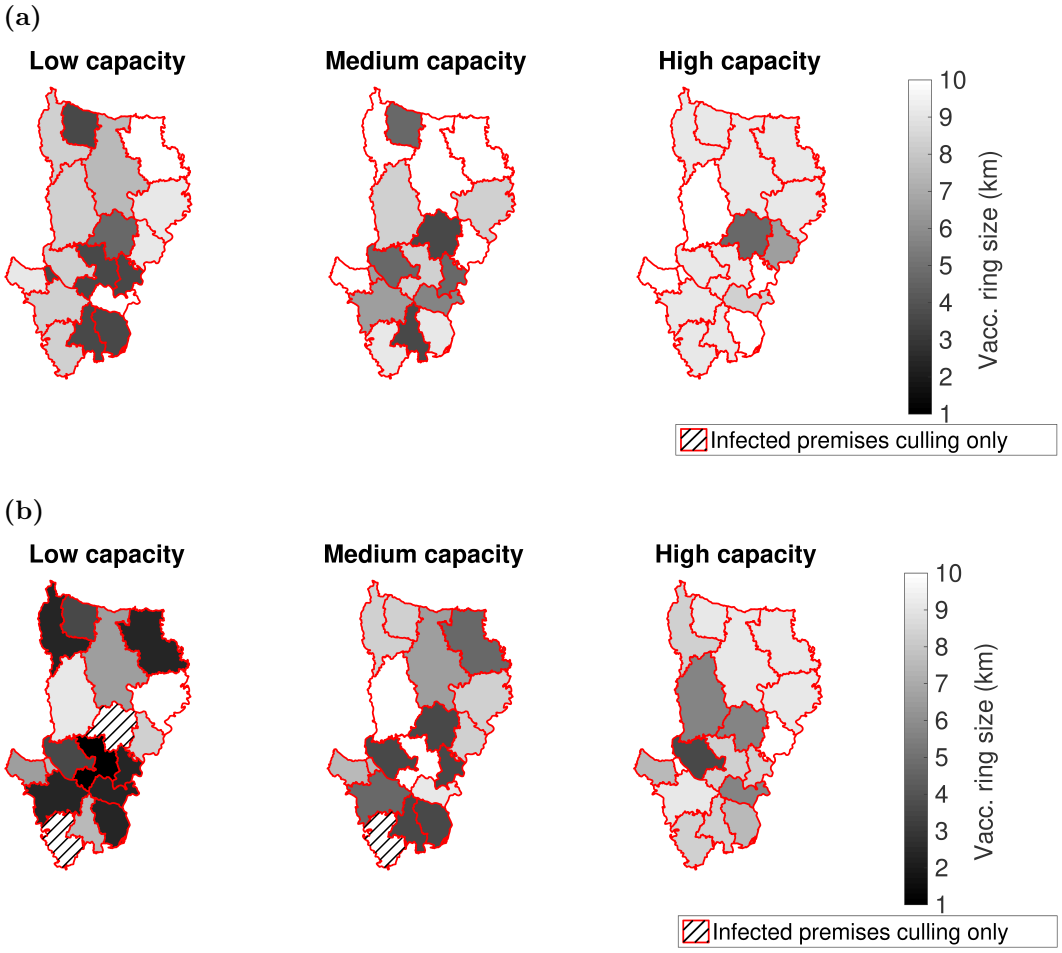

Figure 4.4: Maps displaying the ring vaccination range that optimises the specified control objective with respect to district of outbreak origin and control capacity level. All runs were with the wave 2 transmission model. For each combination of control capacity level, district of outbreak origin and control objective 1,000 simulation runs were performed. Hatching of a district indicates the preferred strategy was culling infected premises only, while solid shading corresponds to the ring vaccination size determined as the optimal response against outbreaks that originally emerged in that district. Lighter shading corresponds to a larger ring culling region. The control objectives were **(a)** minimising average outbreak duration, and **(b)** minimising the probability of an epidemic. For full results see table B.2, with sensitivity of these control metrics to intervention severity for outbreaks beginning within particular districts presented in figures B.1 and B.3.

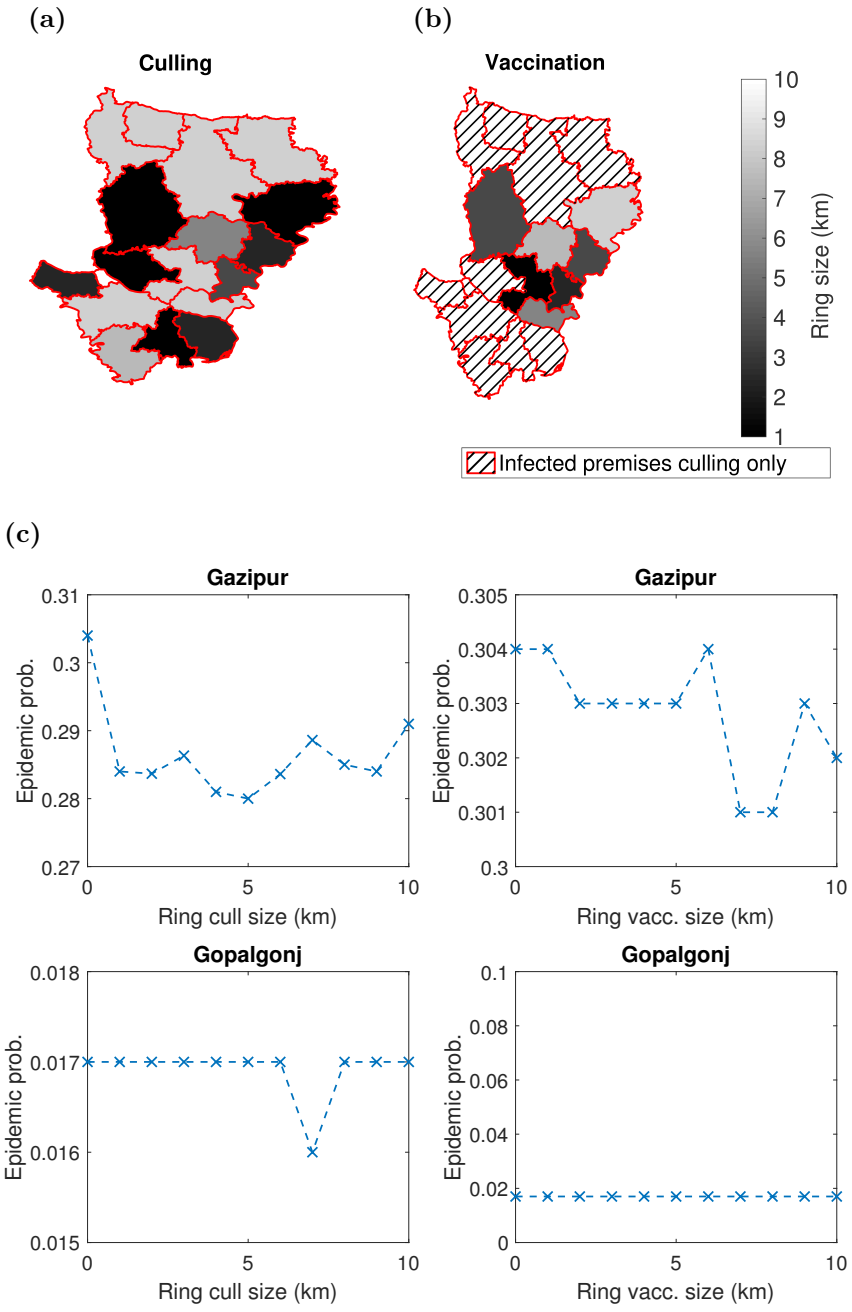

Figure 4.5: Maps displaying optimal ring size to minimise probability of a widespread outbreak for the wave 5 transmission model with a high control capacity. For each combination of intervention method and district of outbreak origin 1,000 simulation runs were performed. Hatching of a district indicates the preferred strategy was culling infected premises only, while solid shading corresponds to the ring size determined as the optimal response against outbreaks that originally emerged in that district. Lighter shading corresponds to a larger intervention region. **(a)** Ring culling; **(b)** ring vaccination. For full results see table B.3. **(c)** Predicted epidemic probabilities against intervention ring size for outbreaks originating in the Gazipur and Gopalgonj districts (for a district locator map see figure 4.1), exposing the minor variations in this control metric across the suite of ring sizes tested. Analogous outcomes were found for outbreaks seeded in the remaining districts.

Active surveillance

In response to outbreaks with either type of transmission dynamics the proactive active surveillance schemes were the top performers across all capacity scenarios and objectives being optimised. Additionally, independent of the source district for the outbreak, higher capacity thresholds usually led to greater reductions in outbreak length and size relative to the scenario where no active surveillance scheme was utilised (figures 4.6 and 4.7).

Concentrating on an outbreak scenario with wave 2 type transmission in conjunction with a control objective targeted at minimising the expected outbreak duration, the chosen active surveillance action differed based on the district of outbreak origin and control capacity conditions. Explicitly, 'proactive by population' was the typical preferred selection, though in a high capacity setting if an outbreak began in the district of Narshingdi then the chosen strategy was 'proactive by density' (figure 4.6(a), for a district locator map see figure 4.1). Switching to a control objective of minimising the spatial spread of the outbreak, the 'proactive by population' strategy was consistently selected as the best choice and reduced tremendously the risk of an epidemic across all capacity settings (figure 4.6(b)).

Similar outcomes were obtained for outbreaks with wave 5 type transmission dynamics where, irrespective of the district where the outbreak originated, the 'proactive by population' strategy was always selected as the optimal action (figure 4.7).

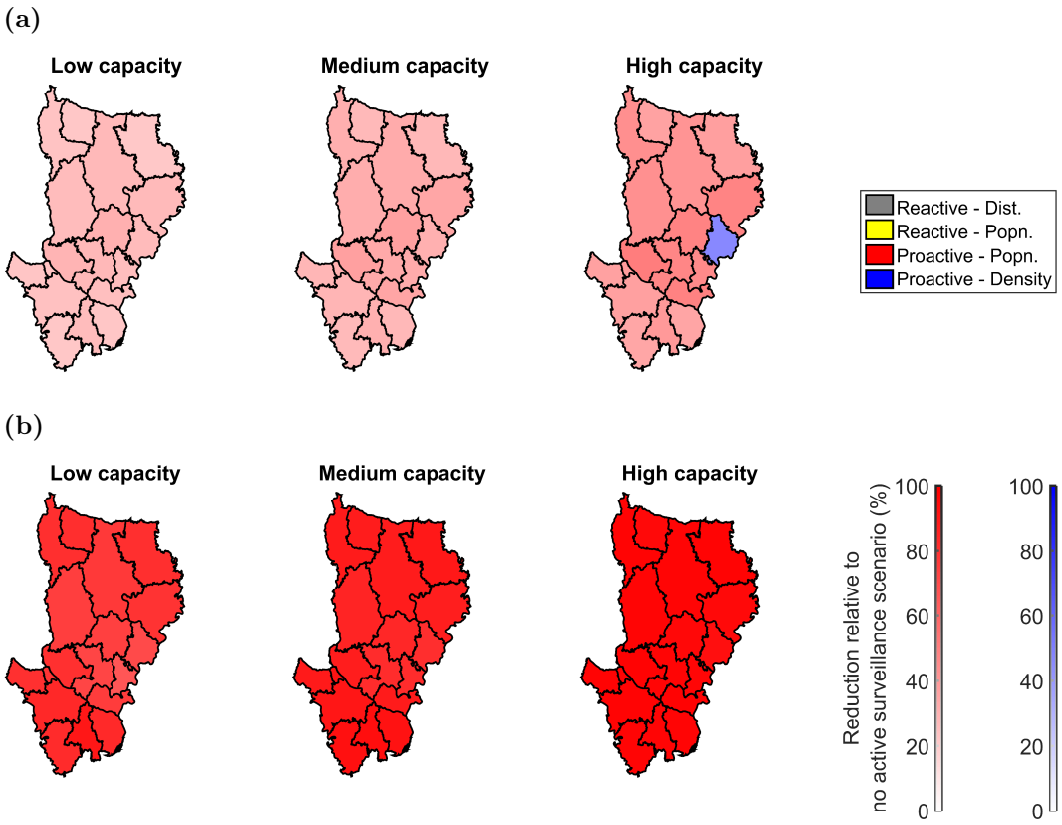

Figure 4.6: Maps displaying the preferred active surveillance strategy to optimise control objectives with respect to district of outbreak origin and capacity setting, for outbreaks with wave 2 type transmission dynamics. For each combination of active surveillance method and district of outbreak origin 1,000 simulation runs were performed. District colour corresponds to the active surveillance strategy determined to be optimal for countering outbreaks originating from that district (grey - 'reactive by distance', yellow - 'reactive by population', red - 'proactive by population', blue - 'proactive by density'). Transparency coincides with the reduction in the objective metric relative to the scenario where no active surveillance was utilised, with completely transparent corresponding to a 0% reduction (no improvement) and completely opaque corresponding to a 100% reduction. **(a)** Minimising average outbreak duration control objective - 'proactive by population' scheme was generally preferred, although we found discrepancies in the best scheme dependent upon the control capacity setting. **(b)** Minimising the probability of an epidemic control objective - 'proactive by population' scheme was found to be preferred in all cases when optimising for this aim. For full results see tables B.4 and B.5.

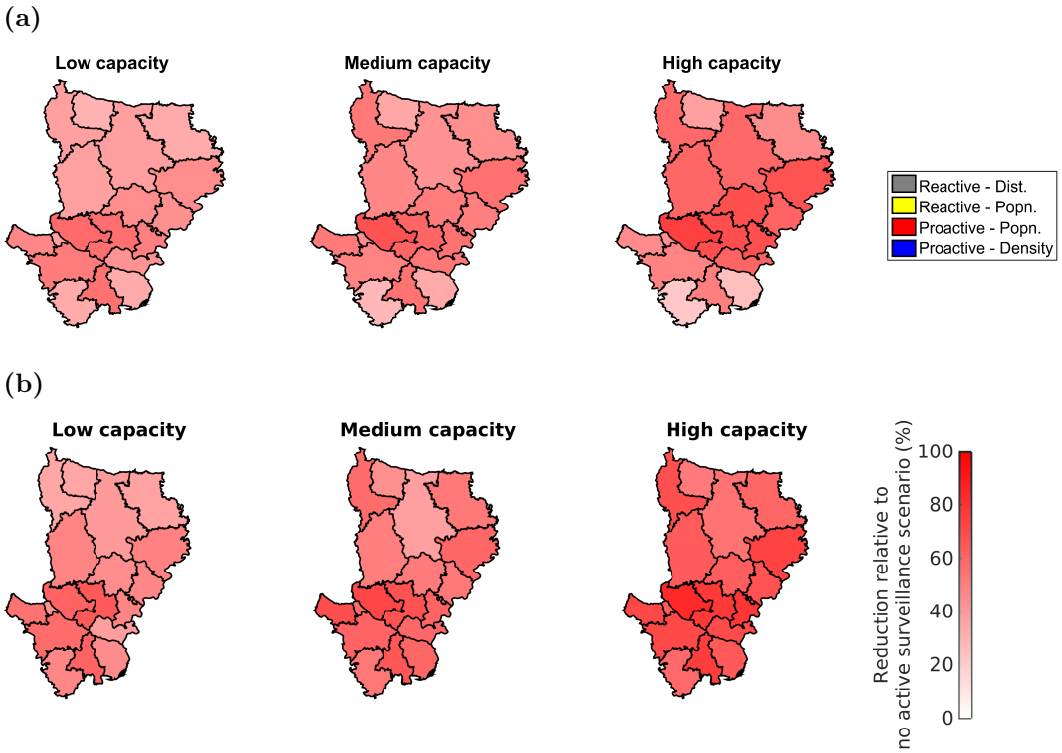

Figure 4.7: Maps displaying the preferred active surveillance strategy to optimise control objectives with respect to district of outbreak origin and capacity setting, for outbreaks with wave 5 type transmission dynamics. For each combination of active surveillance method and district of outbreak origin a minimum of 300 simulation runs were performed. District colour corresponds to the active surveillance strategy determined to be optimal for countering outbreaks originating from that district (grey - 'reactive by distance', yellow - 'reactive by population', red - 'proactive by population', blue - 'proactive by density'). Transparency coincides with the reduction in the objective metric relative to the scenario where no active surveillance was utilised, with completely transparent corresponding to a 0% reduction (no improvement) and completely opaque corresponding to a 100% reduction. The 'proactive by population' scheme was found to be preferred in all cases when optimising either control objective. **(a)** Minimising average outbreak duration, **(b)** minimising the probability of an epidemic. For full results see tables B.6 and B.7.

4.3.2 Optimal control policy in presence of external factors

Culling and vaccination

Recall that we assumed vaccination efficacy was only 70%, in combination with it taking seven days post-administration to become effective. With pre-emptive culling at a premises definitely preventing future infection at that site, we would expect ring culling to bring about slightly greater benefits than ring vaccination with regards to outbreak duration and size. This is indeed what we observed across both transmission models and all three control capacity scenarios (figures 4.8 and 4.9). Likewise, when the goal was linked to minimising the number of poultry culled, the preferential strategy would be vaccination with no additional culling beyond infected premises (figure 4.10).

Considering next the three capacity levels studied we found, perhaps unsurprisingly, evidence of a performance hierarchy. For any given ring size a high capacity allowance generally outperformed a medium capacity allowance, which in turn outperformed a low capacity allowance. As a notable exception to this, for the wave 5 transmission model together with the outbreak duration control aim of $t \leq 90$ we found that using a 1km ring culling intervention caused the standard performance ranking to be reversed (figure 4.8(c)). Furthermore, wanting to minimise the mean number of poultry culled in a wave 2 type outbreak with a ring culling intervention also caused the standard performance ranking to be flipped (figure 4.10(a)).

Focusing on the control objectives associated with outbreak length and magnitude, we note that the conclusions drawn between our models with differing transmission behaviour were qualitatively similar. Under high control capacity resource availability each incremental increase in the radius size generally led to modest improvements in the summary output of interest, at least up to the 10km upper limit in place here. A similar effect was also seen for a medium capacity setting in conjunction with the case size $I \leq 25$ summary statistic (figure 4.9). However, this was generally not the case under both low and medium capacity thresholds. For these situations the optimal radius size varied, depending on the control objective to be optimised and on whether the intervention implemented was ring culling or ring vaccination (figures 4.8 and 4.9).

Optimising the expected number of poultry culled control objective led to contrasting findings being identified between the culling and vaccination measures. Under a

ring vaccination management action, incremental increases in radius size under each set of control capacity conditions were again found to cause modest improvements with regards to the objective, with a 9km or 10km ring selected across all capacities and both transmission models. However, if pursuing a ring culling strategy in combination with this control objective, either no culling beyond infected premises or a ring cull of 1km were deemed optimal. Furthermore, in this case the capacity thresholds deemed most effective were greatly influenced by the disease transmission dynamics (figure 4.10).

Active surveillance

A collection of common trends were obtained across the three control objectives (outbreak duration being 90 days or less, outbreak size not exceeding 25 premises, minimising mean number of poultry culled) and two disease transmission models analysed.

Irrespective of the objective being scrutinised, for each transmission model the 'proactive by population' scheme was found to be the best performing of the strategies under consideration. It was the most effective in increasing the probability of an outbreak not lasting above 90 days (figure 4.11(a)), maximising the probability of an outbreak not consisting of more than 25 infected premises (figure 4.11(b)), and minimising the number of poultry culled (figure 4.11(c)). In addition, increased availability of resources for control led to greater gains. For example, under the wave 2 transmission model the probability of the outbreak duration $t \leq 90$ rose from 0.55 (low capacity) to 0.61 (high capacity), whereas with no active surveillance in use the probability was only 0.51. For case size $I \leq 25$, while with no active surveillance this probability was 0.69, it rose from 0.79 (low capacity) to 0.87 (high capacity). In terms of minimising the mean number of poultry culled in an outbreak, with no active surveillance the average was 3.0×10^5 birds, which dropped to 1.8×10^5 for low capacity conditions, and then to 1.1×10^5 with high capacity conditions enforced. This represented approximate 33% and 66% improvements respectively in minimising the total number of poultry culled. Such effects were even more stark for the wave 5 transmission model, with outbreaks being more likely to take off and having superior longevity under these dynamics. This is highlighted by the outbreak duration $t \leq 90$ being 0.38 when no active surveillance was used, rising to 0.46 for low capacity levels, and reaching 0.58 under high capacity conditions (an improvement of approximately 50% over having no control, and roughly 25% over the low control capacity setting).

Although the 'proactive by density' strategy also had notable improvements with increasing coverage, it was not as effective as a population-based targeting measure. This is exemplified by the discrepancy between the two typically growing with enlarged capacity thresholds. For example, the difference grew from 0.02 (at low capacity) to 0.04 (at high capacity) for $t \leq 90$ using the wave 2 transmission model, and from 0.07 (at low capacity) to 0.11 (at high capacity) for $I \leq 25$ using the wave 5 transmission model. Furthermore, a further drawback of the 'proactive by density' strategy was that under low control capacity levels it struggled to beat either reactive surveillance policy (figure 4.11).

Switching our attention to the two reactive strategies we found there was little to choose between them. Although offering marginal benefits over having no active surveillance at all, they did not bring about noticeable improvements towards the desired goal as the resources utilised increased (figure 4.11). For a full listing of values related to the features raised see tables B.8 to B.10.

The observation of 'proactive by population' outperforming 'proactive by density', and the two reactive strategies only being a slight improvement compared to having no active surveillance, may also be seen when analysing the complete premises case size distributions (figure 4.12).

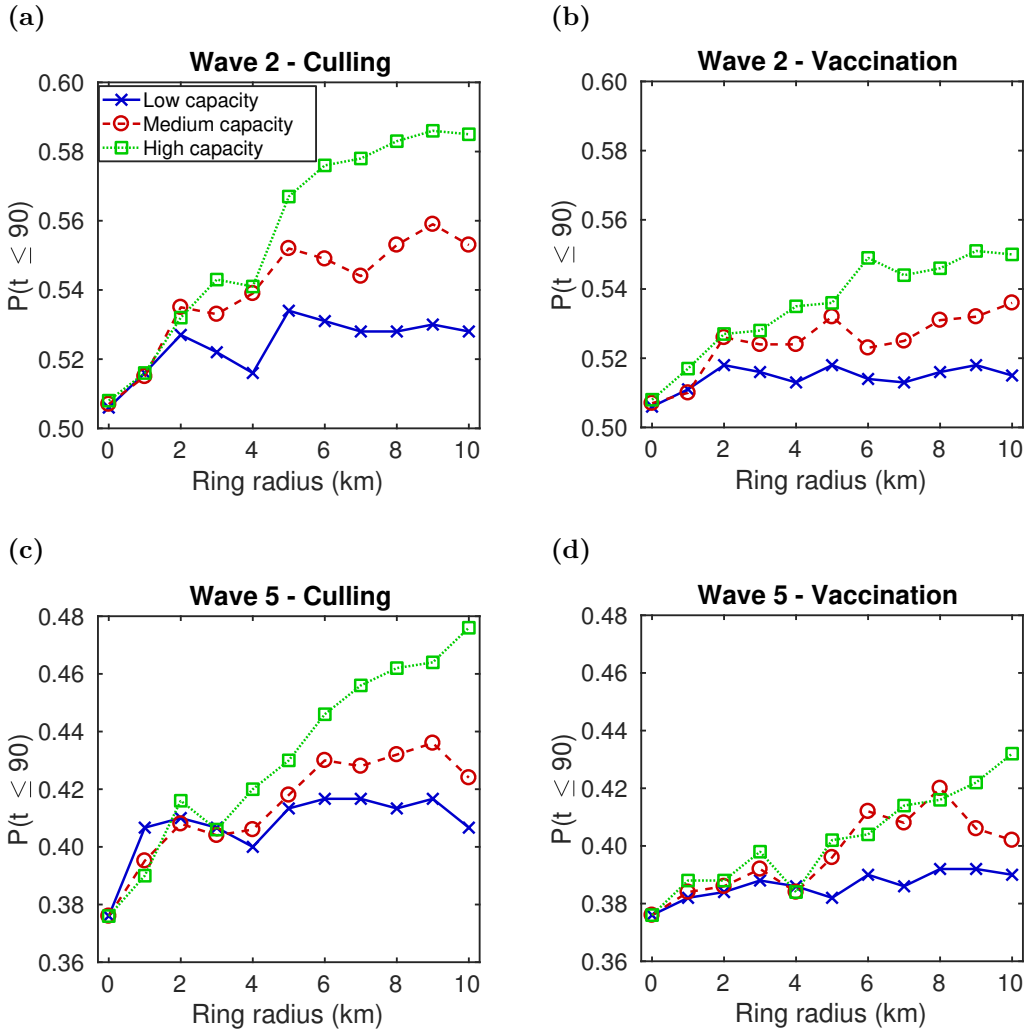

Figure 4.8: Predicted probability of outbreak duration (t) being 90 days or less for different ring culling and vaccination radii. For each transmission model and control method combination, the three capacity settings of interest, low (solid blue line, crosses), medium (dashed red line, circles), and high (dotted green line, squares) displayed disparate behaviour. **(a)** Wave 2 - culling; **(b)** wave 2 - vaccination; **(c)** wave 5 - culling; **(d)** wave 5 - vaccination.

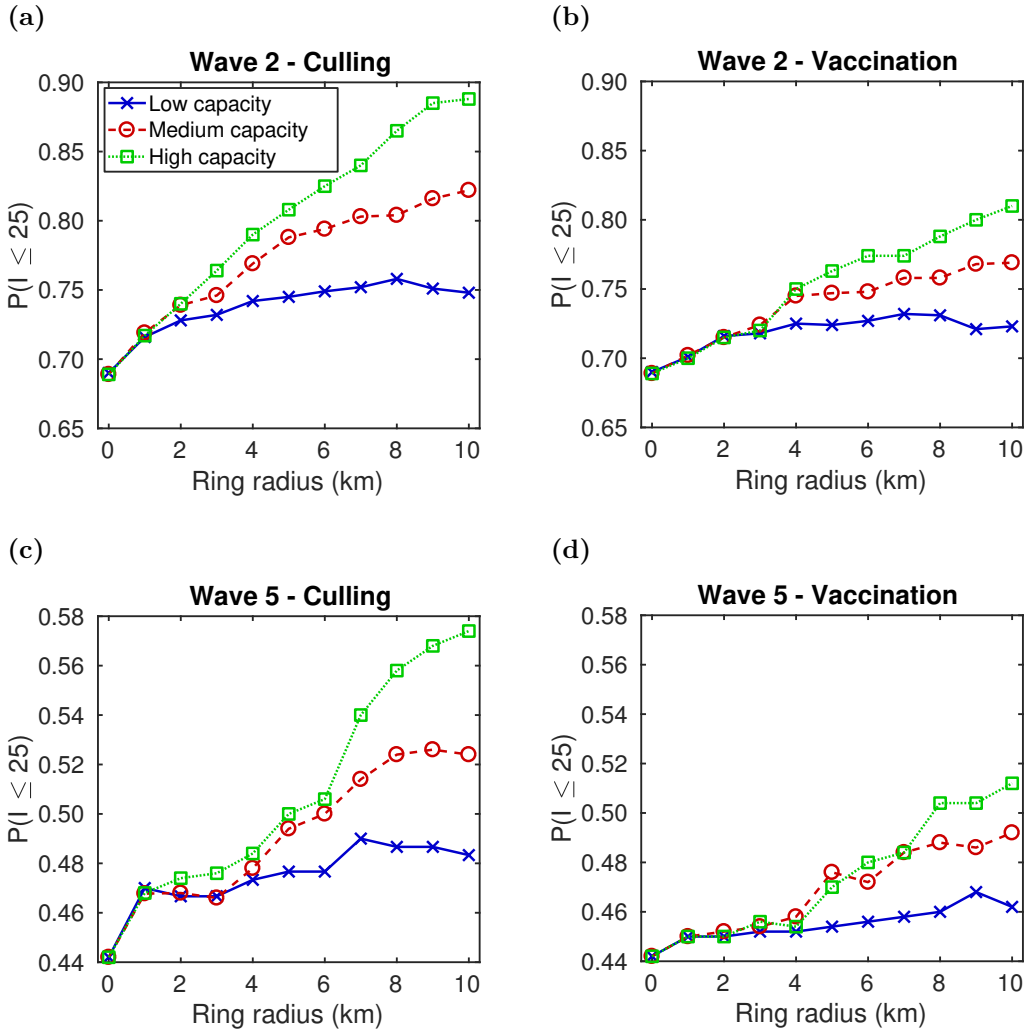

Figure 4.9: Predicted probability of outbreak size (I) not exceeding 25 premises for different ring culling and vaccination radii. For each transmission model and control method combination, the three capacity settings of interest, low (solid blue line, crosses), medium (dashed red line, circles), and high (dotted green line, squares) displayed disparate behaviour. **(a)** Wave 2 - culling; **(b)** wave 2 - vaccination; **(c)** wave 5 - culling; **(d)** wave 5 - vaccination.

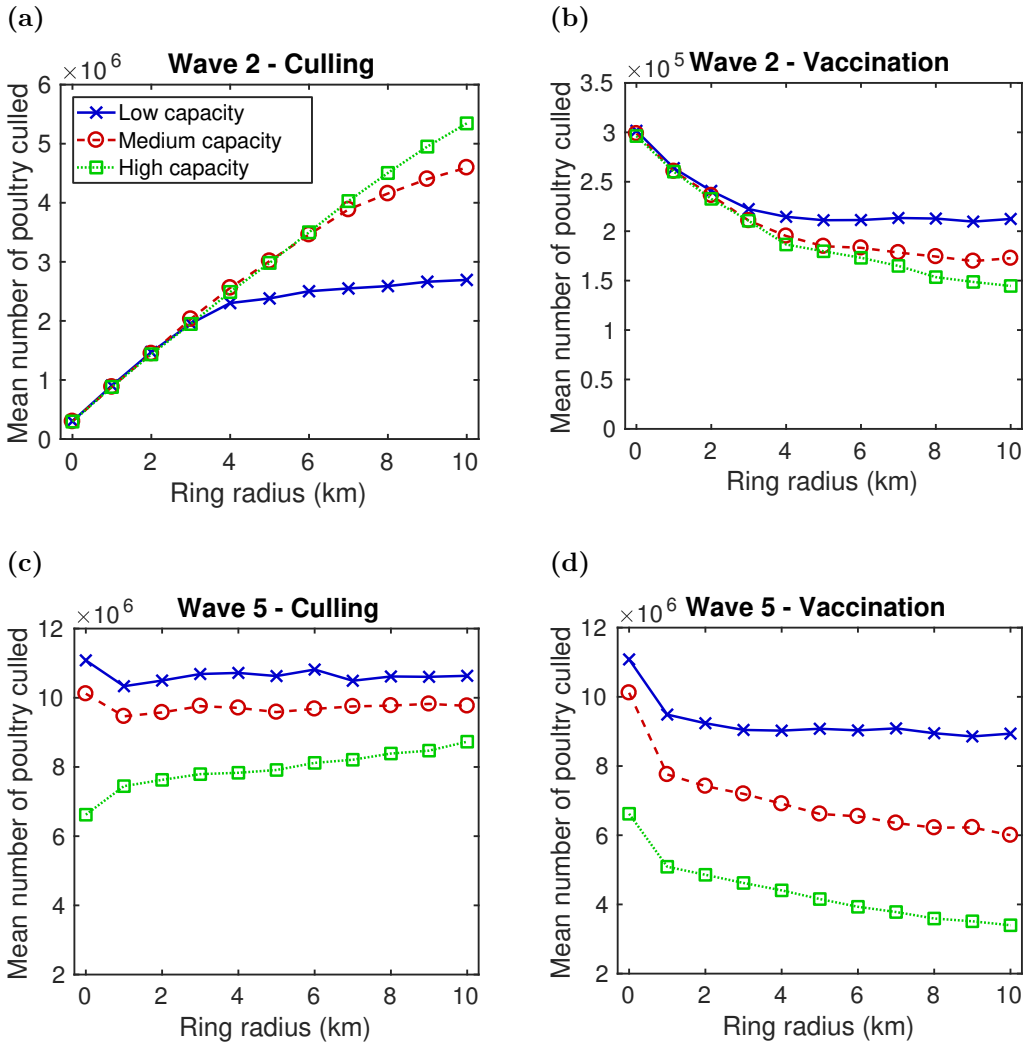

Figure 4.10: Mean number of poultry culled for different ring culling and vaccination radii. The three capacity settings of interest were low (solid blue line, crosses), medium (dashed red line, circles), and high (dotted green line, squares). If pursuing a ring culling strategy, either no culling beyond infected premises or a ring cull of 1km were deemed optimal. For a ring vaccination strategy, a 9km or 10km ring was selected across all capacities. **(a)** Wave 2 - culling; **(b)** wave 2 - vaccination; **(c)** wave 5 - culling; **(d)** wave 5 - vaccination.

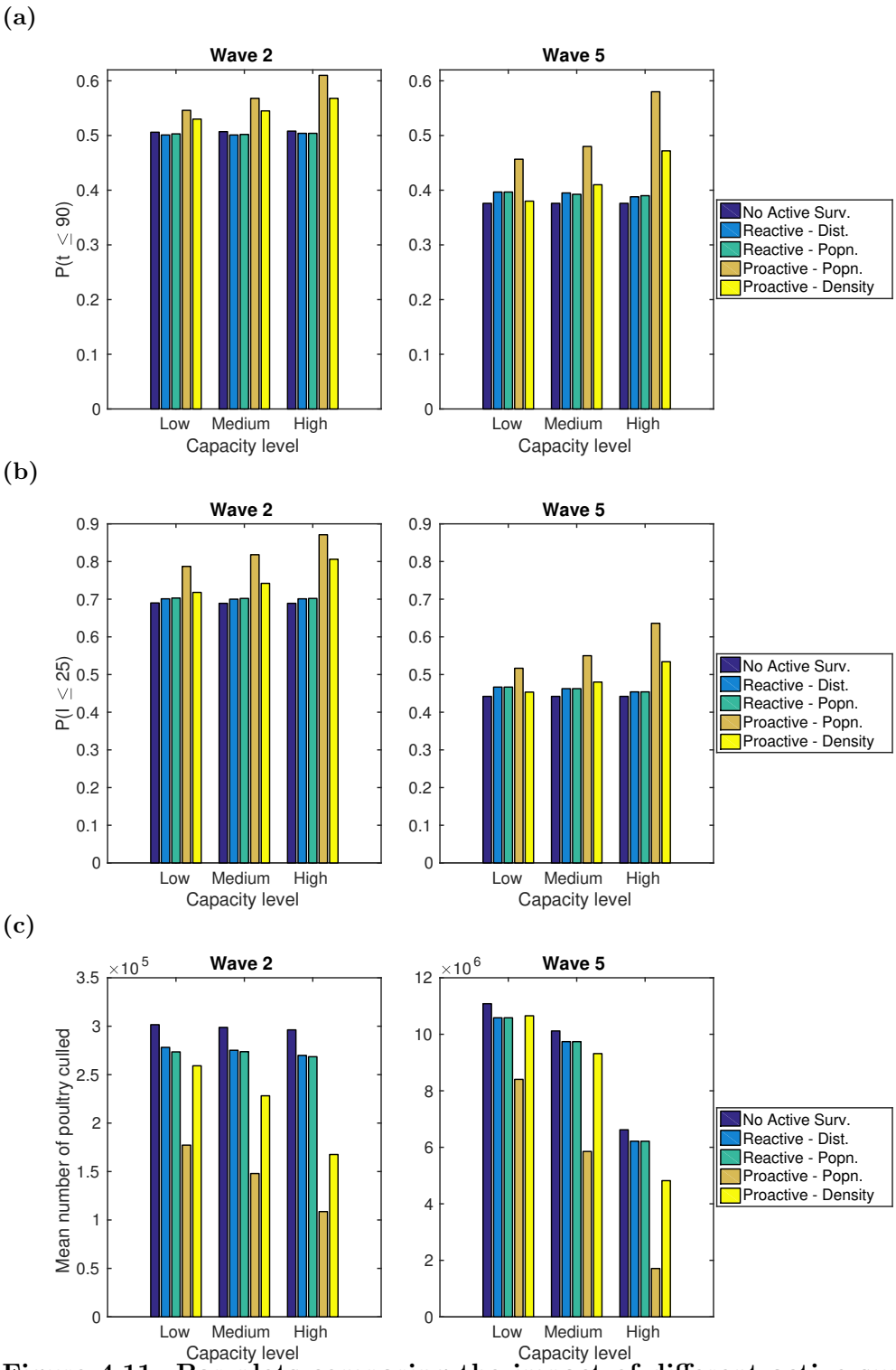

Figure 4.11: Bar plots comparing the impact of different active surveillance strategies on specific control objectives. For each combination of transmission model, resource restrictions and active surveillance strategy we performed a minimum of 300 simulation runs. The control objectives were: **(a)** predicted probability for outbreak duration t being 90 days or less; **(b)** predicted probability for outbreak size I not exceeding 25 premises; **(c)** mean number of poultry culled. For both wave 2 and wave 5 transmission dynamics the 'proactive by population' surveillance strategy was found to be optimal for all control objectives considered, irrespective of the capacity limitations. Full values are given in tables B.8 to B.10.

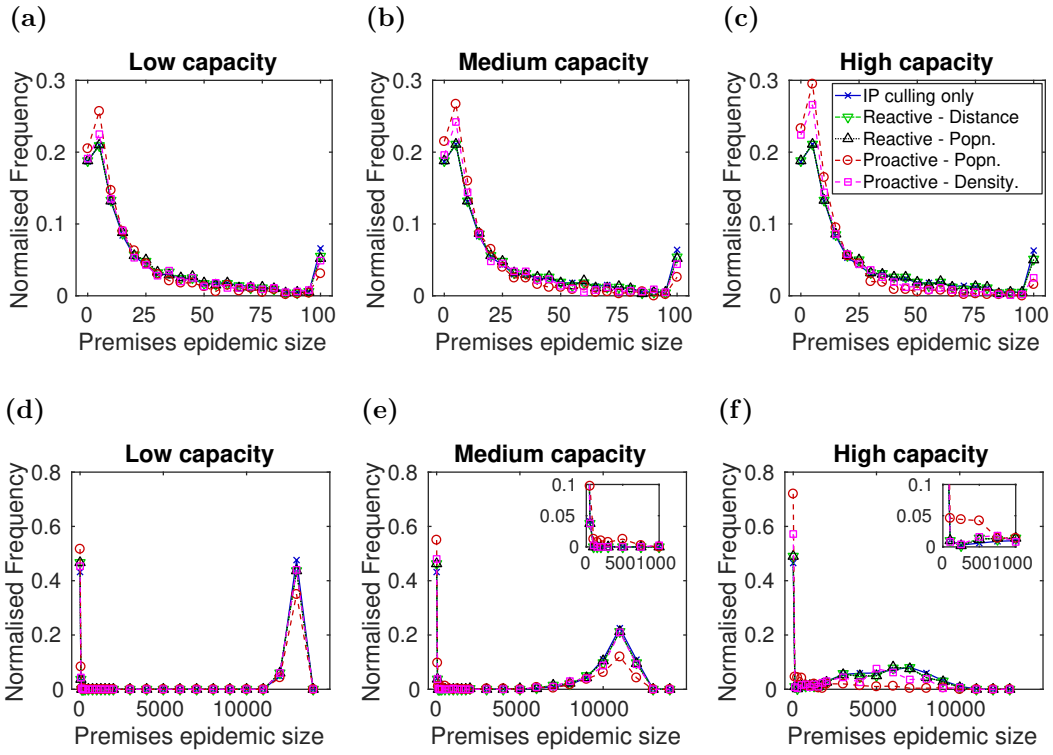

Figure 4.12: Impact on premises epidemic size using different active surveillance strategies with the specified capacity restriction. Four premises targeting strategies were tested and compared to the case where no active surveillance was in place (solid blue line, crosses); reactive by distance (dash-dot green line, inverted triangles), reactive by population (dotted black line, triangles), proactive by population (dashed red line, circles), proactive by density (dashed magenta line, squares). The proactive strategies lead to a decreased chance of an outbreak reaching more than 100 premises, while the reactive strategies offer minor gains compared to having no active surveillance. **(a–c)** Wave 2 transmission model, where the normalised frequency at 100 also includes all epidemic sizes 100 or greater. **(d–f)** Wave 5 transmission model.

Targeting zoonotic spillover

Quantitatively similar results for the proportion of runs with at least one spillover transmission event were obtained regardless of the culling capacity limits in place (figure 4.13(a)). A decrease in this statistic was observed with percentage reduction in ϵ_h (ranging from 50-100%). This was more severe for the wave 2 transmission model, which we previously found to have greater dependence on factors encapsulated by the ϵ_h parameter (see section 3.3.3). More specifically, while a 50% reduction in ϵ_h reduced the number of runs with at least one human case occurrence event by a fifth, the drop became substantial when ϵ_h was fully eradicated (i.e. set to zero), to just 40% of the total compared to the no targeted control of zoonotic spillover scenario (when ϵ_h was left unscaled).

The strength of transmission dependence on factors encapsulated by the ϵ_h parameter was further shown when analysing the average number of days with human cases reported. Once more, quantitatively similar results were obtained regardless of the culling capacity limits in place (figure 4.13(b)), while under wave 2 type transmission we again see a drastic decrease in the objective measure with percentage reduction in ϵ_h. Yet, for wave 5 type transmission dynamics there was no discernible pattern between the two, with predicted values in the range of 90-100% of the estimates obtained when ϵ_h was left unscaled.

4.4 Discussion

This study emphasises how knowledge of both disease transmission dynamics and potential resource limitations for implementing an intervention can alter what are deemed the most effective actions for optimising specific H5N1 HPAI influenza control objectives among poultry within the Dhaka division of Bangladesh. Likewise, we saw differences in policy recommendations when comparing alternative control objectives to one another, corroborating previous work that showed establishing the objective to be optimised is pivotal in discerning the management action that should be enacted [170].

If it is believed transmission is exclusively premises-to-premises, we found considerable variation in the preferred control strategy dependent upon the spatial location of the source of the outbreak, the relationship between risk of transmission and between premises distance (examined here by comparing the wave 2 and wave 5 transmission models), and the capacity restrictions that are in place. This was

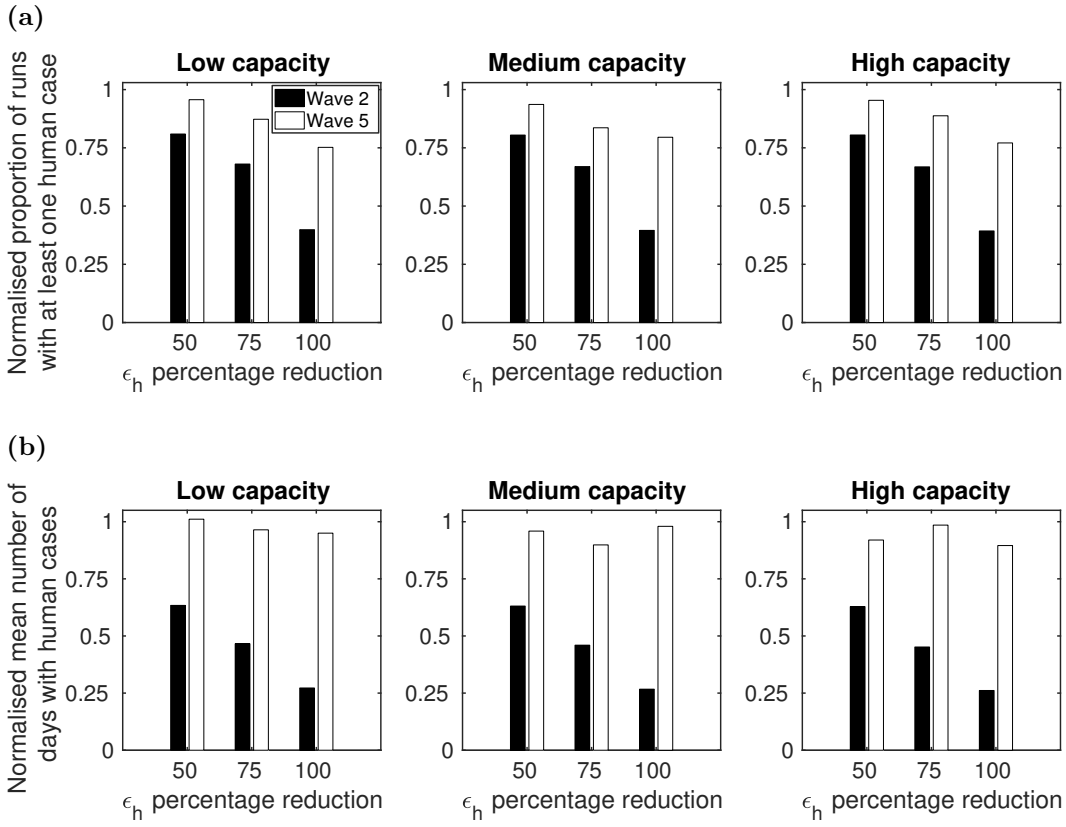

Figure 4.13: Bar plots of the likelihood of spillover transmission occurrence with respect to scaled values of the human case spark term. Bars were obtained by computing normalised proportions when scaled values of the human case spark term ϵ_h are used, relative to the scenario with unaltered ϵ_h values. Outcomes using the wave 2 transmission model are represented by the black bars, and those for wave 5 transmission model in white bars. **(a)** Normalised proportion of runs with at least one human case; **(b)** normalised mean number of days with human cases. Results were averaged over 1,000 simulation runs, with full values given in tables B.11 and B.12. Quantitatively similar results were obtained regardless of the culling capacity thresholds in place.

epitomised for ring culling and ring vaccination. Although there was a common trend of increasing the suggested radius of an intervention ring zone for less stringent capacity settings, solely culling infected premises was sometimes expected to be the best course of action. Given insight into the exact outbreak circumstances, this shows the potential benefits of having flexibility to adapt the intervention that is ratified. Nonetheless, this is under the strong presumption that all required information related to the above three dependencies is known. We also note that increasing intervention measure severity may lead to marginal benefits with regards

to the control metric of interest, but the additional costs in enacting such schemes warrants consideration to enable an informed decision to be made on the scale of ring culling and/or vaccination that should be put into practice.

When it was assumed that external factors had a meaningful impact on the transmission dynamics, capacity considerations became key in ensuring the measure selected was the best suited for the task, while conclusions drawn between our models with differing transmission behaviour were, in general, qualitatively similar. This may be due to the fact that although the relative likelihood of premises-to-premises disease transmission (with respect to distance) was well understood, the threat of any premises within the division becoming infected at any time resulted in this factor becoming less influential. A notable exception to this typical outcome were situations that involved pursuing a ring culling intervention action in conjunction with control objectives linked to number of poultry culled, highlighting once more the great importance attributed to establishing the control objective to be optimised to ensure the action pursued is in fact the most suitable.

It is vital that the area covered by ring based control methods is selected to only be as large as necessary. If set too small then other cases may emerge just outside the intervention zone that would have been dealt with had tougher measures been imposed. However, the use of widespread pre-emptive culling based on defined areas around an outbreak has been shown to be very difficult to implement effectively in developing countries. Enforcing wide area culling can alienate farmers if healthy birds are destroyed, or inadequate compensation is provided or is provided too late. This loss of poultry owner cooperation can be counter-productive, leading to resentment and resistance to further control measures [143]. Additionally, if instead the decision was to invoke a particular ring size for the control measure, then whilst for the three capacity levels considered here there was a general performance hierarchy (high capacity allowance outperforming a medium capacity allowance, which in turn outperformed a low capacity allowance), there were a couple of exceptions to this general rule that require further investigation.

In terms of active surveillance procedures, a consistent outcome over all scenarios was the superior performance of proactive schemes, that constantly monitor a predetermined set of premises based on selective criteria, over reactive schemes, which are only enforced once an outbreak has begun. In particular, proactive schemes focused on monitoring the premises with the largest flocks were the most successful, with

larger coverage levels strengthening performance outcomes. As a cautionary note, the reactive active surveillance strategies tested here assumed a maximum coverage radius of 500m around each premises reporting infection, with the sensitivity of the performance of these schemes to this coverage distance warranting further study.

Surveillance methodology is a discipline requiring greater attention. Only a fraction of studies on animal influenza surveillance have been specifically aimed at detecting influenza viruses with pandemic potential in animals, with calls for increased surveillance globally targeting potentially zoonotic influenza viruses in relevant animal species [54]. An additional hurdle to overcome is a need to standardise sampling, testing, and reporting methods, including full-genome sequencing and sharing of isolates with the scientific community [171]. Improvements in the coverage and regularity of sampling, alongside the standardisation of reporting methods, should help surmount these concerns.

In situations where disease transmission between premises is weak, enforcement of control measures not directly applied to the poultry flocks themselves (e.g. public awareness campaigns) can dramatically cut the risk of spillover transmission of zoonotic influenza at the poultry-human interface, leading to a reduction in H5N1 human case occurrence. Nonetheless, there are certain aspects of controlling spillover transmission that have yet to be considered, including analysing collaborative control policies across both animals and humans. When considering the human-animal interface as a whole, it is possible alternative species-specific strategies become preferable to reduce the risk of spillover transmission in the first place, or further human cases if an outbreak is in progress. This area may be investigated further through additional forms of treatment being built into the model that influence population-level susceptibility to emerging influenza strains. Antivirals that intend to shorten the infectious period are one such option, with an exploratory mathematical study suggesting they are generally a more cost-effective control measure than non-pharmacologic methods [172]. Another possible treatment strategy to dampen the effect of an emerging pandemic is utilising pre-pandemic immunological priming. This involves prior immunisation with a different, immunologically distinct vaccine of the same HA subtype that offers the potential to 'prime' recipients for subsequent protection with a single booster dose of a manufactured vaccine matching the newly emergent pandemic strain [173–178]. Naturally, policy effectiveness will depend critically on how swiftly clinical cases are diagnosed and the speed with which antiviral drugs and vaccines can be distributed.

There are a couple of model framework facets to be addressed that would help support the instruction of H5N1 HPAI control policy. The zoonotic transmission component of our modelling framework is non-spatial. Bearing in mind that individual flocks may have differential infection risk, and differential contact rates with humans, the impact of various control strategies at the poultry level may well have a non-linear effect on human exposure. The presence of such an effect may alter our choice of optimal control strategy might well change. However, there is currently a dearth of knowledge about the intensity and type of contact patterns between livestock and humans that result in micro-organism transmission [89]. Therefore, data collection at the human-animal interface, in particular analysing contact habits with poultry, should be strongly encouraged. This can help quantify the proportion of the human population that has contact with poultry, and how the frequency and length of contacts varies across gender, age and occupation. These findings can then be used to predict the exposure of the human population to H5N1 in poultry flocks, aiding the accuracy of model simulations forecasting the effects of a particular control policy.

Another focal point for control, not explicitly included here, is LBMs. The high density and variety of avian hosts in LBMs supports the maintenance, amplification and dissemination of avian influenza viruses, while providing frequent opportunities for inter-species transmission events. In a meta-analysis of before-after studies, to assess the impact of LBM interventions on circulation of avian influenza viruses in LBMs and transmission potential to humans, Offeddu *et al.* [179] determined that periodic rest days, overnight depopulation and sale bans of certain bird species significantly reduced the circulation of avian influenza viruses in LBMs. Furthermore, prolonged LBM closure reduced bird-to-human transmission risk. Developing a theoretical model with LBMs included would allow us to validate these findings.

This initial analysis can be extended naturally in a number of additional ways to those already mentioned. In this chapter we considered conventional control strategies used to combat avian influenza outbreaks among poultry, namely culling, vaccination and active surveillance. One could compare these traditional schemes with innovative direct interruption strategies that modify the poultry production system [180]. An example would be intermittent government purchase plans, so that farms can be poultry-free for a short time and undergo disinfection. Another is to model restrictions on species composition. This aims to synchronise all flocks

on a premises to the same birth-to-market schedule, allowing for disinfection of the premises between flocks. A separate direction for further study is to understand whether the intensification of farming systems, which can alter the spatial configuration of flock locations, requires the severity of previously established control protocols to be amended to prevent a small-scale outbreak developing into a widespread epidemic.

In situations where the daily control resource thresholds would be exceeded, the extent to which other premises prioritisation schemes for administering the intervention of interest influences the results merits further examination. For the culling and vaccination controls deliberated here we assumed premises were prioritised by distance, from the outer edge of the designated ring control size inwards. Alternative prioritisation strategies that may be considered include ordering by flock size, in either ascending or descending order. Ultimately, public-health decision making generally necessitates the real-time synthesis and evaluation of incoming data. Optimal decision making for management of epidemiological systems is often hampered by considerable uncertainty, with epidemic management practices generally not incorporating real-time information into ongoing decision making in any formal, objective way. An adaptive management approach could be implemented to account for the value of resolving uncertainty via real-time evaluation of alternative models [170, 181].

To conclude, the results of this exploratory analysis illustrate some general principles of how disease control strategies of H5N1 in Bangladesh should be prioritised and implemented when having to account for resource availability. We have highlighted how targeting of interventions varies if it is believed transmission is predominately premises-to-premises, versus the scenario where importations and other external factors are included. Most importantly, based on this consideration, targeted active surveillance can significantly reduce the scale of an epidemic as long as the appropriate choice between reactive and proactive strategies is made. They also indicate that reactive culling and vaccination control policies should pay close attention to this factor, thus ensuring intervention targeting is optimised. Consequently, we advocate that much more attention is directed at identifying ways in which control efforts can be targeted for maximum effect.

Chapter 5

Evidence for history-dependence of influenza pandemic emergence

5.1 Introduction

In the previous two chapters we focused on developing, fitting and simulating mathematical models for H5N1 transmission, at a local scale, among poultry and across the poultry-human interface. In this chapter we take a broader viewpoint, aiming to instruct the type of intervention actions to be employed on an international scale to reduce the risk of influenza A strains with pandemic potential emerging and crossing into the human population.

Influenza A is a cause of considerable morbidity and mortality in humans, in particular being the class of influenza that is capable of causing global pandemics. Occasionally, humans become infected with a strain of influenza derived from non-human sources, which are essentially novel to humans. This can give rise to a localised outbreak that may develop into a worldwide influenza pandemic [12]. Though efforts are being made to improve pandemic influenza risk assessment, it is currently not possible to predict which non-human influenza A virus will cause the next pandemic or when it will emerge [57, 58].

Identifying possible causal (biological) mechanisms for emergence of strains with pandemic potential often relies on indirect measurements from laboratory and field surveillance studies, such as viral sequencing of archaic pathogens [182, 183] and seroprevalence analysis [134], which are undertaken at great risk and cost. Further-

more, recent experimental studies setting out to determine the necessary mutations for specific influenza strains to become transmissible among mammals [184–186] have been controversial [187]. Thus, we argue that adopting a mathematical modelling approach can potentially provide cheap, safe and direct validation of proposed assumptions regarding pandemic emergence. This may subsequently help inform the type of interventions that would have the greatest impact in reducing the risk of further influenza pandemic occurrences.

When referring to influenza, a pandemic signifies a world-wide epidemic caused by an emergent influenza A strain that transmits among humans, was not previously circulating among humans and to which most people do not have immunity [188]. Therefore, only virological techniques allow recognition of a pandemic with certainty. Nevertheless, work on early influenza outbreaks has argued that if the epidemic originated in one place, and from there spread world-wide with high morbidity, it was likely to have been a pandemic [189–192]. This allows a history of presumptive influenza pandemics since 1700 to be constructed, along with the waiting times between pandemic outbreaks.

From this, we can analyse statistically which of two opposing mechanisms can plausibly generate the waiting time data; a memoryless process or a history-dependent process. For the memoryless process, the time to the next pandemic is not influenced by how much time has passed since the previous pandemic. Outbreaks of Ebola are thought to follow such a process and are typically modelled under this assumption [193]. On the other hand, in a history-dependent process, the probability of an event occurring is influenced by the elapsed time since the last such event. Such a process may feasibly arise through a combination of contributing mechanisms. First, a currently circulating strain may need to undergo a required accumulation of mutations to develop pandemic potential, as found for specific H5N1 influenza strains requiring only a small number of mutations to become transmissible among mammals such as ferrets [184, 185]. Second, as a result of population immunity to prior strains, as suggested by the recycling theory for pandemic emergence [194, 195]. The recycling theory hypothesises that influenza A viruses with similar or identical hemagglutinins can re-emerge over time, with only a significant proportion of then-living individuals aged above a particular threshold age immunologically protected from the emergent virus. In effect, the spacing between pandemics is caused by waning population-level immunity to a previous pandemic strain, with the proportion of the population immunologically protected diminished by both age-related mortality

and the influx of immunologically naive newborns.

In this chapter, we utilise waiting times between presumptive influenza pandemics since 1700 and ascertain via a Bayesian analysis whether they are better modelled by a memoryless or history-dependent process. Using prior information gathered from hypotheses related to the emergence of flu strains, we demonstrate from this small but informative dataset evidence that spillover of strains with pandemic potential is likely to be a history-dependent process. With a weaker assumption of uninformed priors, conclusions may only be inferred if we are sure about the legitimacy of the contested pandemics. Forward simulations using the preferred models are then performed to obtain predictions for the number of influenza pandemics we are likely to see in the 100 years following the last pandemic.

5.2 Methods

5.2.1 Historic pandemic influenza data

We considered possible pandemics from 1700 to the present. Three historic pandemic timelines were constructed based on the number of supporting sources for each epidemic being a pandemic. Between-pandemic waiting times were obtained for each timeline.

All historic pandemic timelines proposed here contained at least four post-1900 pandemics, beginning in 1918, 1957, 1968 and 2009 respectively. In addition, seroarcheological observations have compellingly linked the emergence of an H3 influenza virus to a pandemic beginning in 1889 [195]. Thus, we included 1889 in all three timelines. Note there is a reputed influenza pandemic stated to have begun around 1900 [189, 191]. However, this epidemic was thought to have occasioned the emergence of the H3 pandemic virus, which was later found to have emerged in the preceding 1889 pandemic as just described [195]. In light of these findings this outbreak is not universally considered a genuine pandemic, hence we did not include it within our analysis.

For the period 1700 up to 1889, recorded influenza epidemics that may have satisfied the criteria to be classed as a pandemic were obtained from Patterson [189], Beveridge [190] or Taubenberger and Morens [192]. These particular data sources were selected due to each providing a varied account of putative influenza pandemics since 1700, while collectively representing a range of possible occurrences.

'Timeline A' had the most stringent inclusion criteria. Along with the pandemics that have occurred from 1889 onwards, it only included additional epidemics during 1700-1888 agreed on across the multiple sources as being pandemics. However, we had no waiting time information for the pandemic in 1729, which left us with seven waiting times. For 'Timeline B', an additional five possible pandemics were included that were given in either Beveridge [190] or Taubenberger and Morens [192]. Our most inclusive case, labelled 'Timeline C', included 1732 as a separate pandemic as it is given as being distinct from the 1729 pandemic by some of the earlier sources [189, 190], though it is not known whether these were two pandemics separated by a very short interval or one pandemic with a long-delayed recurrence [192]. Furthermore, this timeline also included the 1977 re-emergence of H1N1, which is widely believed to have occurred due to human factors rather than the biological processes that are the main interest here [196, 197]. Note that although only the post-1900 pandemics have been virologically confirmed, efforts to draw useful conclusions would be hindered by the size of a dataset comprising only those pandemics, primarily due to the penalisation of the additional complexity in the history-dependent model when there are few observations. As a consequence, we did not analyse such a scenario here. Complete lists of each pandemic timeline are provided in table 5.1.

5.2.2 Pandemic influenza emergence - model fitting, comparison and suitability

Our investigation into the plausibility of the observed waiting time data being generated via a memoryless or history-dependent process was formulated as a model selection problem between exponential and gamma distributions. Strictly speaking, if we believe that history dependence is generated by a combination of several memoryless events, then we should use a very general class of distributions called *phase-type*, which in fact can provide an arbitrarily good approximation to almost any probability distribution [198, 199]. In practice, however, any phase-type distribution that arises from a realistic combination of events will be very close to a gamma distribution and we therefore consider only gamma distributions as representations of history dependence, while noting that these may not be adequate if an unexpectedly complex set of individually memoryless events drives pandemic emergence.

While visual comparison of the three influenza pandemic timelines versus candidate

Table 5.1: Complete lists of the pandemics included in each of our proposed pandemic influenza timelines.

			Timeline		
Date	**Popular Name**	**Subtype**	**A**	**B**	**C**
1729					
1732			—	—	3 years†
1761			—	32 years*	29 years*
1781			52 years	20 years	20 years
1788			—	7 years*	7 years*
1800			—	12 years†	12 years†
1830			49 years	30 years	30 years
1847			—	17 years†	17 years†
1857			—	10 years†	10 years†
1889	Russian Flu	H3N?	59 years	32 years	32 years
1918	Spanish Flu	H1N1	29 years	29 years	29 years
1957	Asian Flu	H2N2	39 years	39 years	39 years
1968	Hong Kong Flu	H3N2	11 years	11 years	11 years
1977	Russian Flu	H1N1	—	—	9 years‖
2009	Swine Flu	H1N1	41 years	41 years	32 years

Start date (assumed point of new strain) is given, as well as popular name and most likely subtype. The delay dates for each timeline are given, and pandemics assumed not to have happened in the timeline are denoted —.
*Only listed as a probable pandemic by Taubenberger and Morens [192].
†Not listed as a separate and/or probable pandemic by Taubenberger and Morens [192].
‖Widely believed to have occurred due to human factors, thus only considered within the most inclusive timeline.

history-dependent and memoryless model simulation outputs suggest that the memoryless realisations have more clusters and long gaps than the real data (figure 5.1), human assessment of randomness is notoriously poor [200]. We therefore note the importance of performing a formal statistical analysis on the available data to allow us to quantify the level of support for each model.

Both competing distributions were fitted to each timeline using two different techniques.

Maximum likelihood estimation and Akaike information criterion

Firstly, parameter values were obtained using maximum likelihood estimation (MLE). Competing models were assessed using an extension of the Akaike information crite-

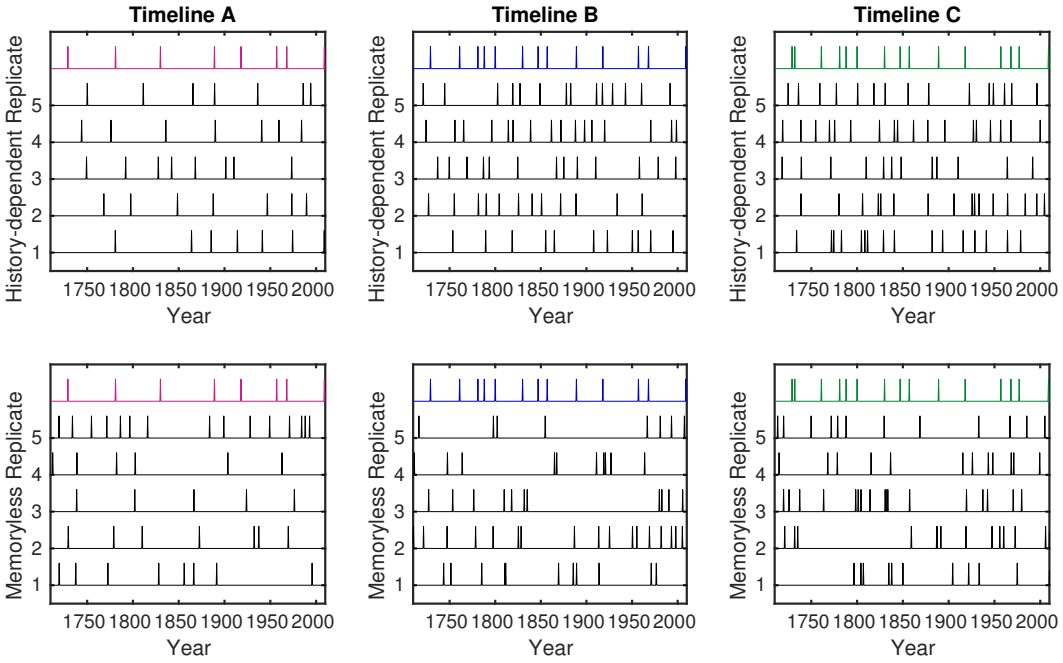

Figure 5.1: Visualisation of the observed timeline data and models. Timeline data is shown as a coloured series at the top of each plot, above five replicates for the fitted history-dependent (top row) and memoryless (bottom row) models.

rion (AIC) [201], namely corrected Akaike information criterion (AICc) [202]. These provide a measure of the relative quality of a collection of models for a given set of data, thus giving a means for model selection.

In detail, the AIC value of a model is given by

$$\mathrm{AIC} = 2k - 2ln(\hat{L}), \tag{5.1}$$

where k denotes the number of model parameters and $\hat{L}$ is the maximised value of the likelihood function for the model (obtained using MLE). This measure rewards goodness of fit, assessed by the likelihood function, but also includes a penalty that is an increasing function of the number of model parameters to discourage overfitting. Given a set of candidate models for the data, the preferred model is the one with the minimum AIC value. Further, the difference in AIC values between models, $\Delta\mathrm{AIC}_i = \mathrm{AIC}_i - \mathrm{AIC}_{\min}$, can indicate the level of empirical support for model i relative to the model giving the minimum AIC value, $\mathrm{AIC}_{\min}$. As a general rule, the larger $\Delta\mathrm{AIC}_i$ is, the less plausible it is that model i is the best model given the data, with $\Delta\mathrm{AIC}_i < 2$ corresponding to model i still having substantial support,

while models giving a $\Delta\text{AIC}_i > 10$ have no support and can be omitted from further consideration [203].

AICc is AIC with a correction for finite sample sizes, which for small sets of data essentially increases the relative penalty for including additional model complexity. Given the size of the datasets being analysed we use AICc, rather than AIC, for model selection purposes in this case. Explicitly,

$$\text{AICc} = \text{AIC} + \frac{2k(k+1)}{n-k-1}, \tag{5.2}$$

with n denoting the sample size. The criteria for selecting the preferred model match those described for AIC.

Reversible Jump MCMC

Second, owing to the small size of the data, we performed a Bayesian analysis using reversible jump Markov chain Monte Carlo (RJMCMC) [113].

In this framework the strength of evidence for each hypothesis is given by a probability. While a typical convention in Bayesian model selection is to use Bayes factors to qualitatively categorise the evidence against a null hypothesis [204], we instead deal directly with the marginal likelihoods due to the small nature of the datasets considered. These marginal likelihoods correspond to the rational subjective probabilities for each hypothesis. Such an approach is more appropriate than simply selecting one model over another when faced with small but informative data set.

Formally, let k denote the model / hypothesis, and D our observed data (the time periods between pandemics). When $k = 1$, the data follows an exponential distribution (parameterised by the rate parameter) so that $D_i \sim \text{Exp}(\lambda_1)$, for $i = 1, 2, \ldots, N$. For $k = 2$, the data follows a gamma distribution (parameterised by shape and rate parameters) so that $D_i \sim \text{Gamma}(\kappa,\lambda_2)$. Note that the shape parameter κ can be interpreted as a proxy for a specified number of events needing to occur before the next pandemic outbreak can begin. For the rate parameter, an uninformative $\text{Uniform}(10^{-3}, 1)$ prior was selected, representing a prior belief that pandemics are neither annual nor extremely infrequent. For κ, we considered three different priors containing different amounts of information about the mechanism behind pandemic emergence. The first of these assumed that the timing of pandemics was not very regularly spaced, or equivalently that the number of mutations required to cause a

pandemic was not particularly large. To describe this, an uninformed 'not clockwork' κ prior of Uniform$(0, 25)$ was used. This captures a prior belief that new pandemics are not very regular (the level of regularity for different values of κ is visualised in figure 5.2), while assuming no prior preference for specific values in the range over others. However, findings from experimental studies allow us to suggest alternatives for the prior informed from hypotheses relating to emergence of influenza strains (e.g. only four or five mutation steps are needed for specific H5N1 influenza strains to become transmissible between ferrets [184, 185]). Two differing informed priors for κ were selected to incorporate this additional mechanistic information, which had the same mean (four) but dissimilar variability to account for contrasting levels of certainty. The first was a 'weakly mechanistic' Exp$(1/4)$ prior, which gives lower prior credibility to large values of κ than the uniform prior. The second was a 'strongly mechanistic' Gamma$(\tilde{\kappa}, \tilde{\lambda})$ prior, with shape parameter $\tilde{\kappa} = 2$ and rate parameter $\tilde{\lambda} = 1/2$, which gave little credibility both for one influenza pandemic being able to immediately follow another (i.e. small values of κ) as well as low credibility to very large, regular spacing between pandemics (large values of κ, see figure 5.2). For each of the three choices of prior for κ we fit the data for each of the three timelines separately, giving nine different sets of modelling assumptions.

For each dataset the RJMCMC sampler was run until we obtained 10^5 samples, thinning by a factor of twenty sweeps with a burn-in period of 10^4 sweeps. Within each sweep within-model moves for each parameter were tried, before a trans-dimensional move was attempted. In the within-model moves, each of the relevant parameters λ_1, λ_2, κ were updated using the Metropolis-Hastings algorithm [107, 108] (algorithm 1), with a Gaussian proposal distribution q; for example, $\lambda_1' \sim N(\lambda_1, \sigma')$, with σ' 'tuned' to give an adequate acceptance rate.

Here we describe how the reversible jump moves were implemented. Firstly, consider the move from model 1 (exponentially distributed model) to model 2 (gamma distributed model). This move requires an auxiliary random variable U, in order to 'match dimensions' between the two model states. We generated u from a Gamma(α, β) distribution (parameterised by shape and scale parameters). For the uninformed 'not clockwork' and 'weakly mechanistic' κ prior cases, we fixed $\alpha = 2$ and $\beta = 2$. These were altered for the 'strongly mechanistic' κ prior case, with $\alpha = 4$ and $\beta = 1$. These were chosen so as to optimise mixing between the two models. Then set $\kappa = u$ and leave the parameter λ as it is. Note that the reverse move, from model 2 to model 1, requires no auxiliary random variable as model

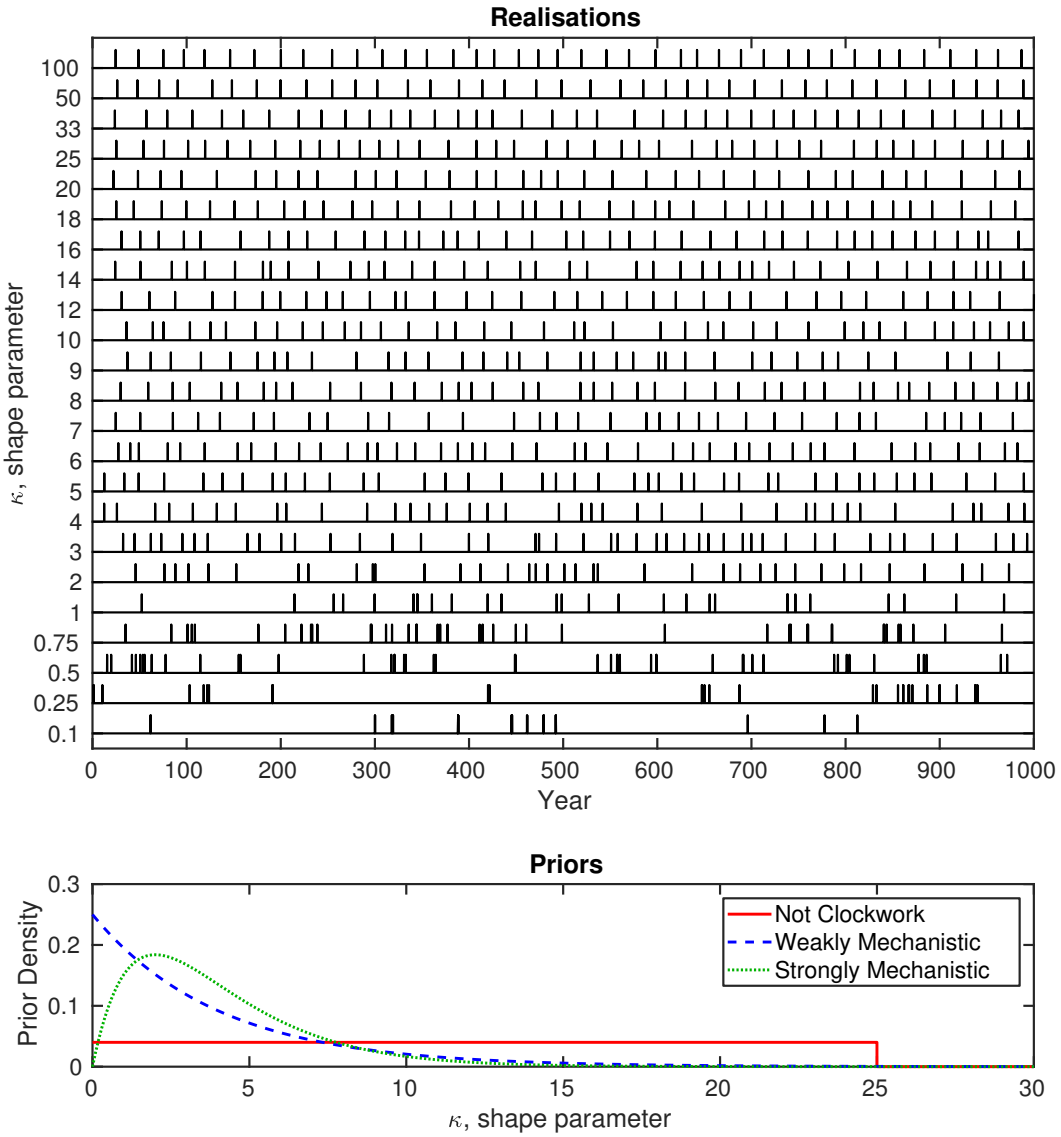

Figure 5.2: (Top) Visualisation of the implications of different gamma-distributed times (each with mean 25) between pandemics. **(Bottom)** Credibility given by the three κ priors to differing levels of regularity.

1 has fewer parameters than model 2. Instead, we simply maintain the value of the parameter λ. The Jacobian factor for these model transformations is 1. The acceptance probability for the proposed move from model 1 to 2 is min $\{1, A_{1,2}\}$, where

$$A_{1,2} = \frac{L(D|\lambda,\kappa)p(\lambda,\kappa|k=2)p(k=2)}{L(D|\lambda)p(\lambda|k=1)p(k=1)} \left(\frac{u^{\alpha-1}e^{-\frac{u}{\beta}}}{\beta^{\alpha}\Gamma(\alpha)}\right)^{-1}$$

and from model 2 to 1 is $\min\{1, A_{2,1}\}$, where

$$A_{2,1} = \frac{L(D|\lambda)p(\lambda|k=1)p(k=1)}{L(D|\lambda,\kappa)p(\lambda,\kappa|k=2)p(k=2)} \left(\frac{\kappa^{\alpha-1}e^{-\frac{\kappa}{\beta}}}{\beta^{\alpha}\Gamma(\alpha)} \right)$$

with $p(k=i)$ corresponding to the prior assigned to model i. Our priors on each model were $p(k=1) = p(k=2) = 0.5$. Note that $A_{1,2}$ and $A_{2,1}$ are reciprocals after change of notation. For the interested reader, a generalised description of the RJMCMC algorithm from Green [113] is outlined below (algorithm 7).

The acceptance rates for these cross-model moves were recorded to track sampler performance. Higher acceptance rates lead to lower autocorrelation in the k chain, while near-zero acceptance rates would indicate the sampler was not performing adequately due to being highly inefficient in exploring the parameter space.

Testing model adequacy

To assess goodness-of-fit Monte Carlo simulations were performed for each of our fitted memoryless and history-dependent models, using 10^4 samples generated from the RJMCMC procedure. A simulation duration equivalent to 300 years was chosen to approximately match the timespan of our historic pandemic lists. Empirical cumulative distribution functions (ECDF) of the inter-event time between pandemics were computed for each simulation run. These were compared to the observed data through survival functions (1 - ECDF).

5.2.3 Validation - Ebola outbreak analysis

The validity of the RJMCMC method was checked by performing a similar analysis on the waiting time data for Ebola outbreaks (the month and year of each outbreak obtained from sources listed by the CDC [205]), with the expectation being the memoryless model would be strongly preferred since Ebola is too virulent in humans to expect to see phenomena such as population-level immunity or long-term evolution. Parameter priors used in this case were Uniform$(10^{-3}, 1)$ for λ_1, λ_2 and Uniform$(0, 25)$ for κ.

5.2.4 Forward simulation outline

For each pandemic influenza timeline and prior type combination, 10^4 samples generated from the RJMCMC procedure were used to simulate forward 100 years (from 2010 to 2110). The number of samples drawn from each model was weighted by

Algorithm 7 Single iteration of the Reversible Jump MCMC

Input: $m_{t-1}, \tilde{\theta}, D, \pi, p$
▷ Last model state, current parameter state following within-model move, observed data,parameter prior, model prior
Output: m_t, θ_t ▷ Next model and parameter states
$m' \sim j(m'|m_{t-1})$ ▷ Sample move-type

Sample auxiliary random variable to 'match dimensions'
$u \sim g_{m_{t-1}\to m'}(u)$

Propose new model parameters using mapping h
$(\theta', u') = h_{m_{t-1}\to m'}(\tilde{\theta}, u)$

Sample auxiliary random variables to 'match dimensions' for the reverse move
$u' \sim g_{m'\to m_{t-1}}(u')$

Calculate the acceptance probability
$\alpha_{m_{t-1}\to m'} = \min\left\{1, A_{m_{t-1},m'}\right\}$,
where $A_{m_{t-1},m'} = \min\left\{1, \frac{\mathcal{L}(D|m',\theta')\pi(\theta'|m')p(m')j(m_{t-1}|m')g_{m'\to m_{t-1}}(u')}{\mathcal{L}(D|m_{t-1},\tilde{\theta})\pi(\tilde{\theta}|m_{t-1})p(m_{t-1})j(m'|m_{t-1})g_{m_{t-1}\to m'}(u)}J\right\}$,
with $J = \left|\frac{\partial}{\partial(\tilde{\theta},u)}h_{m_{t-1}\to m'}(\tilde{\theta}, u)\right|$.

$u \sim \text{Unif}[0,1]$
if u $< \alpha_{m_{t-1}\to m'}$ **then** ▷ Accept proposed states with probability $\alpha_{m_{t-1}\to m'}$
    $m_t = m'$
    $\theta_t = \theta'$
else
    $m_t = m_{t-1}$
    $\theta_t = \tilde{\theta}$
end if

Table 5.2: Fitted parameter values, 95% confidence intervals and AICc values for each timeline. Exponential distribution was parameterised by rate parameter λ_1, Gamma distribution was parameterised by shape parameter α and scale parameter β.

	Model	Param.	Value	(95% CI)	AICc
1. Timeline A	Exponential	λ_1	0.0250	(0.0101, 0.0466)	68.4
	Gamma	α	4.77	(1.73, 13.14)	66.5
		β	8.39	(2.88, 24.4)	
2. Timeline B	Exponential	λ_1	0.0429	(0.0221, 0.0703)	102.0
	Gamma	α	3.56	(1.66, 7.67)	97.3
		β	6.55	(2.88, 14.9)	
3. Timeline C	Exponential	λ_1	0.0500	(0.0273, 0.0794)	114.2
	Gamma	α	2.52	(1.26, 5.07)	111.7
		β	7.92	(3.66, 17.1)	

the posterior model probabilities for the given timeline. Distributions of the proportion of simulations giving a specified number of influenza pandemic events in this time period were constructed, allowing for between-timeline comparisons. All calculations and simulations were performed with MATLAB®.

5.3 Results

5.3.1 AICc results

We conclude from the AICc values (equation (5.2)) that the gamma distributed model was modestly preferred across all three timelines, suggesting influenza pandemic inter-event times could be plausibly generated by a history-dependent process. We do note, however, that for each timeline the data was insufficient to categorically support selecting one model over the other, with differences in AICc values between the two model hypothesis being less than five (see table 5.2).

5.3.2 Pandemic influenza emergence - RJMCMC model fitting, comparison and suitability

Running our RJMCMC sampler on each timeline assuming an uninformed 'not clockwork' κ prior gave good levels of between-model mixing, with cross-model jump acceptance rates between 30% and 60%. In contrast, the informed 'mechanistic' κ prior cases had cross-model jump acceptance rates between 15% and 40%, though the thinning used reduced autocorrelation in the k chain to ensure the samples drawn could be considered as independent.

Table 5.3: Posterior probabilities given to the history-dependent hypothesis for each timeline and choice of κ prior.

	κ prior		
Timeline	**Not clockwork**	**Weakly mechanistic**	**Strongly mechanistic**
A	0.70	0.82	0.86
B	0.77	0.89	0.92
C	0.45	0.73	0.77

When using the 'not clockwork' κ prior, the gamma distributed model (history-dependent hypothesis) had higher posterior probabilities for Timelines A and B (0.70 and 0.77 respectively). Nevertheless, for Timeline C the history-dependent hypothesis had less support than the memoryless hypothesis (0.45 vs 0.55). Consequently, this prior assumption only lets us infer something if we are sure which pandemic timeline is correct. On the other hand, using 'mechanistic' priors informed by studies related to emergence of flu strains resulted in model probabilities between 0.70 and 0.92 for the history-dependent hypothesis across our range of exclusive to inclusive histories of pandemic influenza (table 5.3).

Comparing across the different κ priors, the median and 95% credible intervals for each model parameter are quantitatively similar. Unsurprisingly, the main exceptions to this were the 'weakly mechanistic' and 'strongly mechanistic' prior resulting in tighter posterior distributions for κ compared to when the 'not clockwork' prior was used (figure 5.3 and table 5.4). Timeline A exhibits the widest posterior distributions for κ, along with a greater median value for κ and lower λ median values (figure 5.3 and table 5.4). This coincided with a broader distribution for the standard deviation σ (table 5.5), implying longer and greater variability in spacing between influenza pandemics relative to Timelines B and C.

For each of our nine sets of modelling assumptions the posterior simulated survival function distributions for pandemic event waiting times indicated a reasonable model fit to the data when a history-dependent hypothesis for influenza pandemic emergence was assumed. A good level of agreement was observed in particular under this hypothesis in combination with the 'strongly mechanistic' prior assumption for κ, with the predicted median waiting time survival function closely tracking the empirical data. On the other hand, the memoryless model exhibited greater inter-event time variability, with the predicted survival function distributions generally declining sooner and having a longer tail relative to the data (figure 5.4). This

reaffirms the fact that such an assumption lends greater credibility for both long interludes between pandemics (exceeding 100 years) and for one influenza pandemic being able to immediately follow another relative to what truly occurred. Similar outcomes were obtained using the alternative κ prior distributions (figures 5.5 and 5.6).

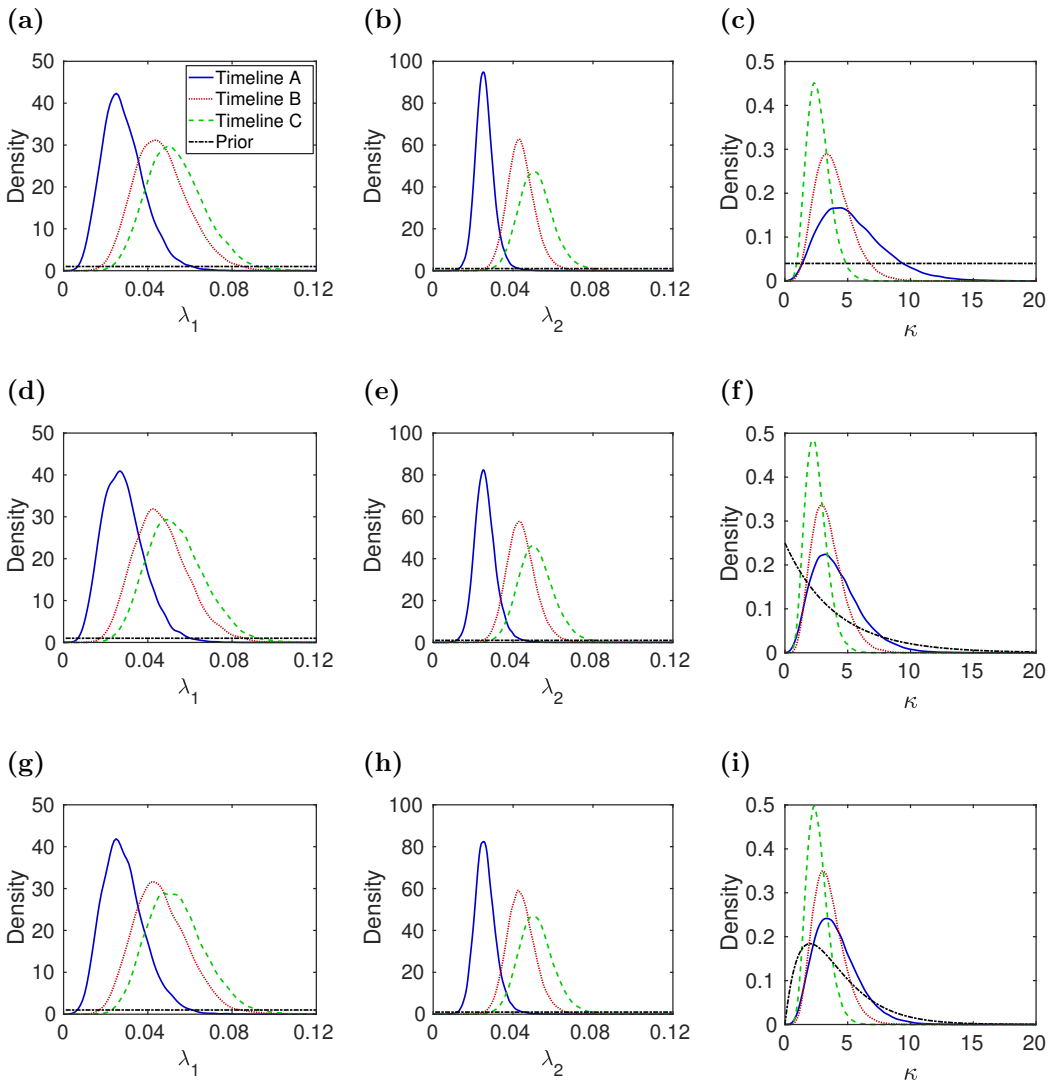

Figure 5.3: Parameter probability distribution functions estimated from RJMCMC output when using the specified prior for κ: (a-c) Not clockwork; (d-f) weakly mechanistic; (g-i) strongly mechanistic. (a,d,g) Density estimate for λ_1 where $k = 1$; **(b,e,h)** density estimate for λ_2 where $k = 2$; and **(c,f,i)** density estimate for κ where $k = 2$.

Table 5.4: Model Fitting I for each κ prior and timeline.

κ prior	Timeline	Model	Param.	Value	(95% CI)	π
Not clockwork	**A**	Exp.	λ_1	0.0273	(0.0122, 0.0515)	0.30
		Gamma	λ_2	0.0255	(0.0175, 0.0362)	0.70
			κ	5.03	(1.61, 11.8)	
	B	Exp.	λ_1	0.0453	(0.0247, 0.0751)	0.23
		Gamma	λ_2	0.0435	(0.0315, 0.0588)	0.77
			κ	3.68	(1.58, 7.19)	
	C	Exp.	λ_1	0.0522	(0.0300, 0.0837)	0.55
		Gamma	λ_2	0.0510	(0.0357, 0.0707)	0.45
			κ	2.59	(1.21, 4.80)	
Weakly mechanistic	**A**	Exp.	λ_1	0.0275	(0.0123, 0.0512)	0.18
		Gamma	λ_2	0.0256	(0.0166, 0.0383)	0.82
			κ	3.81	(1.24, 8.84)	
	B	Exp.	λ_1	0.0452	(0.0248, 0.0753)	0.11
		Gamma	λ_2	0.0436	(0.0308, 0.0600)	0.89
			κ	3.23	(1.40, 6.32)	
	C	Exp.	λ_1	0.0523	(0.0299, 0.0832)	0.27
		Gamma	λ_2	0.0510	(0.0353, 0.0714)	0.73
			κ	2.40	(1.12, 4.46)	
Strongly mechanistic	**A**	Exp.	λ_1	0.0274	(0.0124, 0.0520)	0.14
		Gamma	λ_2	0.0257	(0.0168, 0.0379)	0.86
			κ	3.85	(1.41, 8.30)	
	B	Exp.	λ_1	0.0451	(0.0245, 0.0747)	0.082
		Gamma	λ_2	0.0436	(0.0311, 0.0596)	0.92
			κ	3.31	(1.52, 6.22)	
	C	Exp.	λ_1	0.0524	(0.0301, 0.0835)	0.23
		Gamma	λ_2	0.0511	(0.0356, 0.0708)	0.77
			κ	2.50	(1.22, 4.50)	

Fitted parameter values and 95% credible intervals for each parameter in the rate / shape parameterisation (to 3 s.f.) and marginal posterior π for each model (to 2 s.f.) in the three timelines considered.

Table 5.5: Model Fitting II for each κ prior and timeline.

κ prior	Timeline	Model	Param.	Value	(95% CI)	π
Not clockwork	**A**	Exp.	μ_1	36.6	(19.4, 81.8)	0.30
		Gamma	μ_2	39.3	(27.6, 57.1)	0.70
			σ	17.4	(10.8, 35.7)	
	B	Exp.	μ_1	22.1	(13.3, 40.5)	0.23
		Gamma	μ_2	23.0	(17.0, 31.8)	0.77
			σ	11.9	(7.99, 20.9)	
	C	Exp.	μ_1	19.1	(12.0, 33.4)	0.55
		Gamma	μ_2	19.6	(14.1, 28.0)	0.45
			σ	12.1	(8.18, 21.2)	
Weakly mechanistic	**A**	Exp.	μ_1	36.3	(19.5, 81.3)	0.18
		Gamma	μ_2	39.0	(26.1, 60.2)	0.82
			σ	19.7	(12.2, 41.8)	
	B	Exp.	μ_1	22.1	(13.3, 40.4)	0.11
		Gamma	μ_2	22.9	(16.7, 32.4)	0.89
			σ	12.7	(8.44, 22.5)	
	C	Exp.	μ_1	19.1	(12.0, 33.4)	0.27
		Gamma	μ_2	19.6	(14.0,28.3)	0.73
			σ	12.6	(8.40, 22.1)	
Strongly mechanistic	**A**	Exp.	μ_1	36.5	(19.2, 80.8)	0.14
		Gamma	μ_2	39.0	(26.4, 59.4)	0.86
			σ	19.6	(12.4, 39.5)	
	B	Exp.	μ_1	22.2	(13.4, 40.8)	0.082
		Gamma	μ_2	22.9	(16.8, 32.2)	0.92
			σ	12.5	(8.48, 21.6)	
	C	Exp.	μ_1	19.1	(12.0, 33.2)	0.23
		Gamma	μ_2	19.6	(14.1, 28.1)	0.77
			σ	12.3	(8.36, 21.1)	

Fitted parameter values and 95% credible intervals for each parameter in the mean / standard deviation parameterisation (to 3 s.f.) and marginal posterior π for each model (to 2 s.f.) in the three timelines considered.

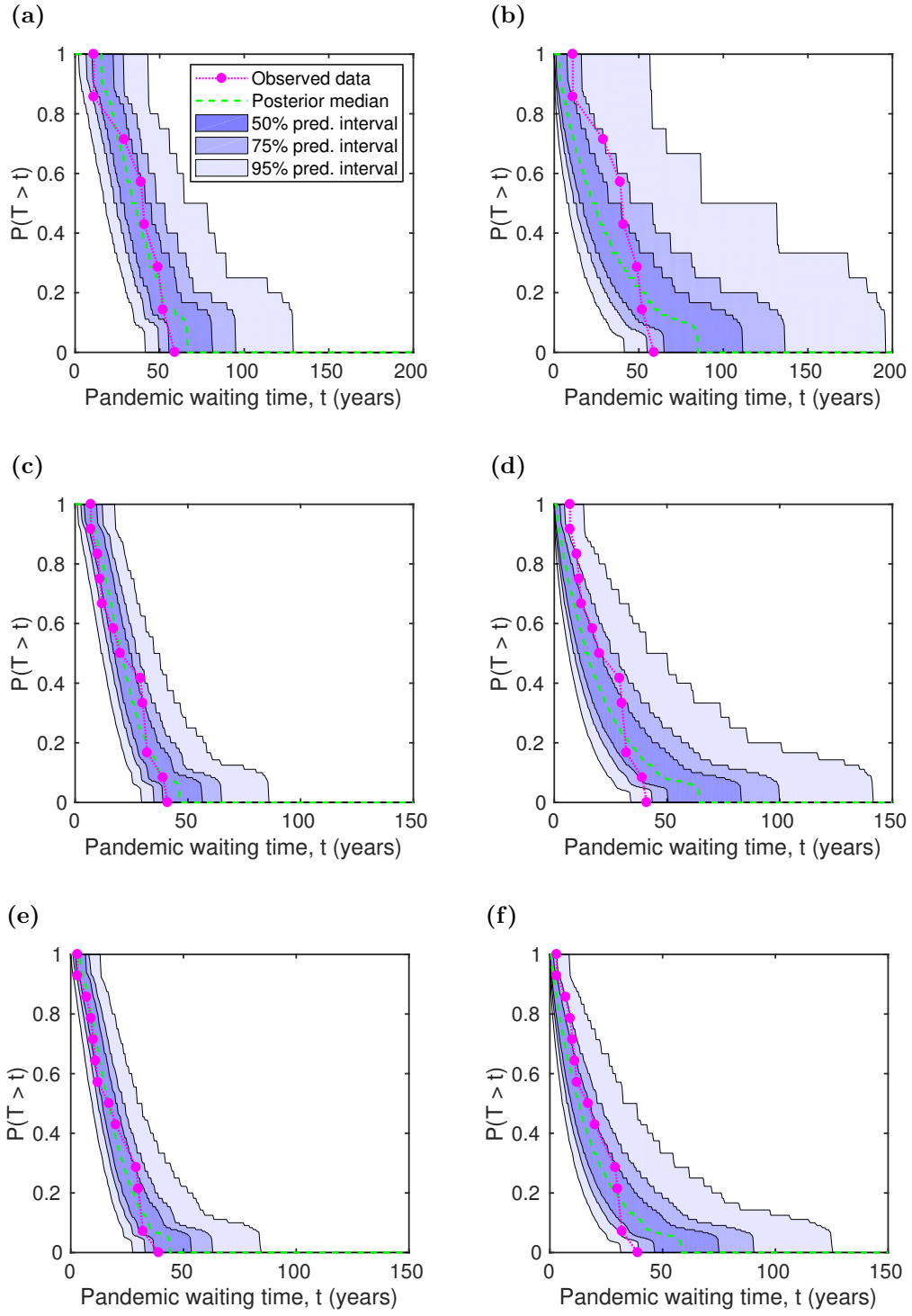

Figure 5.4: Predicted posterior influenza pandemic inter-event time survival functions versus the empirical survival function, under the 'strongly mechanistic' prior assumption. Waiting time model fits relative to the observed data (magenta dotted line) under the following hypothesis: **(left)** history-dependent; **(right)** memoryless. Across all proposed historic pandemic lists the history-dependent hypothesis corresponds adequately with the observed data. **(a,b)** Timeline A; **(c,d)** timeline B; **(e,f)** timeline C. Median posterior survival functions are given by the green dashed lines, with prediction intervals of 50%, 75% and 95% represented by shaded blue regions (moving from darkest to lightest).

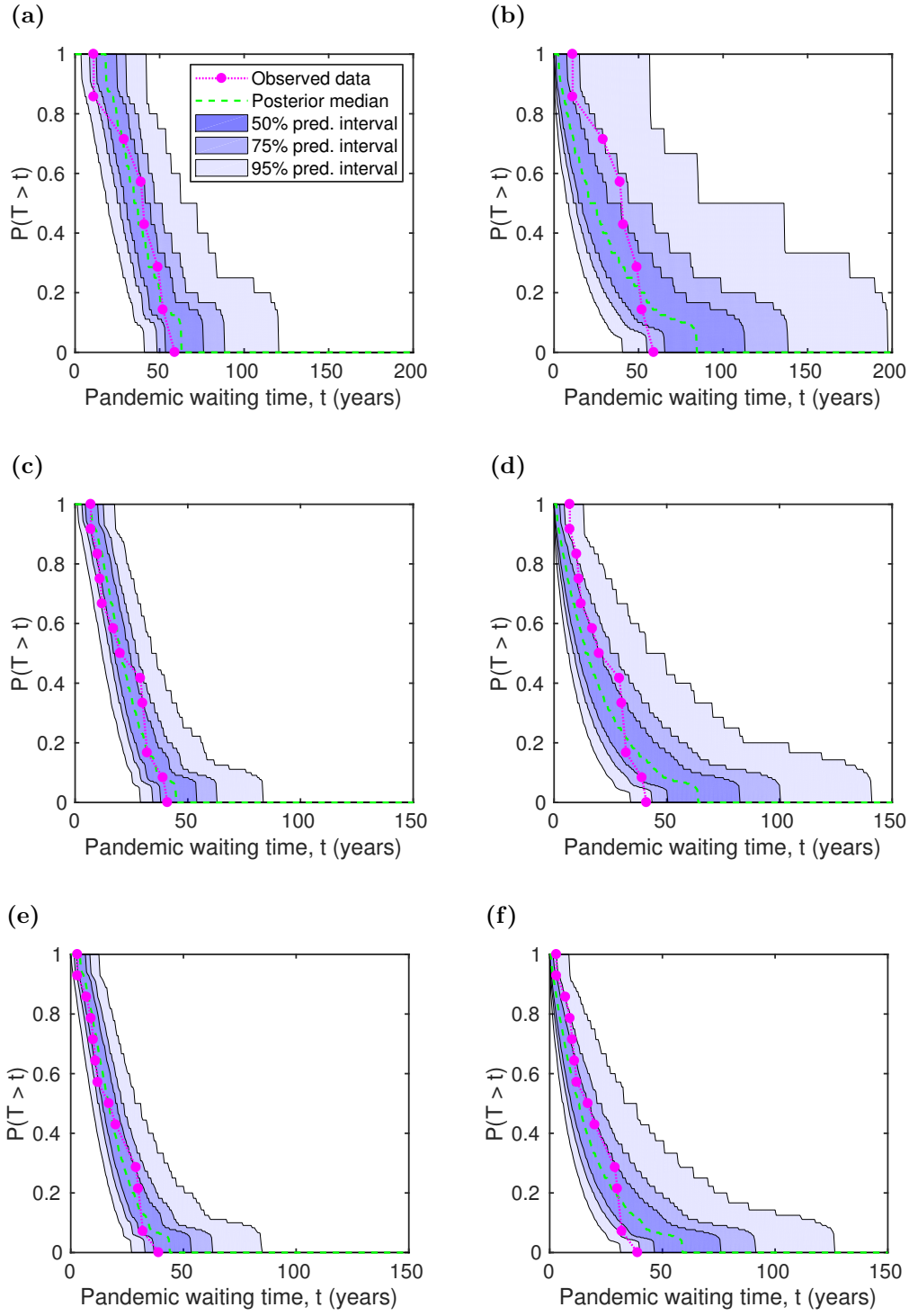

Figure 5.5: Predicted posterior influenza pandemic inter-event time survival functions versus the empirical survival function, under the 'not clockwork' prior assumption. Waiting time model fits relative to the observed data (magenta dotted line) under the following hypothesis: **(left)** history-dependent; **(right)** memoryless. Across all proposed historic pandemic lists the history-dependent hypothesis corresponds adequately with the observed data. **(a,b)** Timeline A; **(c,d)** timeline B; **(e,f)** timeline C. Median posterior survival functions are given by the green dashed lines, with prediction intervals of 50%, 75% and 95% represented by shaded blue regions (moving from darkest to lightest).

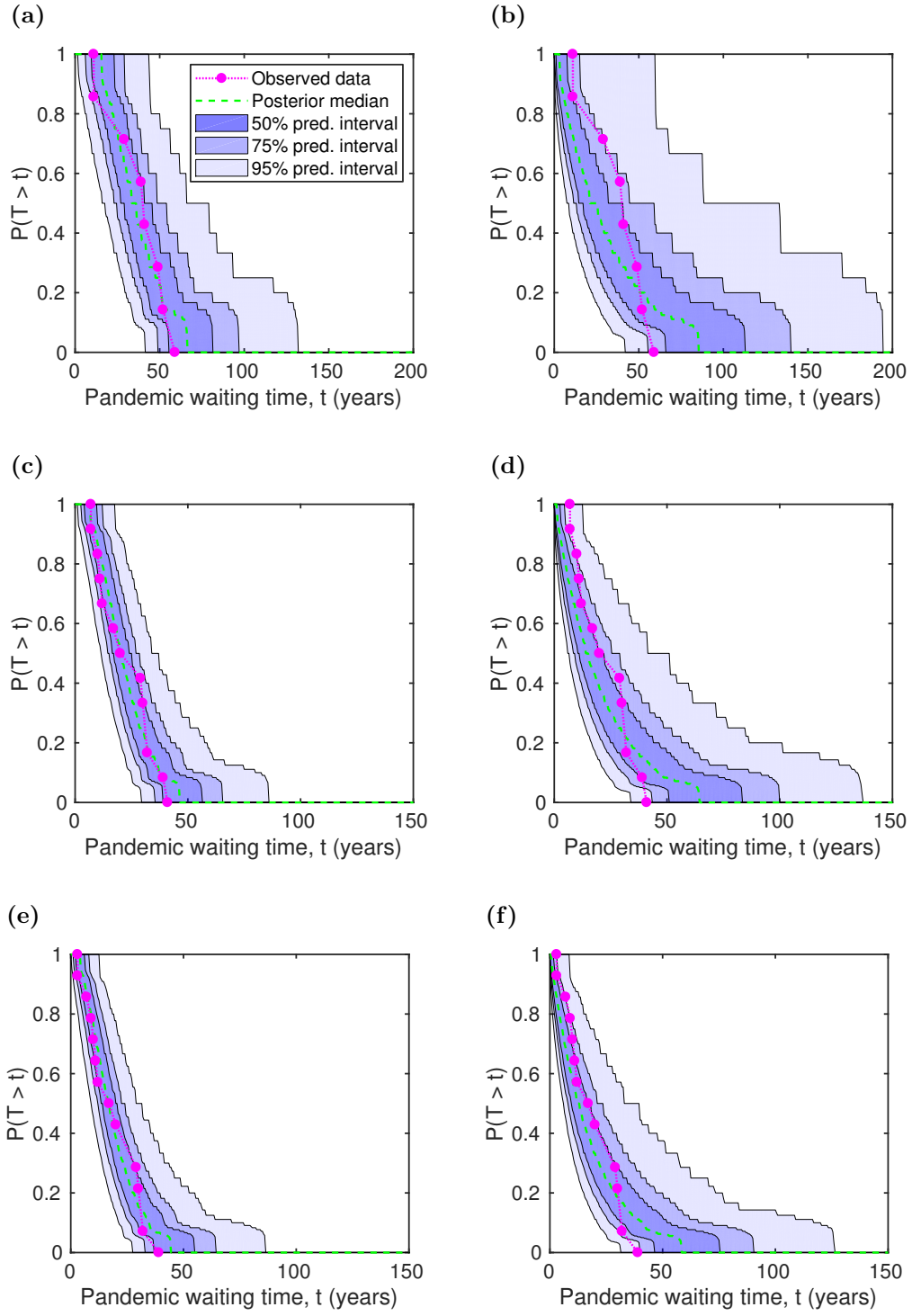

Figure 5.6: Predicted posterior influenza pandemic inter-event time survival functions versus the empirical survival function, under the 'weakly mechanistic' prior assumption. Waiting time model fits relative to the observed data (magenta dotted line) under the following hypothesis: **(left)** history-dependent; **(right)** memoryless. Across all proposed historic pandemic lists the history-dependent hypothesis corresponds adequately with the observed data. **(a,b)** Timeline A; **(c,d)** timeline B; **(e,f)** timeline C. Median posterior survival functions are given by the green dashed lines, with prediction intervals of 50%, 75% and 95% represented by shaded blue regions (moving from darkest to lightest).

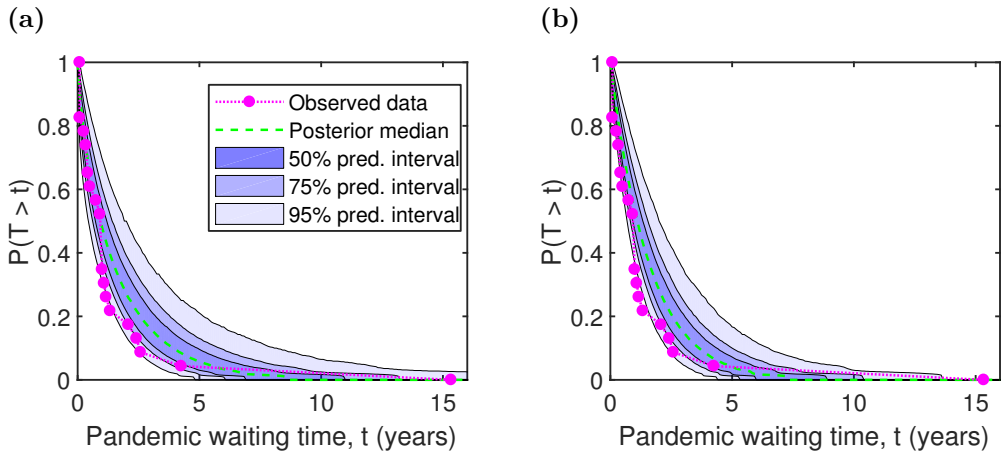

Figure 5.7: Predicted posterior Ebola outbreak inter-event time survival functions versus the empirical survival function. Waiting time model fits relative to the observed data (magenta dotted line) under the following hypothesis: **(a)** history-dependent; **(b)** memoryless. The model hypothesis generated similar waiting time survival function profiles, both being heavier tailed than the observed data. Median posterior survival functions are given by the green dashed lines, with prediction intervals of 50%, 75% and 95% represented by shaded blue regions (moving from darkest to lightest).

5.3.3 Validation - Ebola outbreak analysis

When fitting our models to the Ebola outbreak waiting times we obtained 10^5 samples, thinning by a factor of 100 sweeps with a burn-in period of 10^4 sweeps. As expected, the memoryless model was strongly favoured compared to the history-dependent model, with posterior model probabilities of 0.95 and 0.05 respectively. Furthermore, this weight of evidence was supplemented by the inferred gamma model being exponential-like in nature, with the following parameter median and 95% credible intervals obtained: $\kappa = 0.719$ (0.429,1.15); $\lambda_2 = 0.618$ (0.366,0.925). Similar values were found for the exponential model rate parameter, $\lambda_1 = 0.619$ (0.403,0.890). As a result, the model-specific simulated survival function distributions for Ebola outbreak waiting times strongly resembled one another, both being heavier tailed than the empirical survival function (figure 5.7).

5.3.4 Forward simulation analysis

Across our range of inclusivity criteria for proposed pandemics, the lowest mode for the number of predicted pandemics during 2010-2110 was obtained by Timeline A (two). In contrast, Timelines B and C gave modes of four or five (figure 5.8). Further, Timeline A displayed less variability in the predicted number of pandemic

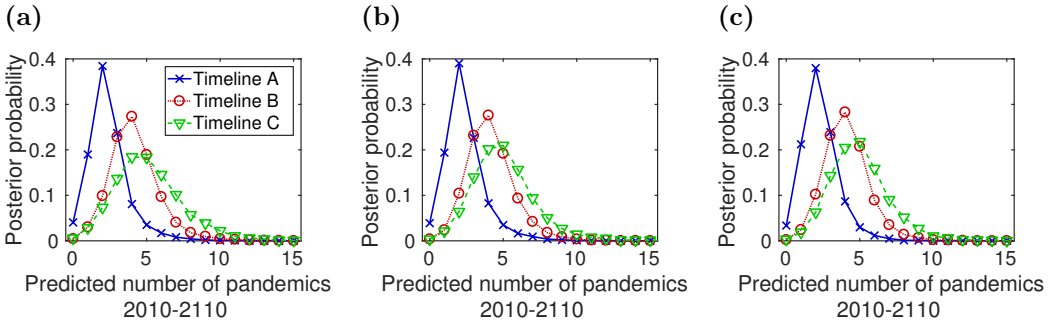

Figure 5.8: Posterior predictive distributions for the number of pandemics between 2010-2110, for each timeline and choice of κ prior. (a) 'Not clockwork' κ prior case; **(b)** 'weakly mechanistic' κ prior case; **(c)** 'strongly mechanistic' κ prior case.

events, with the majority of simulations predicting two or fewer pandemics between 2010-2110. When the 'not clockwork' κ prior was used, the proportion of simulations predicting two or fewer pandemics was 62% for Timeline A, compared to just 13% and 11% for Timelines B and C respectively. Quantitatively similar results were obtained using either 'mechanistic' κ prior (for the 'weakly mechanistic' prior in particular, Timeline A: 62%, Timeline B: 13%, Timeline C: 8.9%).

Timeline C had a more diffuse distribution than Timeline B, despite including only a couple of additional waiting times. Explicitly, Timeline C attributed a greater weight of support to there being 10 or more pandemics between 2010 and 2110 (4.6% versus 1.1% of simulations). This discrepancy between the predictive pandemic event distributions remained using the 'mechanistic' κ priors, although these choices of prior did result in reduced absolute probabilities of an extreme number of pandemics (10 or more) occurring in the time period of interest (for the 'weakly mechanistic' prior, Timeline B: 0.63%, Timeline C: 3.2%).

5.4 Discussion

Assessing influenza pandemic emergence risk is an ongoing public health concern [57, 58]. This study was motivated by the view that mathematical approaches could provide cheap ground-truthing and validation of expensive and sometimes hazardous laboratory and field surveillance efforts to understand pandemic emergence. Our findings suggest evidence for the emergence of influenza pandemics being plausibly generated by a history-dependent process, although this does depend on the prior assumptions made. Even with the data being limited, utilising a RJMCMC approach

to model selection (rather than solely relying on AICc) indicates in a mathematically rigorous sense the strength of evidence for each hypothesis, improving upon merely selecting one model over another. Our analysis of the Ebola outbreak waiting time data, finding the memoryless hypothesis to be strongly preferred, provides confirmation that the RJMCMC model selection method works as expected for another real system.

An additional advantage of using RJMCMC as a tool for model selection is the ability to compare inferred model probabilities under various prior assumptions. For our particular model selection problem, under an uninformed 'not clockwork' prior assumption the preferred model mechanism depended on the strictness of the proposed pandemic inclusion criteria, only letting us infer model preference if we are sure which pandemic timeline is correct. We note, however, that advances in the recovery and sequencing of genetic information has led to analyses of ancient specimens, meaning genomic data from archaic pathogens (obtained from corpses) can be gathered. Recent examples include characterisation of the 1918 pandemic influenza virus [182, 183] and reconstruction of the bubonic plague genome from victims of the Black Death [206]. Further discovery and sequencing of past influenza viruses may weaken or strengthen the claims of the contested pre-20th century pandemics, culminating in the accuracy of the list of historic influenza pandemics being enhanced.

Nevertheless, using a 'mechanistic' prior informed by experimental studies suggests much stronger evidence for predictability of pandemics, with the spillover of strains with pandemic potential from the animal reservoir to humans likely to be a history-dependent process. This would be consistent with strains needing to accumulate specific mutations and predictions of, for example, the recycling theory [194, 195], with the history-dependent process arising as a result of population immunity to prior strains. Further support is given by previous studies determining that all the emergent viruses that caused the post-1900 influenza pandemics resembled those that had circulated previously within the lifespan of then-living people [207]. Specifically, the 1968 pandemic virus was similar to the one that caused the 1889 pandemic [194, 195], while the H1 2009 pandemic virus was most similar to viruses that circulated before 1947 [208]. Such an immunity argument could also explain why there was very strong support for Ebola outbreak inter-event times being memoryless, with each of these outbreaks not being capable of establishing population-wide immunity due to being relatively small and having a high mortality rate [205]. With

the majority of the population remaining susceptible, we would expect the likelihood of the next outbreak occurring within a given time (originating from zoonotic transmission events from animal hosts) to be unaffected by the time passed since the preceding outbreak.

In the case of influenza, this suggests there would be significant merits in having active surveillance in humans and zoonotic hosts to reveal the diversity of influenza strains in circulation and the immunological profile of the population. Current immune landscape studies, such as Fluscape [209, 210] and seroprevalence analyses of live bird market workers [134], provide example frameworks that could be extended to investigate the immune profiles of the population to strains widely circulating in livestock populations. Such knowledge would aid classification of currently circulating strains with respect to their pandemic potential, determining those that pose a considerable threat. Going a step further, earlier detection of strains with pandemic causing capability could allow the containment of the strain to a localised area through, for example, more aggressive culling controls and movement restrictions in farmed poultry. Additionally, active surveillance outputs may inform the composition of influenza virus vaccines to be developed and administered to the at-risk human population as part of pandemic preparedness procedures. However, current influenza virus vaccines have limitations, requiring frequent updates to reflect the antigenic changes that occur in the pool of circulating virus strains. Circumventing this drawback is a key area of ongoing research through the development of 'universal' influenza vaccines that are less sensitive to the antigenic evolution of the virus, therefore giving broader protection against emergent strains [211, 212].

For all three suggested pandemic influenza histories, forward simulations implied a wide variability in the possible number of expected pandemics in the next 100 years following the last pandemic. Although Timeline A offers greater support for there being two or fewer pandemics between 2010-2110 compared to either Timeline B or C, additional virological information is required to determine which of these timelines is representative of the actual history of pandemic influenza. If this is achieved, the relevant timeline findings can be used to quantify the risk of a set number of pandemic events occurring in the next 100 years. Note that these predictions could be viewed as a worst case scenario, due to not accounting for the impact of future intervention measures. We anticipate however that the implementation of interventions that successfully determine and prevent the spread of strains posing the largest pandemic risk, possibly through the strategies discussed above, will di-

minish the likelihood of them realising their full pandemic potential. The number of influenza pandemics would drop as a result.

Our modelling framework posed the following potential issues. First, there was data uncertainty as a result of conflicting accounts of what were true historic influenza pandemics within the literature. As a consequence, to account for a range of possible pandemic histories, we constructed our three distinct pandemic timelines with strict, moderate and relaxed inclusion criteria based on the amount of corroboration across sources for an outbreak being a pandemic. Second, result robustness is likely weakened due to the small nature of the dataset and sensitivity with respect to the timeline used. We argue, however, that the observed dates of influenza pandemics are the ultimate test of our understanding of novel strain emergence, and with this data the more complex hypothesis was still found to be preferred when including a varied number of pre-20th century pandemics. Third, for the modelling framework a number of subjective decisions had to made. For our history-dependent model, alternative distributions could have been selected (e.g. Weibull distribution), though due to any phase-type distribution arising from a realistic combination of events being very close to a gamma distribution we only considered gamma distributions as representations of history dependence here. Additionally, other priors for κ could be used since, as expected, our prior specification for κ impacts on the posterior support for the history-dependent model. Although our 'strongly mechanistic' prior may be overly restrictive in assuming low credibility for one influenza pandemic being able to immediately follow another, such assumptions were not made with the 'weakly mechanistic' prior. Further, it is noteworthy that the weakly informed prior had no qualitative impact on the findings when compared to the findings using the 'strongly mechanistic' κ prior.

In summary, our analysis of influenza pandemic waiting times found support for the hypothesis of influenza pandemic emergence being history-dependent, rather than a memoryless process, under eight out of nine sets of modelling assumptions. Although the approach utilised here is reliant on limited and uncertain data, it corroborates findings from indirect measurements gathered via expensive and sometimes dangerous laboratory and field surveillance efforts that aim to further our understanding of pandemic emergence.

Chapter 6

Modelling the spread of mood in adolescent friendship networks

6.1 Introduction

In chapters 3 to 5 we used both pre-existing and novel infectious disease mathematical models to answer epidemiological questions concerning influenza A. While standard epidemic models are well established in the mathematical modelling of infectious disease, there is an increasing usage of these techniques for modelling other complex spreading processes that occur among society. One particular application is to investigate the possible spread of behavioural-linked health problems, with this chapter focusing on addressing this question with regards to depressive disorders.

Depression and other mood disorders are major and growing contributors to mortality and morbidity worldwide [62]. These mood disorders are widespread, with the World Health Organisation estimating that, globally, there are currently more than 300 million people affected by depression [63, 64, 213]. There is evidence that social support is important for the mental well-being of adolescents [214] and that befriending can have a positive effect on mental health [215]. Recent experiments suggest that peoples expression of negative or positive emotions is influenced by the level of negative or positive news from their friends and associates [216]. An improved understanding of the social processes that drive the epidemiology of these disorders therefore has the potential to bring highly significant public health benefits.

It is now very common to model infectious diseases as spreading processes on net-

works [77]. This approach is increasingly applied to behaviours, for example those related to infectious risk [3], and non-infectious diseases that are linked to behaviours that can spread socially (e.g. obesity and smoking [78, 79]).

Previous work relating to spread of depression on friendship networks has generally made at least one of the following key assumptions: (i) low mood and / or depression spreads like an infectious agent; (ii) healthy mood (non-depression) does not spread like an infectious agent; (iii) the information to distinguish between transmission and no-transmission models can be found in differences in static network measures such as clustering of disease [7, 8, 217], or in coarse population-level measures such as web-search over time [218]. Here we allow more flexibility in our model by making no prior assumption as to whether it is low mood or healthy mood that spreads. In addition, we use the dynamical behaviour of mood over time, allowing us to distinguish directly between transmission and no transmission.

6.2 Methods

6.2.1 The data

We consider data from the in-home interview survey of the Add Health study [219], which recorded adolescents in-school friends in addition to their CES-D (Center for Epidemiologic Studies Depression Scale) scores [220]. The CES-D scale is a 20-item measure that asks the respondent to rate how often over the past week they experienced symptoms associated with depression, with the in-home interview survey including 18 of these 20 questions. Four possible response options are given for each item, worth a score of 0, 1, 2 or 3, with a higher score indicating more severe depressive symptoms. The total score across all items was used to classify individuals as either having depressive symptoms (low mood) or not being depressed (healthy mood) according to the score cut-off associated with a clinical diagnosis of depression [221]. For the complete 20-item measure the thresholds found were 24 and 22 for females and males respectively. To compensate for the in-home interview survey including only 18 of the 20 items, we used modified thresholds of 22 for females and 20 for males.

We took data from two time points (waves), 6 to 12 months apart, from students in saturated schools (in a saturated school all students were selected to have an in-home interview, eliminating selection bias). In each wave, respondents were asked to either nominate up to five male and five female friends, or up to one male and one

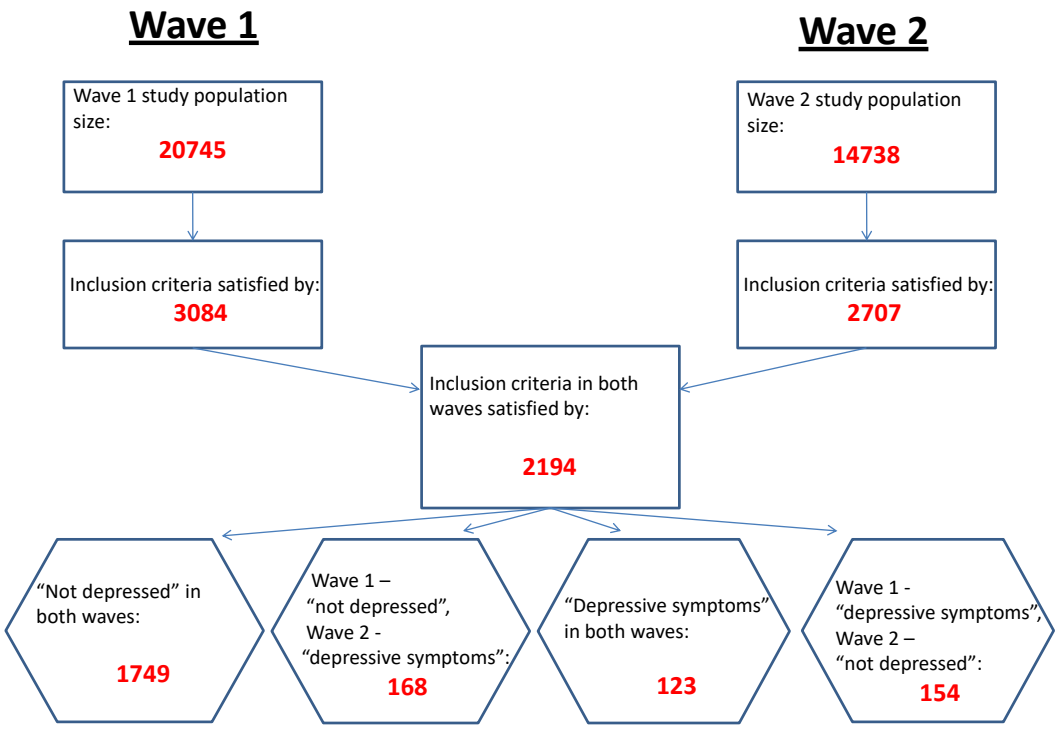

Figure 6.1: Sampling flow diagram for our study sample. At the top is the whole study population for wave 1 and wave 2, with the hexagon boxes giving the sample from the whole population that we used.

female friend. To be included in our study sample, for both time points the student must have provided complete answers to all the CES-D survey related questions and be the least restricted in the number of school friends they were allowed to give (i.e. allowed to list up to five male and five female friends). These respondents could be split into four groupings that cover all possible mood state combinations across the two waves (figure 6.1).

6.2.2 Model construction, fitting and selection

Throughout, we write N for healthy mood and D depressive symptoms. Letters $A, B, \cdots \in \{N, D\}$ and overlining is used as follows:

$$\overline{A} = \begin{cases} N & \text{if } A = D, \\ D & \text{if } A = N. \end{cases} \tag{6.1}$$

We let individuals be labelled with indices $i, j, \cdots \in \{1, \ldots, n\}$. At (discrete) time t individual i has state $X_i^t \in \{N, D\}$. From the friendship data, the respondents were connected in a network with adjacency matrix $\mathbf{G}$ with elements

$$G_{ij} = \begin{cases} 1 & \text{if individual } i \text{ named } j \text{ as a friend,} \\ 0 & \text{otherwise.} \end{cases} \tag{6.2}$$

We modelled depression status as a discrete-time Markov chain. This model is specified by two probabilities: the probability $p = Pr[X_i(t+1) = D | X_i(t) = N]$ of developing depressive symptoms, and the probability $q = Pr[X_i(t+1) = N | X_i(t) = D]$ of recovering from depression. Following Centola and Macy [90, 222], we considered a model in which these probabilities depend on the number of friends of an individual who have value N or D, with this dependence taking the form of an S-shaped function. These models are referred to as N-transmits and D-transmits respectively. Formally, our general N-transmits model was a discrete-time Markov chain $\mathbf{X}^t = (X_i^t)$ with transition probabilities

$$\begin{aligned} \mathbb{P}[X_i^{t+1} = D | X_i^t = N] &= p_{\sum_j G_{ij} I\{X_j^t = N\}}, \\ \mathbb{P}[X_i^{t+1} = N | X_i^t = D] &= q_{\sum_j G_{ij} I\{X_j^t = N\}}, \end{aligned} \tag{6.3}$$

where I is an indicator function

$$\mathrm{I}\{\omega\} = \begin{cases} 1 & \text{if } \omega \text{ is true,} \\ 0 & \text{otherwise.} \end{cases} \tag{6.4}$$

In the D-transmits model we exchange D for N on the right-hand side of (6.3).

We fit the above models to the Add Health data moving from wave 1 to wave 2, and compare to the no-transmission model where the probabilities do not depend on the mood states of an individual's friends (i.e. p_k and q_k are independent of k). Parameter values for our transmission and no-transmission models were found using Maximum Likelihood Estimation (MLE) by minimising the negative log-likelihood $\log(L)$ with respect to p and q using the MATLAB fmincon() function. The following likelihood function was constructed for the development of depressive symptoms scenario, with respect to either the number of N or D friends,

$$L(y|p, N) = \prod_k \binom{N_k}{y_k} p_k^{y_k} (1 - p_k)^{N_k - y_k},$$

where y_k was the number of respondents with k not depressed friends (friends with depressive symptoms) who were classified as not depressed at the first time point and having depressive symptoms at the second time point. N_k was the total number of respondents classified as not depressed at the first time point with k not depressed friends (friends with depressive symptoms). An equivalent likelihood function was constructed for the recovery from depressive symptoms scenario,

$$L(y|q,N) = \prod_k \binom{D_k}{y_k} q_k^{y_k}(1-q_k)^{D_k-y_k},$$

with y_k corresponding to the number of respondents with k not depressed friends (friends with depressive symptoms) who were classified as having depressive symptoms at the first time point and not depressed at the second time point. D_k was the total number of respondents classified as having depressive symptoms at the first time point with k not depressed friends (friends with depressive symptoms).

Confidence intervals were obtained through calculation of the Hessian matrix at the MLE parameters and use of standard asymptotic formulas. Competing models were assessed using Akaike information criterion (AIC) [201], with a description of this measure outlined in section 5.2.2.

6.2.3 Simulation outline

The Markov chains defining each of our candidate models could be simulated using Monte Carlo methods. Since we consider a situation where

$$0 < p_k, q_k < 1, \forall k, \tag{6.5}$$

we can get from any state to any other in a finite number of steps (the chain is irreducible) and the expected time to return to any state will be finite (all state are positive recurrent). Hence by e.g. Theorems (6.4.3) and (6.4.17) of Grimmett and Stirzaker [223], there will be a unique stationary distribution π that describes the behaviour of the chain at large times.

We performed discrete-time Monte Carlo simulations of the no-transmission and N-transmits models on a directed network of named friends constructed from the 3084 individuals in the dataset satisfying our inclusion criteria at the first time point (wave 1). Depending on the simulated output required, we either took 10^5 time-separated samples of node-level values at a single time point or 10^4 temporally

adjacent node-level state transitions from the stationary distribution for each model.

6.2.4 Goodness-of-fit

Static summary statistics

Performing simulations as detailed above (section 6.2.3), we took 10^5 independent samples from the stationary distribution for each model to calculate static summary statistics of interest (prevalence of D individuals, prevalence of various network pairs, stratification by contacts). We assessed uncertainty in the observed quantities through bootstrapping. The Bonferroni method was used to account for multiple testing of statistically significant differences between models and observed data [224].

Calculating p-values required comparison of Monte Carlo simulation output with uncertain data. We used the expression

$$p = \sum_x 2\pi_x \times \min(E_x, 1 - E_x),$$

where π_x is the density of the value x in the bootstrap sample from data and E_x is the empirical cumulative distribution function of the simulation output.

Residual errors

For logistic regression a standard approach to assess goodness-of-fit is the Hosmer-Lemeshow (HL) test, which is based on the distribution of residual errors [225] - i.e. the differences between the observed and the model values. Our model was not a standard regression, and so we tested goodness-of-fit in a similar manner to the HL test but with assumptions more appropriate for our model. In particular, we defined a residual error function stratified by number of friends,

$$\varepsilon_A = \left(\sum_{k=0}^{10} \left(Y_k^{A\to\overline{A}} - X_k^{A\to\overline{A}}(\theta) \right)^2 \right)^{1/2}, \tag{6.6}$$

where $Y_k^{A\to\overline{A}}$ is the observed number of state transitions from A to $\overline{A}$ of individuals with k friends in state N or D, and $X_k^{A\to\overline{A}}(\theta)$ is the modelled number of such events given parameters θ; explicitly

$$Y_k^{A\to B} = \sum_i \mathrm{I}\{\textstyle\sum_j G_{ij} = k\} \mathrm{I}\left\{Y_i^{t+1} = B\right\} \mathrm{I}\left\{Y_i^t = A\right\}, \tag{6.7}$$

and

$$X_k^{N\to D} = p_k \times \sum_i \mathrm{I}\{\textstyle\sum_j G_{ij} = k\}\mathrm{I}\left\{X_i^t = N\right\},$$
$$X_k^{D\to N} = q_k \times \sum_i \mathrm{I}\{\textstyle\sum_j G_{ij} = k\}\mathrm{I}\left\{X_i^t = D\right\}. \tag{6.8}$$

The quantity ε_A is positive definite and will tend to zero for a model that perfectly captures the data.

The distribution for ε_A was not analytically available. Thus, a parametric bootstrap approach was used, simulating from the model once it had been fitted to observed data by maximum-likelihood estimation (MLE), giving MLE parameter estimates $\hat{\theta}$.

We performed simulations as detailed in section 6.2.3. We took 10^4 temporally adjacent node-level state transitions from the stationary distribution for each model, extracting the proportion of individuals who recover from depressive symptoms/develop depressive symptoms within a year dependent on the number of friends they had of state N or D at the initial time point (as appropriate for the calculation of interest). This sampling process was repeated many times as for other bootstrap methods to obtain an accurate estimate of the distributions of ε_D and ε_N.

6.2.5 Parameter identifiability

We now turn to the question of how accurately model parameters can be inferred from data. To analyse this we performed model simulations as described in section 6.2.3, taking 10^4 temporally adjacent node-level state transitions from the stationary distribution. Each set of simulated data was then fitted to the same model that it had been generated from using MLE.

6.3 Results

6.3.1 Fitted parameter values

No-transmission models

We obtained the following no-transmission deterioration model for transitioning from healthy mood to low mood within a year (figures 6.2(a) and 6.2(c)),

$$p_k = 0.088[0.075, 0.10],$$

and no-transmission recovery model for transitioning from low mood to healthy mood within a year (figures 6.2(b) and 6.2(d)),

$$q_k = 0.56[0.50, 0.61].$$

N-transmits models

We obtained the following N-transmits deterioration model for transitioning from healthy mood to low mood within a year,

$$p_k = \alpha + \beta \sum_{l=0}^{k} \binom{10}{l} \gamma^l (1-\gamma)^{10-l},$$

with $\alpha = 0.13[0.088, 0.16], \beta = -0.064[-0.10, -0.025]$ and $\gamma = 0.25[0.070, 0.44]$ (figure 6.2(c)). Note that here and elsewhere numbers such as 10 appear as the limits in the data on number of friends; k is the number of friends in the transmitting state, and the parameters estimated are a simple way to parameterise a discrete sigmoidal function as suggested by a complex contagion model. The N-transmits recovery model, for transitioning from low mood to healthy mood within a year, was

$$q_k = \alpha + \beta \sum_{l=0}^{k} \binom{10}{l} \gamma^l (1-\gamma)^{10-l},$$

with $\alpha = 0.53[0.46, 0.59], \beta = 0.46[0.099, 0.83]$ and $\gamma = 0.71[0.45, 0.98]$ (figure 6.2(d)).

D-transmits models

We obtained the following D-transmits deterioration model for transitioning from healthy mood to low mood within a year,

$$p_k = \alpha + \beta \sum_{l=0}^{k} \binom{4}{l} \gamma^l (1-\gamma)^{4-l},$$

with $\alpha = 0.087[0.074, 0.10], \beta = 0.021[-0.27, 0.31]$ and $\gamma = 0.84[-0.62, 2.3]$ (figure 6.2(a)). The D-transmits recovery model was

$$q_k = \alpha + \beta \sum_{l=0}^{k} \binom{4}{l} \gamma^l (1-\gamma)^{4-l},$$

with $\alpha = 4.1 \times 10^{-5}[-5.0, 5.0], \beta = 0.64[-4.6, 5.9]$ and $\gamma = 0.050[-0.42, 0.52]$ (figure 6.2(b)).

6.3.2 Model comparisons

We fit our proposed models to the $n = 2194$ data points given by our inclusion criteria. For the dependence of probabilities p and q on the number of friends with depressive symptoms (no-transmission model against D-transmits model), AIC values (computed using equation (5.1)) showed the no-transmission model was the preferred choice (figures 6.2(a) and 6.2(b)). When considering the no-transmission model against the N-transmits model, the N-transmits model was the preferred choice in both cases (figures 6.2(c) and 6.2(d)).

6.3.3 Goodness-of-fit

Static summary statistics

Comparing D prevalence and edge type summary statistics to those obtained for the observed data there were significant differences between the no-transmission model and data, while the N-transmits model had no statistically significant differences (figure 6.3). In particular, the $N \to N$, $N \to D$ and $D \to N$ edge statistics (where $A \to B$ denotes an individual in state A naming an individual in state B as a friend) were found to be statistically significantly different between the no-transmission model and the data (figure 6.3(b-d)). The observed stratification by contacts could be plausibly generated by both the no-transmission and N-transmits models (figure 6.4).

Residual errors

For the following comparisons between the observed data residual error values and the parametric bootstrap residual error distributions special attention should not be paid to any particular threshold of p value; rather, a larger p value simply denotes a better fit. The calculated observed data residual error values could be plausibly generated by the N-transmits recovery model, $p = 0.944$ (figure 6.5(a)), and N-transmits deterioration model, $p = 0.942$ (figure 6.5(b)). In contrast, while the no-transmission recovery model could plausibly generate the calculated observed data residual error values, $p = 0.816$ (figure 6.5(c)), there was less support for the no-transmission deterioration model, $p = 0.153$ (figure 6.5(d)). The analysis of this simple summary statistic, while not intended to provide a criterion for model selection in the way that AIC does, supports our broad conclusion that N-transmits should be preferred to no transmission.

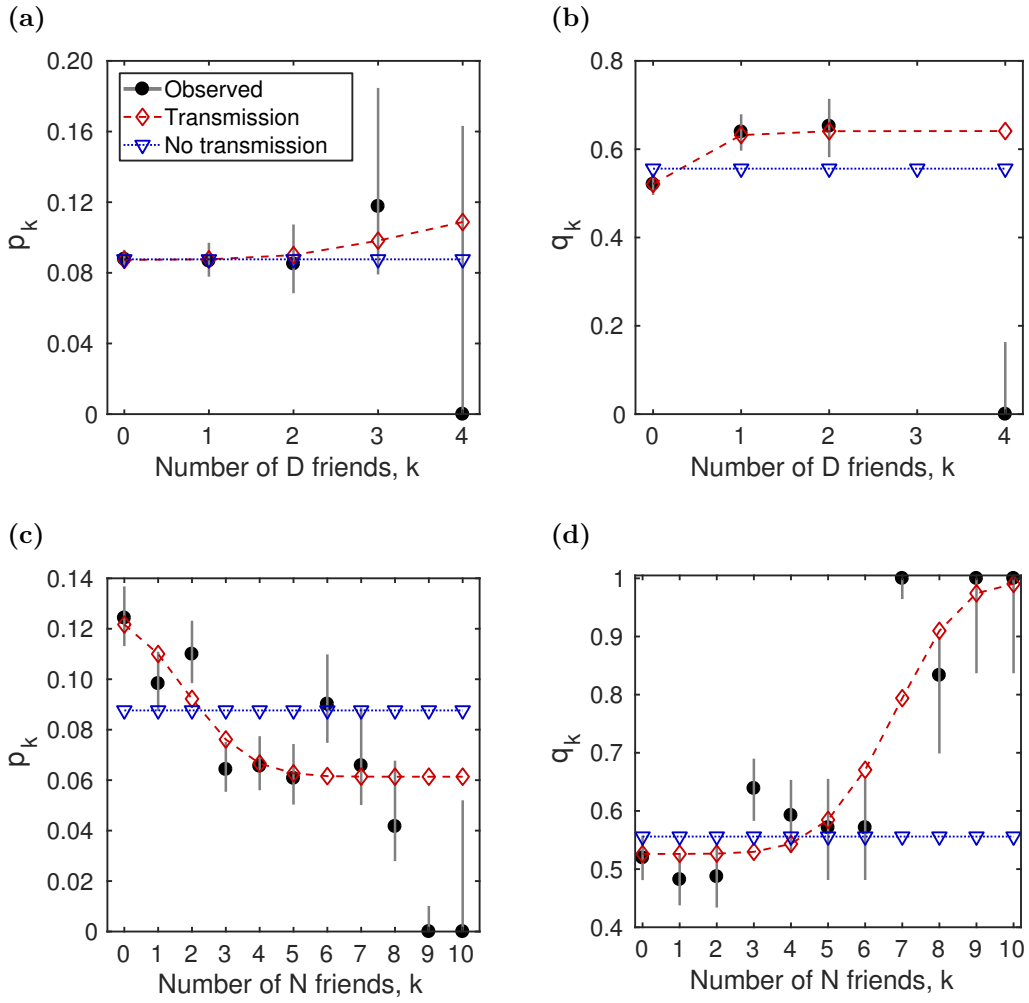

Figure 6.2: Dynamical behaviour of depression status between samples as a function of N friends or D friends for observed data (and 50% CI), transmission and no-transmission models. Uncertainty in the observed values was quantified using Jeffreys intervals [226]. The ΔAIC value is calculated by subtracting the no-transmission AIC value from the transmission AIC value. **(a)** Probability of transitioning from healthy mood to low mood against number of D friends - transmission is not preferred to no transmission (ΔAIC ≈ -4); **(b)** probability of recovering from low mood against number of D friends - transmission is not preferred to no transmission (ΔAIC ≈ -0.9); **(c)** probability of transitioning from healthy mood to low mood against number of N friends - transmission is preferred to no transmission (ΔAIC ≈ 8.4); **(d)** probability of recovering from low mood against number of N friends - transmission is preferred to no transmission (ΔAIC ≈ 4.5).

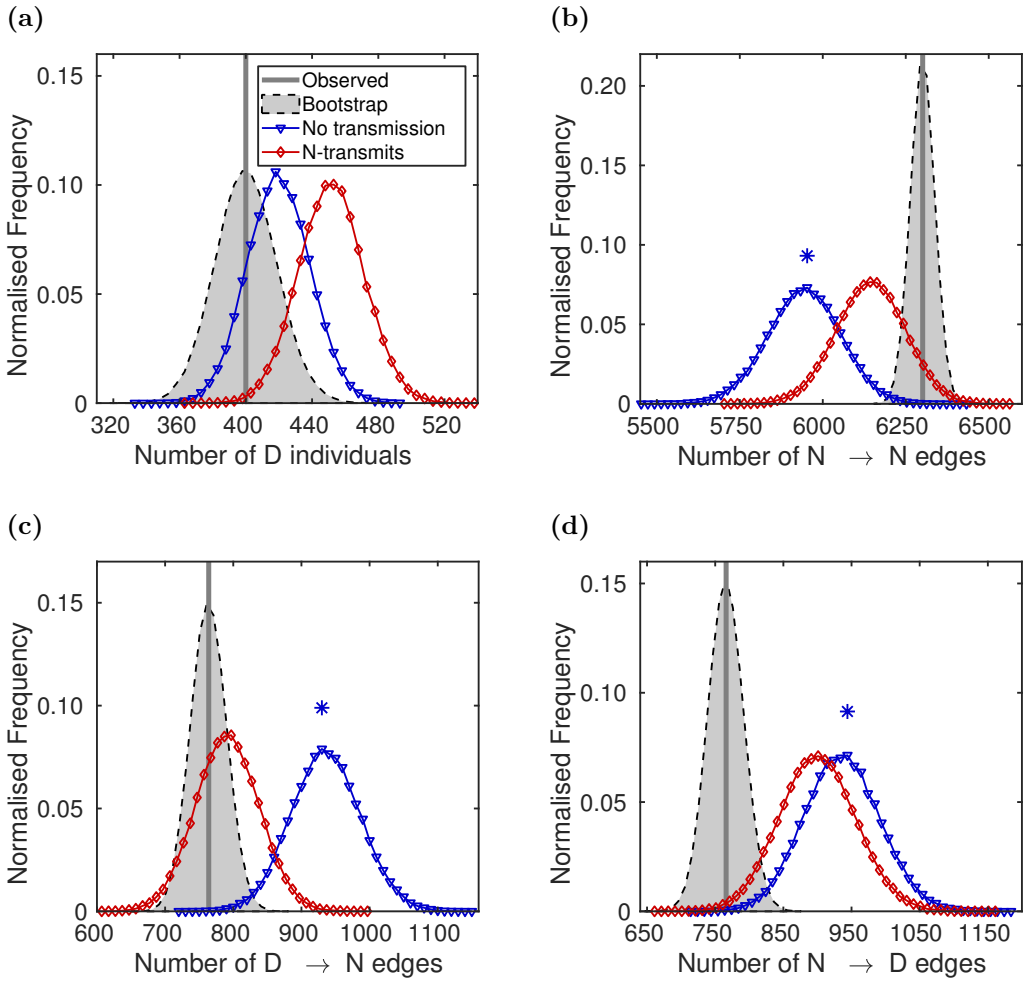

Figure 6.3: Static summary statistics for the stationary distributions of the models versus real data. Asterisks above a plot denote a significant statistical difference at the 5% level, corresponding to $p < 0.01$ using the Bonferroni method to account for multiple testing. **(a)** Prevalence of individuals with depressive symptoms - observed data could be plausibly generated by both N-transmits ($p = 0.058$) and no-transmission ($p = 0.41$) models. **(b)** Number of $N \to N$ edges - observed data could be plausibly generated by the N-transmits model ($p = 0.15$) but not by the no-transmission model ($p = 0.0014$). **(c)** Number of $D \to N$ edges - observed data could be plausibly generated by the N-transmits model ($p = 0.54$) but not by the no-transmission model ($p = 0.0035$). **(d)** Number of $N \to D$ edges - observed data could be plausibly generated by the N-transmits model ($p = 0.027$) but not by the no-transmission model ($p = 0.0067$). The fifth test and plot is for $D \to D$ edges (figure C.1). See section 6.2.4 for the p-value calculation method.

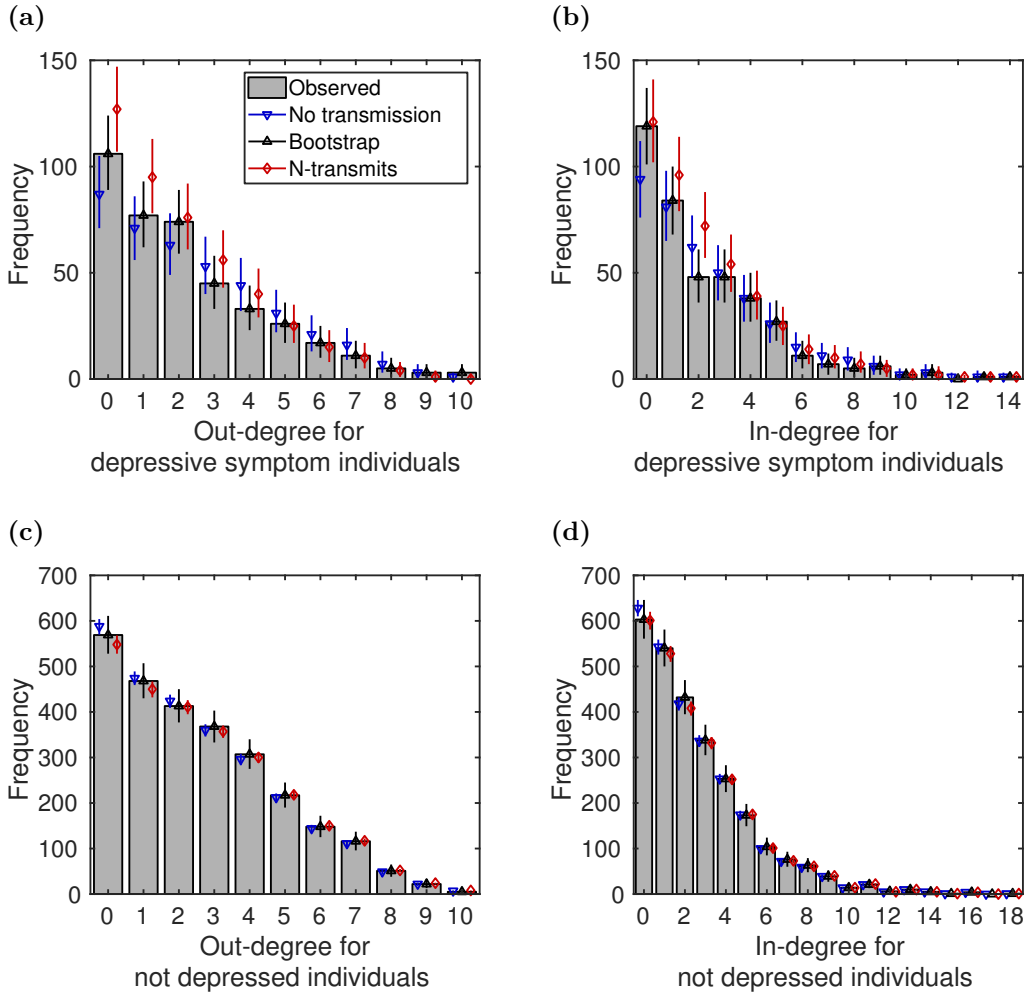

Figure 6.4: Stratification by contacts for the N-transmits and no-transmission models as compared to empirical degree distributions. Grey bars correspond to observed values, with black line segments giving bootstrapped 95% confidence intervals. Blue and red line segments give the 95% prediction intervals for the no-transmission model (median value represented by the upside down triangle) and N-transmits model (median value represented by the diamond) respectively. The observed stratification by contacts could be plausibly generated by both the no-transmission and N-transmits models. **(a)** Out-degree of individuals with depressive symptoms. **(b)** In-degree of individuals with depressive symptoms. **(c)** Out-degree of not depressed individuals. **(d)** In-degree of not depressed individuals.

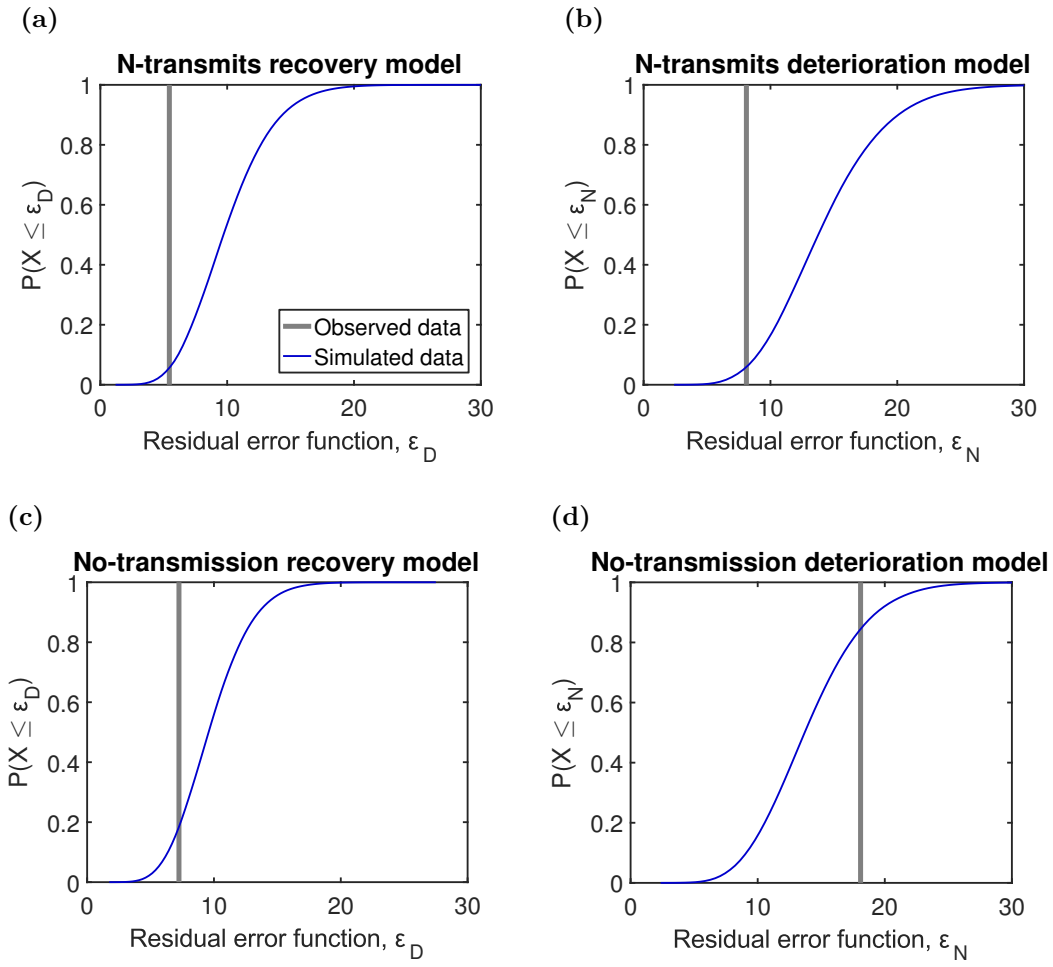

Figure 6.5: Cumulative distribution functions for residual errors, obtained via parametric bootstrapping versus the observed data residual error. **(a)** N-transmits recovery model residual error function cumulative distribution ($p = 0.944$). **(b)** N-transmits deterioration model residual error function cumulative distribution ($p = 0.942$). **(c)** No-transmission recovery model residual error function cumulative distribution ($p = 0.816$). **(d)** No-transmission deterioration model residual error function cumulative distribution ($p = 0.153$).

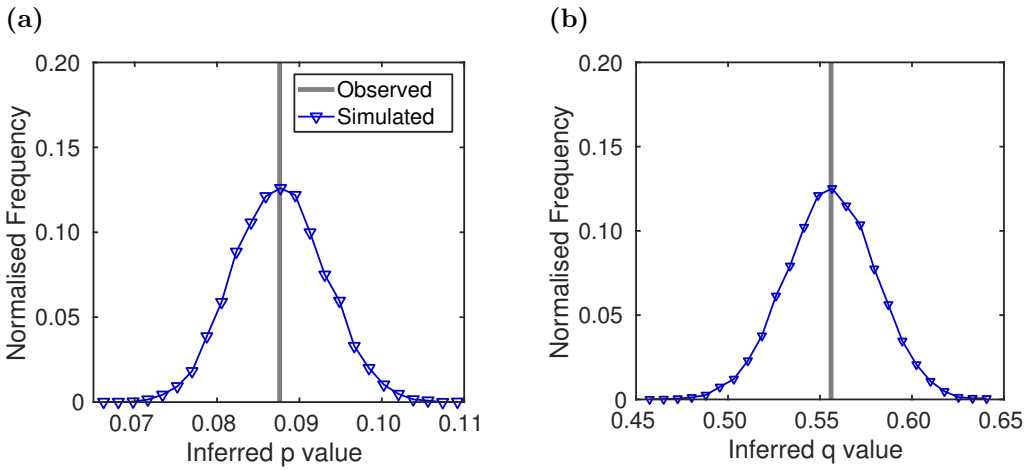

Figure 6.6: Normalised frequency of inferred no-transmission model parameters from 10^4 simulated samples. **(a)** Transition from N to D (p), and **(b)** transition from D to N (q) parameters inferred (by MLE) from simulated data versus chosen model values when fitted to the observed data. A high level of identifiability was observed for each parameter.

6.3.4 Parameter identifiability

Inferred parameter distributions for our three models refitted to simulated data were compared to the parameter values inferred when fitting to the observed data. A high level of parameter identifiability was observed for both the no-transmission (figure 6.6) and N-transmits models (figure 6.7), giving extra confidence to our results. Additionally, the majority of parameters for our D-transmits model could be reliably recovered (figure C.2).

6.4 Discussion

A major benefit of the dynamical approach that we have taken is that it avoids the problems of confounding that have been controversial in other studies of social contagion [8]. Figure 6.8 shows the model schematically, to provide intuitive insight into why this is the case.

In this model, there is transmission of D if the probability of event (3) happening (given the initial state) is bigger than the probability of event (1) happening (given the initial state): Pr(event (3)) $>$ Pr(event (1)). There is also transmission of D if Pr(event (6)) $<$ Pr(event (4)). We did not find evidence for transmission of D based on this criterion, as shown in figures 6.2(a) and 6.2(b). Such transmission would also be expected to lead to more $D \to D$ pairs and fewer $D \to N$ and $N \to D$ pairs

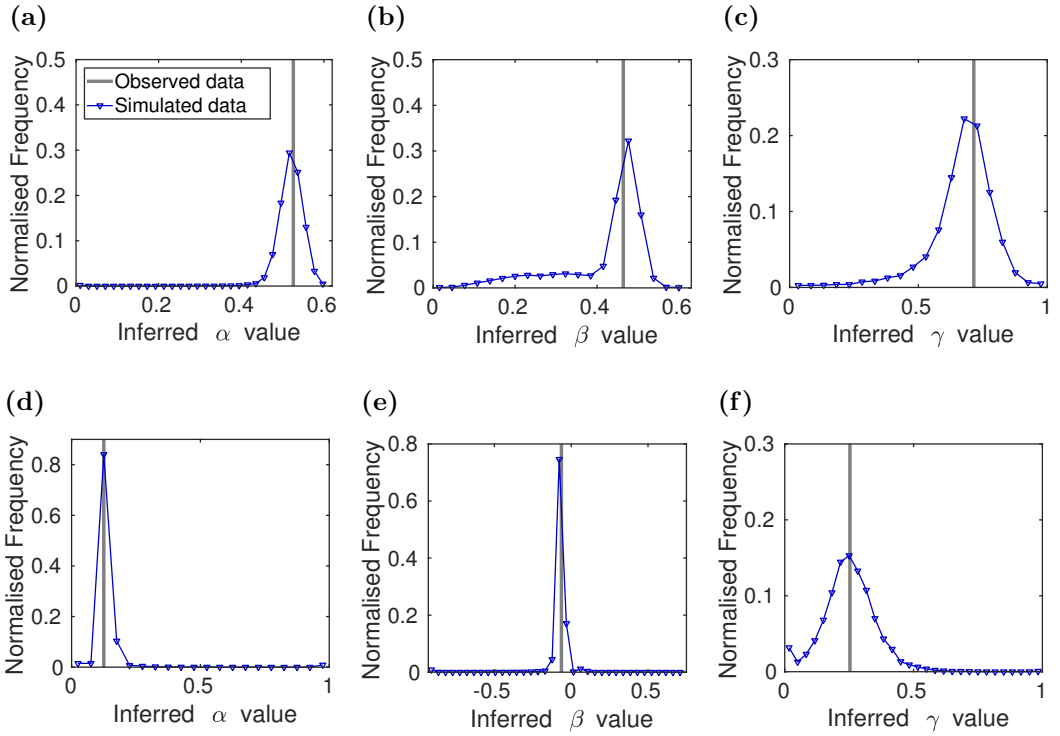

Figure 6.7: Normalised frequency of inferred N-transmits model parameters from 10^4 simulated samples. (a,b,c) Transition from N to D and **(d,e,f)** transition from D to N parameters inferred (by MLE) from simulated data versus chosen model values when fitted to the observed data. **(a,d)** α; **(b,e)** β; **(c,f)** γ. A high level of identifiability was observed for each parameter.

than a null model. This pattern was not observed (see figure 6.3 and figure C.1).

There is transmission of N in our model if Pr(event (2)) $<$ Pr(event (1)) and also if Pr(event (5)) $>$ Pr(event (4)). We found statistically significant evidence for transmission of N on the basis of this criterion as seen in figures 6.2(c) and 6.2(d). Such transmission would also be expected to lead to more $N \to N$ pairs and fewer $D \to N$ and $N \to D$ pairs than a null model. This pattern was observed (see figure 6.3).

Suppose that there is homophily (similar individuals naming each other as friends) at work in the friendship network, either in terms of depressive symptoms, or a latent variable that is correlated with depressive symptoms. This will tend to increase the number of $D \to D$ or $N \to N$ pairs in the absence of any transmission effect, meaning that these tests (shown in figure 6.3) can be confounded by homophily. When working with two waves of data, however, such homophily will simply lead to fewer individuals in the initial states associated with events (3) and (5) than events

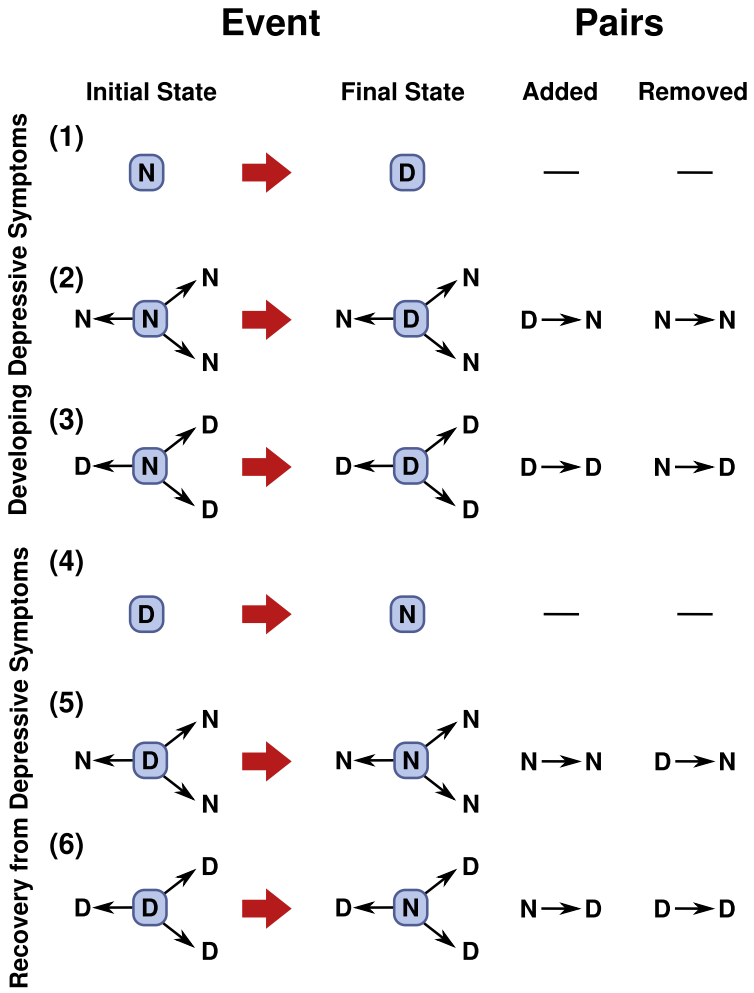

Figure 6.8: Pictorial representation of the possible events in our model. Developing or recovering from depressive symptoms; in the absence of friends, with friends with healthy mood, or with depressed friends. The changes in pairs produced are also shown.

(2) and (6), but in our approach we fit to the probability of moving to a final state given an initial state. This means that since there is still sufficient data to find a statistically significant effect, homophily cannot confound the results presented in figure 6.2. Our verbal argument here can be made in a more mathematically precise manner as shown in appendix C.1.

In summary, we have shown the epidemiological impact of such mood transmission in a large adolescent population, giving statistically significant evidence for spreading of healthy mood, but not for spreading of depressive symptoms. Once discovered, this behaviour is in fact plausible through a number of mechanisms. Depression has

been associated with social withdrawal [227], and so depressed individuals would be expected to exert less social influence than adolescents with healthy mood. However, each individual may need sufficient exposure to others with a healthy mood in order to stay well, or become well if depressed. In support of this, there is evidence from psychology of mechanisms by which mood is transmitted between people. Automatic transmission of mood between people has been demonstrated [228]. Further, unconscious mimicry enhances social rapport [229], while those feeling positive towards the person with whom they are interacting socially are more likely to mimic and so build rapport [230], thus giving the opportunity for transmission of healthy mood. People who are (or have a tendency to be) depressed are less able to maintain a positive outlook from moment to moment [231], a deficit potentially compensated by interaction with healthy friends.

The static network measures provide indirect evidence of spreading of healthy mood through analysis of clustering. It shows that the no-transmission model is significantly different from the observed data, while the data and N-transmits model are in agreement. Such clustering, while supportive of a transmission effect, can have other causes and so we recommend that future empirical work measures changes in mood over time where possible.

Our results offer implications for improving adolescent mood. In particular, they suggest the hypothesis that enabling networks of friendship between adolescents has the potential to reduce both incidence and prevalence of depression. Our complex contagion model suggests that adolescents with five or more healthy friends have half the probability of becoming depressed over a 6-12 month period compared with adolescents with no healthy friends, and that adolescents with 10 healthy friends have double the probability of recovering from depressive symptoms over a 6-12 month period compared with adolescents with three healthy friends. If such an effect were demonstrated in an intervention study it would massively outperform existing interventions.

Chapter 7

Conclusion and Outlook

In this thesis we have outlined our extensions to pre-existing epidemiological models, in addition to our development of novel mathematical modelling frameworks to be utilised in this field. These were applied to two epidemiological problems that greatly afflict public health; zoonotic influenza and depression. Our goal was to increase understanding of the mechanistic processes and determinants that drive these diseases and disorders, in addition to providing insights into how best to prevent and/or treat them.

In **chapter 2**, we stressed the need for building upon pre-existing mathematical modelling frameworks, in addition to developing new ones, to better capture the dynamics of the epidemiological problems of interest in this work. In particular, for zoonotic influenza there is a lack of specific, mechanistic frameworks with cross-species spillover transmission included, while studies of possible social contagion mechanisms require model concepts capable of distinguishing between transmission and non-transmission mechanisms. We also outlined here the array of parameter inference and simulation techniques that would be used to parameterise our models, verify their compatibility with the information available and, where appropriate, make predictions on the likely repercussions of imposing particular prevention and/or intervention strategies for the health-related event of interest, thus fulfilling each part of the mathematical modelling cycle (figure 1.1).

A question of ongoing concern is determining the regions that are likely sources of newly emerging influenza strains with pandemic causing potential. A suitable candidate is Bangladesh, due to being one of the most densely populated countries in the world and having an intensifying farming system. One particular subtype of

HPAI that is poses a persistent public health threat with the capability of causing infection in humans with a high mortality rate, in addition to negatively impacting the livestock industry, is H5N1. It is therefore vital to establish the key factors for this subtype, specific to Bangladesh, that enable both continued transmission within poultry and spillover across the human-animal interface. In **chapter 3** we developed and statistically fitted using a Bayesian inference scheme a modelling framework, consisting of a poultry transmission component and a zoonotic transmission component, to two H5N1 epidemics in the Dhaka region of Bangladesh (occurring from 2007 onwards). Analysis of our fitted models revealed a consistent outcome of non-optimal reporting of infected premises in each poultry epidemic of interest, while the zoonotic transmission component found the main contributor to spillover transmission differed from one poultry epidemic to another. This indicated that shortening delays in reporting infected poultry premises, alongside reducing contact between humans and poultry, may help reduce risk to human health.

Furthermore, simulation-based goodness-of-fit assessments unveiled that the successful identification of a poultry-to-poultry transmission model of minimal complexity to capture the size and spatial distribution of the outbreaks depended on the administration level being analysed.

Resolving the key transmission-dynamic mechanisms and risk factors behind historical disease epidemics can then inform what should be done in a future outbreak. We applied, in **chapter 4**, our previously fitted H5N1 influenza transmission models to examine how H5N1 disease management actions within the Dhaka division of Bangladesh should be prioritised and implemented when having to account for resource availability. In this exploratory analysis we highlighted how intervention selection may be influenced based on whether it is believed transmission is predominately premises-to-premises, allowing other factors to be omitted, or if importations and other external factors should be taken into account. Crucially, proactive active surveillance can seemingly reduce the scale and duration of an outbreak in either case. For the latter, in situations where transmission between premises is weak the enforcement of control measures not directly applied to poultry flocks themselves (e.g. public awareness campaigns) can dramatically cut the risk of spillover transmission of zoonotic influenza at the poultry-human interface taking place, leading to a reduction in H5N1 human case occurrence. To enhance control policy impactfulness, the suggestion is that much more attention should be directed towards identifying ways in which intervention efforts can be targeted and tailored to combat the trans-

mission dynamics specific to the outbreak, while also accounting for the availability of resources to ensure the chosen intervention measure can be capably implemented.

While the spillover of influenza strains with pandemic potential at the human-animal interface initially arises at a localised scale, such events pose a great risk to the health of the public at a global level. Understanding the biological mechanisms behind the appearance of strains that successfully cause global pandemics may subsequently help inform the type of interventions that would have the greatest impact in reducing the risk of pandemic emergence. To this end, we analysed in **chapter 5** the time periods between influenza pandemics since 1700 under different assumptions to determine whether the emergence of new pandemic strains is a memoryless or history-dependent process. Under a Bayesian model selection approach, eight out of nine sets of modelling assumptions gave support to the hypothesis of history-dependence. Although the approach we took relies on limited data, so is uncertain, it provides cheap, safe and direct evidence relating to pandemic emergence, a field where indirect measurements are often made at great risk and cost. In addition, we performed forward simulations using our fitted models and showed a high level of variability in the predicted number of pandemics from 2010-2110 across all modelling assumptions. Reductions in this uncertainty may occur as the legitimacy of disputed historic influenza pandemics is clarified.

In addition to infectious diseases, a growing affliction upon the health of modern society are non-communicable diseases that are linked to modifiable behaviours. This encompasses mental disorders such as depression, which is a major public health concern worldwide. There is evidence that social support and befriending influence mental health, and an improved understanding of the social processes that drive depression has the potential to bring significant public health benefits. The possible spread of behavioural-linked health problems, such as depression, are amenable to being represented and analysed with methodological approaches typically used to model infectious disease epidemics. Thus, in **chapter 6** we investigated whether there was evidence for transmission of mood on a school friendship network of adolescents, constructing a model framework with the flexibility of making no prior assumption as to whether it is low mood or healthy mood that spreads. We additionally utilised the dynamical behaviour of mood over time, with such an approach letting us distinguish directly between transmission and no transmission. We showed that while depression does not spread, healthy mood among friends is associated with significantly reduced risk of developing and increased chance of re-

covering from depression. In particular, the spreading of healthy mood could be captured using a non-linear complex contagion model. These findings suggest that promotion of friendship between adolescents can reduce both incidence and prevalence of depression.

The future directions we could take this research are wide-ranging, with a limited selection outlined here. We can extend our H5N1 model framework by incorporating additional complexity. Specific goals include treating layer and broiler poultry separately, explicitly incorporating live bird markets into the model framework, and the development of a spatially dependent zoonotic transmission component. We can then study whether proposed models incorporating these additional factors provide an improved fit to historical outbreaks. Our H5N1 influenza model framework may also be adapted and applied to other regions and/or strains, with H7N9 in China being a possible example. Furthermore, such additions to the model framework would let use explore the benefits and drawbacks of novel schemes for optimal intervention. Directions to pursue include accounting specifically for the role of live bird markets, the impact of intensification in farming practices, and exploring the relationship between control compliance and intervention effectiveness through the incorporation of opinion dynamics.

On the analysis of waiting times between influenza pandemics, as a definitive history of historic influenza pandemics is established we can refit our models to the available data. Subsequently, we may attempt to make quantitative predictions of the likelihood of a pandemic outbreak occurring within a specified period of time.

Research into mathematical modelling of social contagion may be taken forward in a couple of ways. The first is using the model framework to investigate whether there is evidence for other behaviour-linked health problems spreading in a contagion-like manner. We note that undertaking such analyses requires the availability of a suitable longitudinal dataset tracking the trait of interest, plus information on the social network. The second direction is further development of the mathematical tools to study such questions. While our work focused on a binary variable, a natural extension would be considering cases where more than two states are possible.

While there is still much to be done we have made contributory steps in a broad range of epidemiological topics, and we will endeavour to continue to help inform these goals going forward.

Appendix A

Appendix to chapter 3

A.1 Additional tables

Table A.1: Comparison of DIC values for the district datasets fitted to our models with various fixed infected to reporting times. For each model, the minimum DIC value across all possible infected to reporting times is highlighted in bold.

	Epidemic dates	**Inf. to rep. time**	**Model A**	**Model B**	**Model C**
Wave 2	Region	2 days	421.5	418.7	433.5
		4 days	419.5	419.1	439.1
		7 days	**413.8**	**410.8**	**409.3**
	Country	2 days	446.7	436.6	429.2
		4 days	444.1	437.3	434.9
		7 days	**430.0**	**423.3**	**413.0**
Wave 5	Region	2 days	451.9	452.6	470.4
		4 days	450.1	448.2	466.4
		7 days	**444.6**	**438.7**	**452.8**
	Country	2 days	462.9	461.6	473.6
		4 days	458.6	455.9	466.1
		7 days	**449.7**	**443.4**	**452.9**

Table A.2: Comparison of DIC values for the division datasets fitted to our models with various fixed infected to reporting times. For each model, the minimum DIC value across all possible infected to reporting times is highlighted in bold.

	Epidemic dates	Inf. to rep. time	Model A	Model B	Model C
Wave 2	Region	2 days	2156.2	2149.6	2105.2
		4 days	2150.3	2144.2	2104.6
		7 days	**2148.9**	**2142.1**	**2100.5**
	Country	2 days	2167.3	2158.3	2115.4
		4 days	2159.9	2152.7	2110.6
		7 days	**2157.2**	**2149.1**	**2104.9**
Wave 5	Region	2 days	1574.5	1543.5	1538.9
		4 days	1566.8	1521.1	1511.5
		7 days	**1557.5**	**1503.5**	**1499.9**

Table A.3: DIC values for the different data sets fitted to our nested models. For each model, the fixed infected to reporting time that minimised the DIC was used (see tables A.1 to A.2). The DIC value of the preferred model is highlighted in bold. Out of the two ecological models (D,E), if one is strongly preferred over the other (having a DIC value at least 3 lower) it is highlighted in italics.

	Epidemic dates	Model A	Model B	Model C	Model D	Model E
Wave 2 district	Region	413.8	410.8	409.3	411.1	***408.0***
	Country	430.0	423.3	413.0	414.4	**412.0**
Wave 2 division	Region	2148.9	2142.1	**2100.5**	2107.7	*2102.2*
	Country	2157.2	2149.1	**2104.9**	2109.1	2106.3
Wave 5 district	Region	444.6	**438.7**	452.8	443.8	*440.2*
	Country	449.7	**443.4**	452.9	448.5	*444.9*
Wave 5 division	Region	1557.5	1503.5	**1499.9**	1506.0	*1502.9*

Table A.4: Zoonotic transmission model parameter mean, median and 95% credible intervals (CI) fitted to both Dhaka district and division data using region specific epidemic dates. Distributions calculated using 4×10^4 samples obtained from MCMC, thinning by a factor of 40 samples with a burn-in period of 10^5 steps. Examples of MCMC diagnostic realisations when fitting to the wave 6 division data are presented in figures A.9 and A.10.

	Param.		**Wave 2**	**Wave 5**	**Wave 6**
District	β	Mean	2.56×10^{-6}	9.15×10^{-6}	3.58×10^{-6}
		Median	1.92×10^{-6}	7.72×10^{-6}	2.52×10^{-6}
		(95% CI)	(7.97×10^{-8},8.64×10^{-6})	(5.38×10^{-7},2.61×10^{-5})	(9.99×10^{-8},1.30×10^{-5})
	$\epsilon_\mathbf{h}$	Mean	0.0161	0.0229	0.0340
		Median	0.0124	0.0176	0.0300
		(95% CI)	(6.11×10^{-4},0.0749)	(7.16×10^{-4},0.0707)	(0.0043,0.0871)
Division	β	Mean	2.27×10^{-7}	1.12×10^{-6}	3.42×10^{-7}
		Median	1.69×10^{-7}	9.31×10^{-7}	2.38×10^{-7}
		(95% CI)	(6.96×10^{-9},7.79×10^{-7})	(5.19×10^{-8},3.23×10^{-6})	(9.27×10^{-9},1.26×10^{-6})
	$\epsilon_\mathbf{h}$	Mean	0.00889	0.0132	0.00780
		Median	0.00703	0.0103	0.00710
		(95% CI)	(3.48×10^{-4},0.0279)	(4.35×10^{-4},0.0425)	(0.0019,0.0176)

A.2 Transmission kernel figures

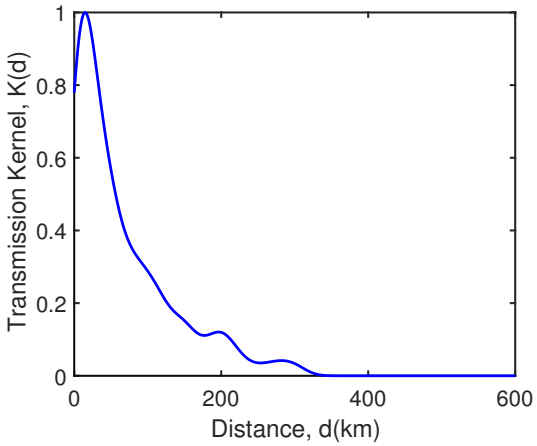

Figure A.1: Baseline model (model A) transmission kernel. The transmission kernel K as a function of distance (d), derived from the Dhaka division poultry case data using the following procedure. For each infected premises we found the nearest premises that was infected within the previous two days before that farm was reported. The distance between the infecting premises and the newly infected premises was calculated, with the process repeated for each infected premises.

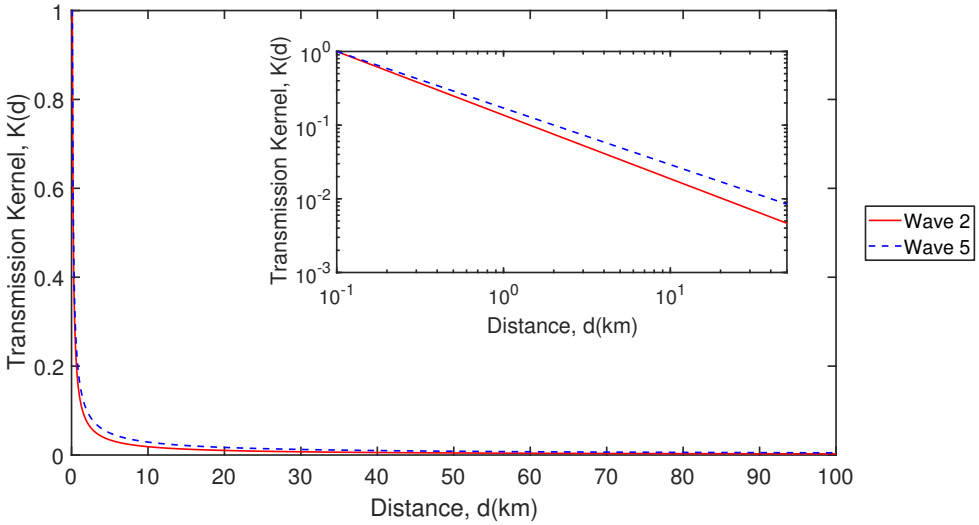

Figure A.2: Preferred district level model transmission kernels when fitting to country-wide epidemic dates. The transmission kernel K as a function of distance (d) for each of the district level models, using the median α parameter values inferred when performing the MCMC fitting procedure using country-wide epidemic dates (see table 3.5). The transmission kernel profiles exhibit similar characteristics across both epidemics waves. Comparable kernels were obtained when fitting to the district level data using region-specific epidemic dates.

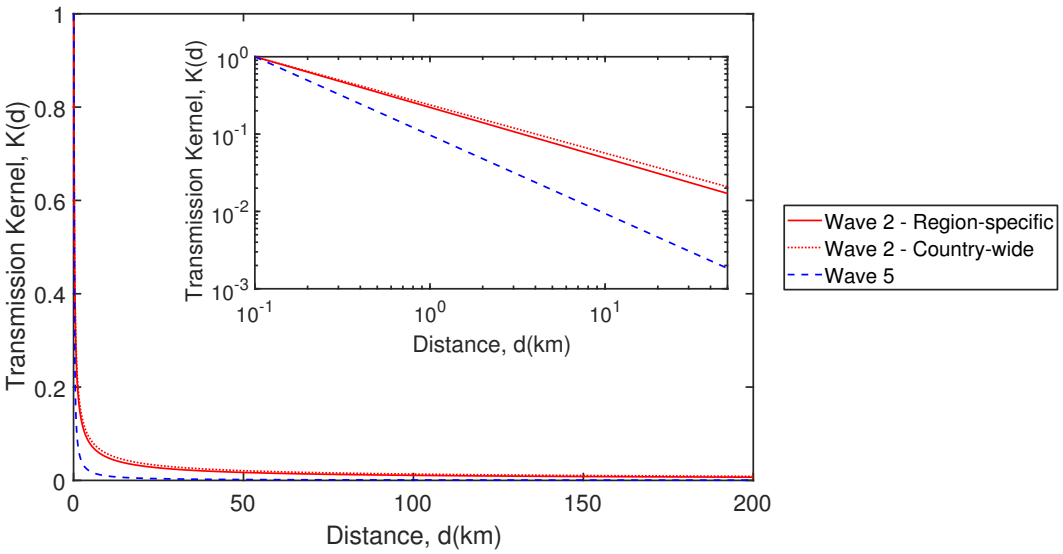

Figure A.3: Preferred division level dataset transmission kernels. The transmission kernel K as a function of distance (d) for each of the division level datasets, using the inferred median α parameter values (see table 3.6). The estimated transmission kernel for the wave 5 division dataset (blue dashed line) strongly preferred short-range transmission, while the kernels for both wave 2 division datasets attributed a higher relative likelihood to transmission events between premises separated by distances in excess of 10km (region-specific epidemic dates denoted by the red solid line, country-wide epidemic dates denoted by the red dotted line).

A.3 MCMC diagnostics

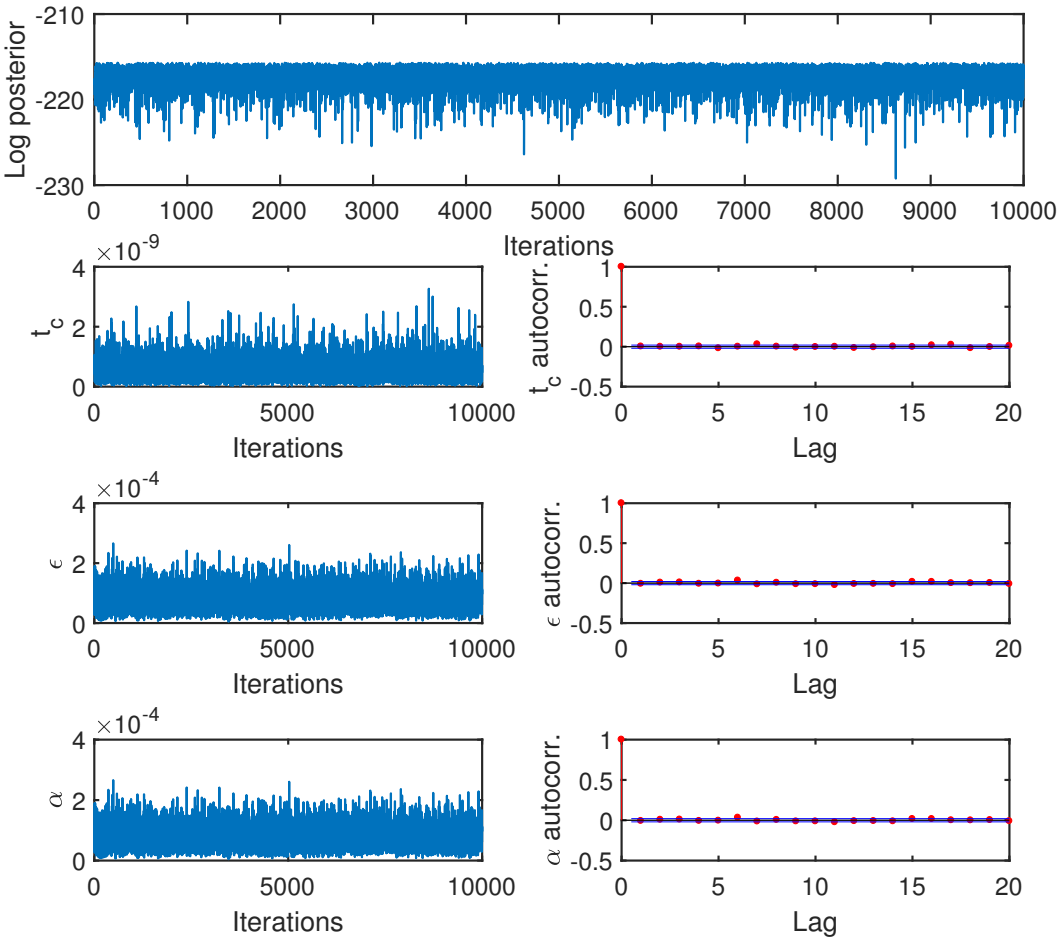

Figure A.4: MCMC scheme parameter trace and autocorrelation plots for poultry transmission model B fitted to the wave 5 district dataset using region-specific dates. Run for 10^6 iterations (ignoring a burn-in period of 10^5 steps), with these replicates thinned by a factor of 100 (to lower autocorrelation in the parameter chains) giving 10^4 samples. **Top row:** log posterior trace plot. **Rows two-four:** Trace plots (left column) and autocorrelation plots (right column) for the model parameters t_c (second row), ϵ (third row) and α (fourth row).

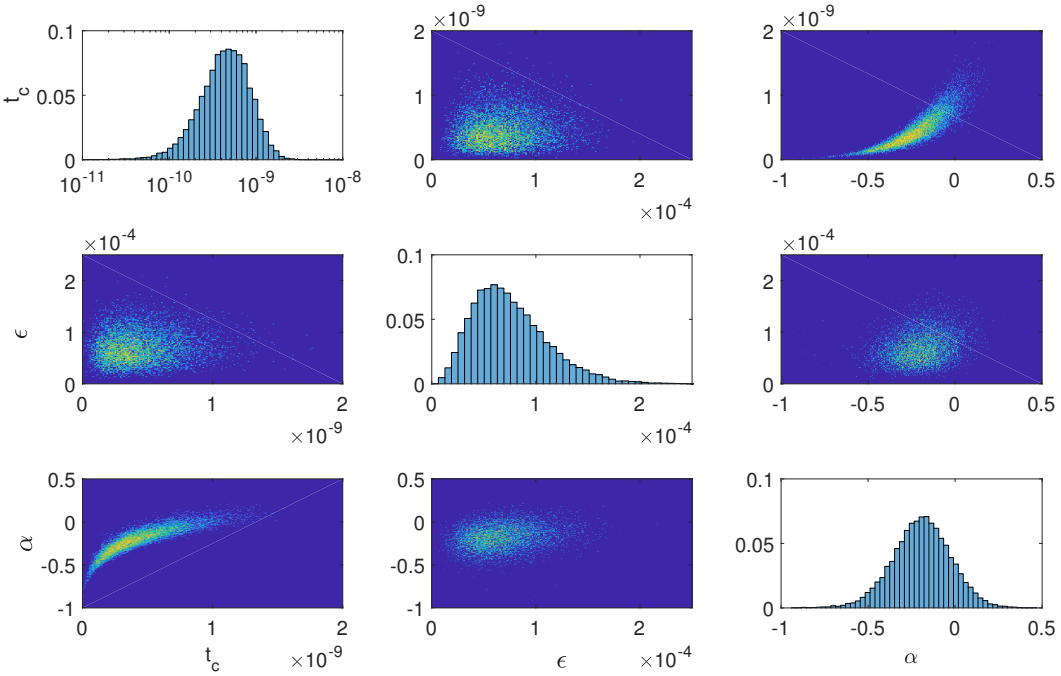

Figure A.5: Cloud plot for inferred parameter distributions when fitting poultry transmission model B to the wave 5 district dataset using region-specific dates. Histograms along the main diagonal panels give the proportion of parameter samples (generated via the MCMC scheme) within the specific bins. Off diagonal panels give two-dimensional histograms showing the density distribution of the data when plotted against one another, with lighter colours corresponding to higher densities.

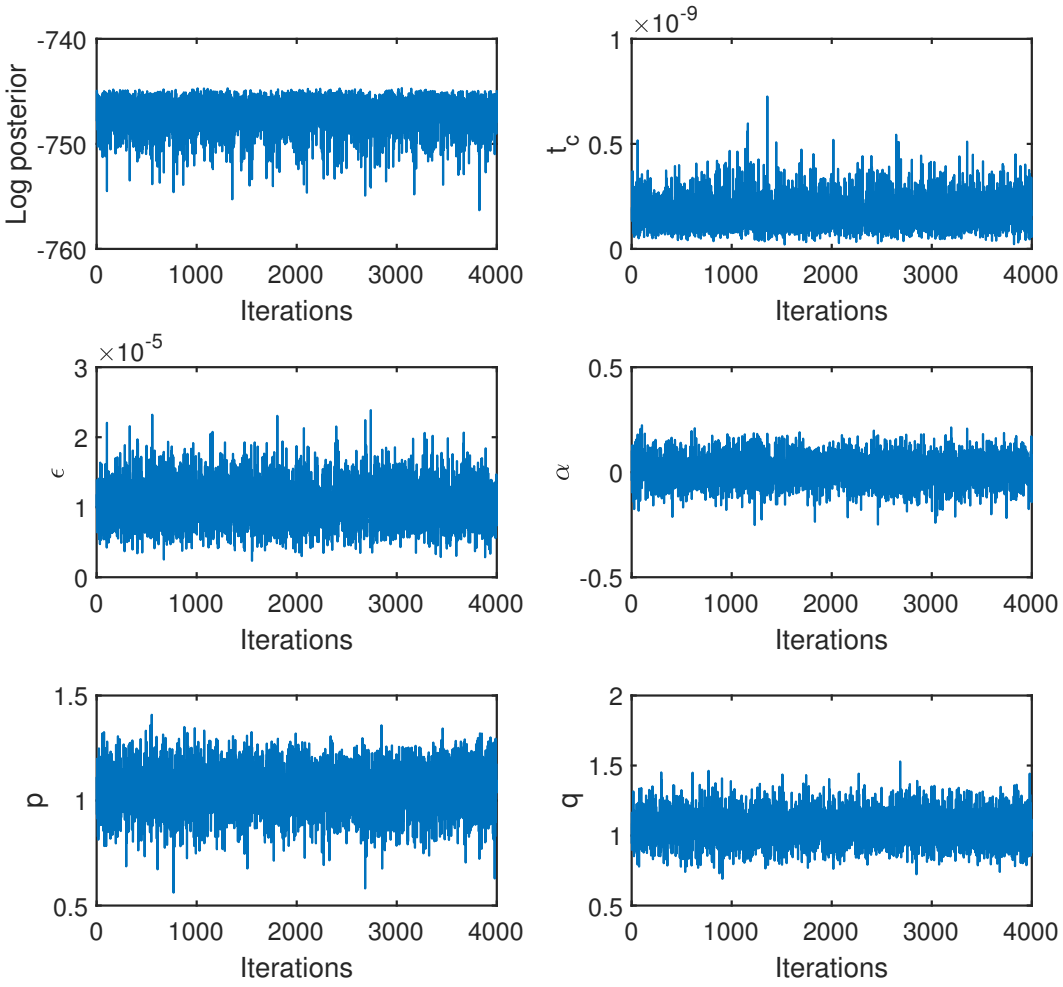

Figure A.6: MCMC scheme parameter trace plots for poultry transmission model C fitted to the wave 5 division dataset. Run for 4×10^5 iterations (ignoring a burn-in period of 10^5 steps), with these replicates thinned by a factor of 100 (to lower autocorrelation in the parameter chains) giving 4×10^3 samples. Trace plots are as follows: log posterior, t_c, ϵ, α, p and q.

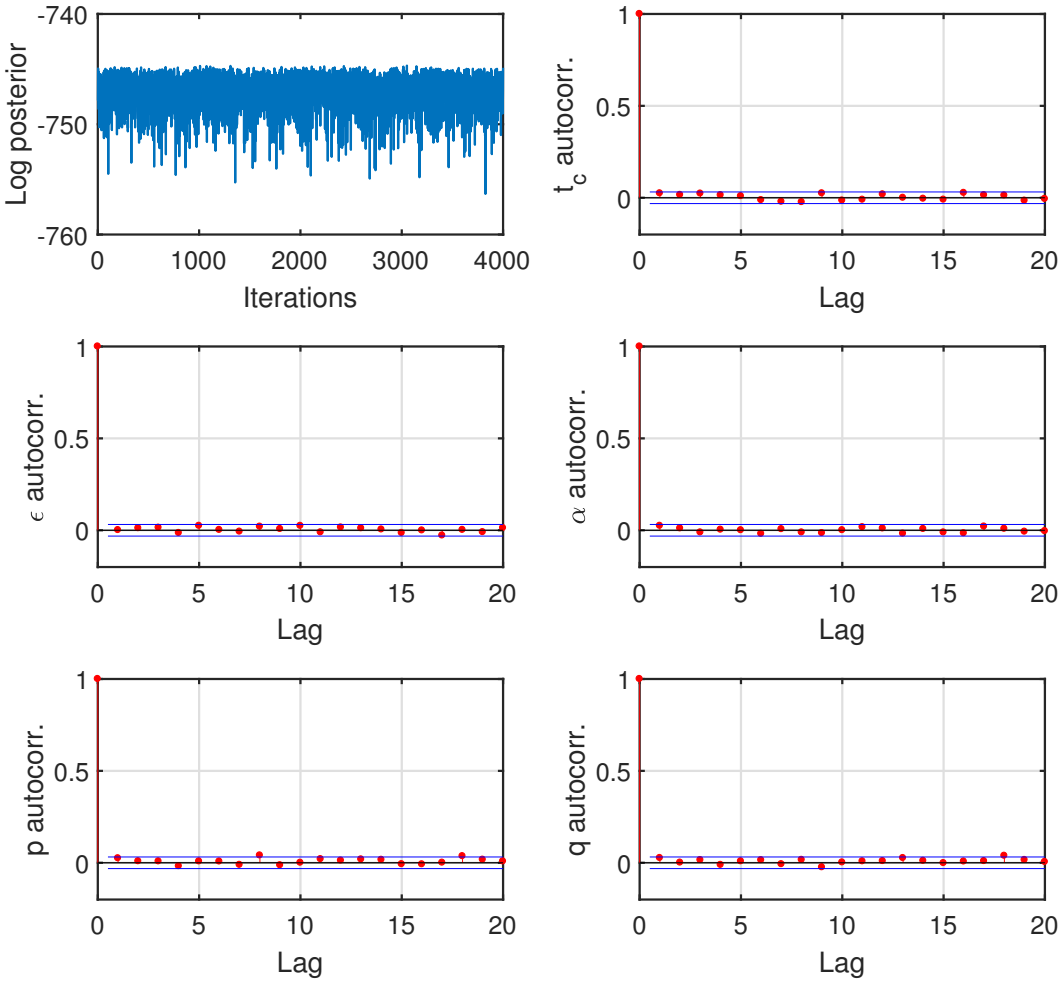

Figure A.7: MCMC scheme parameter autocorrelation plots for poultry transmission model C fitted to the wave 5 division dataset. Run for 4×10^5 iterations (ignoring a burn-in period of 10^5 steps), with these replicates thinned by a factor of 100 (to lower autocorrelation in the parameter chains) giving 4×10^3 samples. Plots are as follows: log posterior trace plot, autocorrelation plots for model parameters t_c, ϵ, α, p and q.

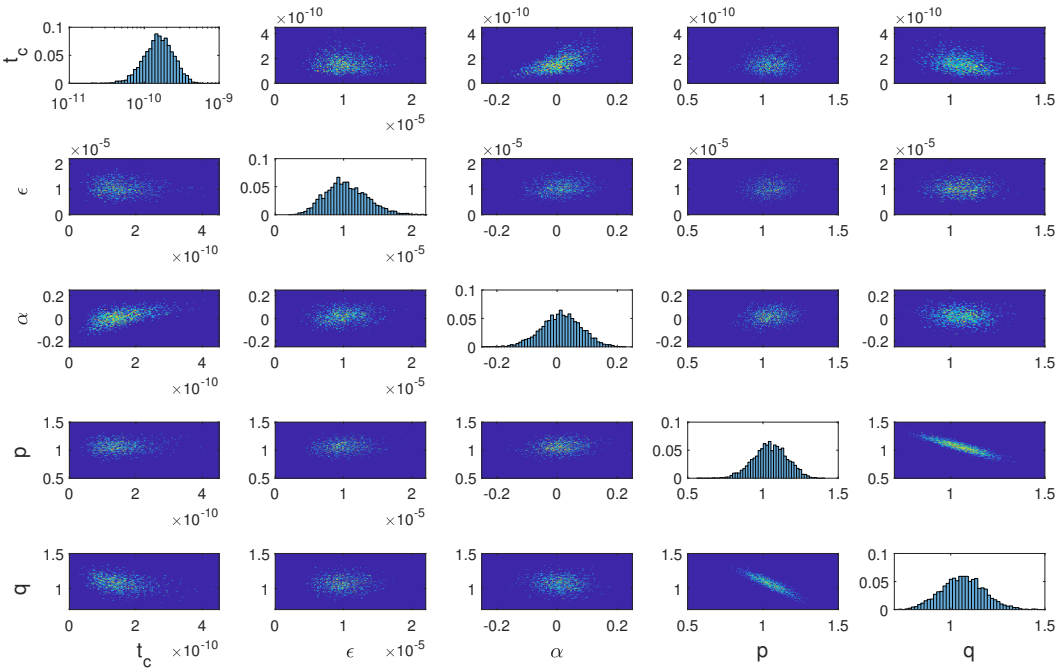

Figure A.8: Cloud plot for inferred parameter distributions when fitting poultry transmission model C to the wave 5 division dataset. Histograms along the main diagonal panels give the proportion of parameter samples (generated via the MCMC scheme) within the specific bins. Off diagonal panels give two-dimensional histograms showing the density distribution of the data when plotted against one another, with lighter colours corresponding to higher densities.

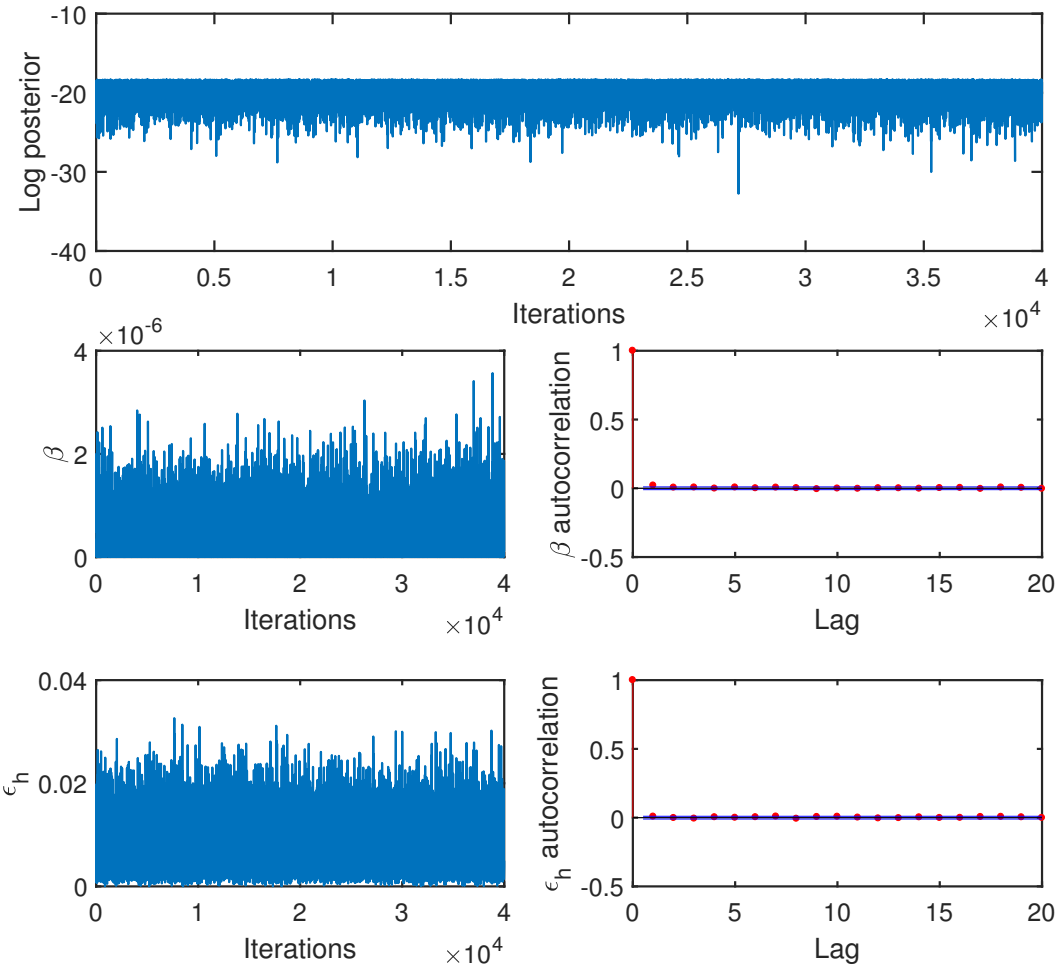

Figure A.9: MCMC scheme parameter trace and autocorrelation plots for the zoonotic transmission model fitted to the wave 6 division dataset. Run for 1.6×10^6 iterations (ignoring a burn-in period of 10^5 steps), with these replicates thinned by a factor of 40 (to lower autocorrelation in the parameter chains) giving 4×10^4 samples. **Top row:** log posterior trace plot. **Rows two-three:** Trace plots (left column) and autocorrelation plots (right column) for the model parameters β (second row) and ϵ_h (third row).

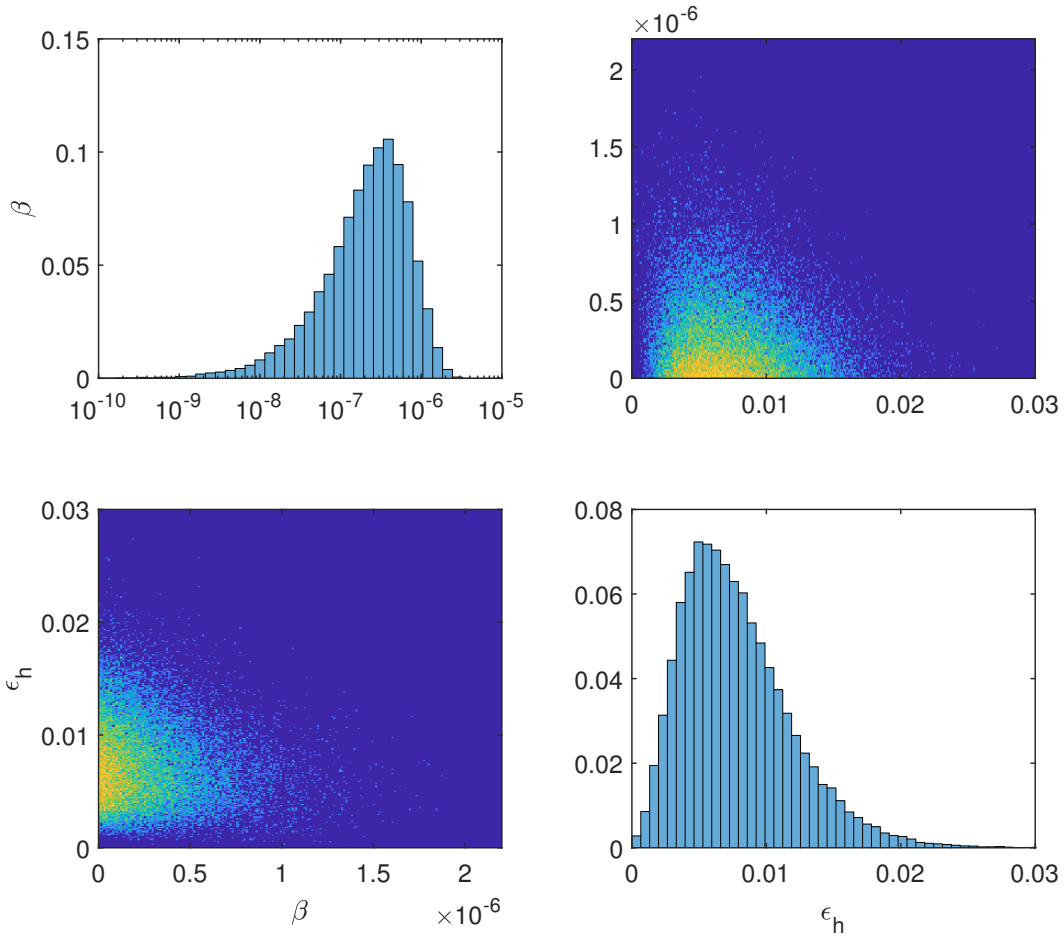

Figure A.10: Cloud plot for the zoonotic transmission model parameters inferred when fitting to the wave 6 division dataset. Histograms along the main diagonal panels give the proportion of parameter samples (generated via the MCMC scheme) within the specified bins. Off diagonal panels give two-dimensional histograms showing the density distribution of the data when plotted against one another, with lighter colours corresponding to higher densities.

Appendix B

Appendix to chapter 4

B.1 Additional tables

Table B.1: Ring culling range to optimise control objectives with respect to district of outbreak origin (seed district) and capacity setting, under wave 2 type transmission dynamics. For each combination of control capacity level, district of outbreak origin and control objective 1,000 simulation runs were performed. The two control objectives were minimising average outbreak duration and minimising the probability of an epidemic. Infected premises culling only (with no additional ring culling) being selected as the optimal strategy is denoted by —.

	Seed district		Capacity setting		
Objective	**ID**	**Name**	**Low**	**Medium**	**High**
Average duration	1	Jamalpur	2 km	3 km	7 km
	2	Sherpur	1 km	6 km	7 km
	3	Nasirabad	2 km	7 km	8 km
	4	Netrakona	1 km	7 km	4 km
	5	Tangail	2 km	2 km	4 km
	6	Gazipur	8 km	3 km	4 km
	7	Kishoreganj	1 km	3 km	3 km
	8	Narshingdi	2 km	3 km	4 km
	9	Manikgonj	1 km	1 km	5 km
	10	Dhaka	3 km	4 km	5 km
	11	Naray Angonj	1 km	4 km	5 km
	12	Munshigonj	1 km	7 km	6 km
	13	Rajbari (west)	2 km	1 km	6 km

Continued on next page

Table B.1: Continued from previous page

	Seed district		Capacity setting		
Objective	**ID**	**Name**	**Low**	**Medium**	**High**
	14	Rajbari (east)	1 km	4 km	7 km
	15	Faridpur	1 km	3 km	7 km
	16	Gopalgonj	1 km	1 km	4 km
	17	Madaripur	1 km	2 km	6 km
	18	Shariatpur	2 km	3 km	5 km
Epidemic probability	1	Jamalpur	1 km	2 km	2 km
	2	Sherpur	1 km	1 km	2 km
	3	Nasirabad	2 km	7 km	8 km
	4	Netrakona	1 km	1 km	3 km
	5	Tangail	1 km	5 km	5 km
	6	Gazipur	1 km	3 km	5 km
	7	Kishoreganj	1 km	3 km	5 km
	8	Narshingdi	2 km	3 km	5 km
	9	Manikgonj	1 km	1 km	6 km
	10	Dhaka	3 km	5 km	7 km
	11	Naray Angonj	6 km	4 km	7 km
	12	Munshigonj	2 km	1 km	7 km
	13	Rajbari (west)	1 km	2 km	2 km
	14	Rajbari (east)	1 km	2 km	10 km
	15	Faridpur	1 km	2 km	2 km
	16	Gopalgonj	1 km	1 km	3 km
	17	Madaripur	1 km	2 km	2 km
	18	Shariatpur	2 km	2 km	3 km

Table B.2: Ring vaccination range to optimise control objectives with respect to district of outbreak origin (seed district) and capacity setting, under wave 2 type transmission dynamics. For each combination of control capacity level, district of outbreak origin and control objective 1,000 simulation runs were performed. The two control objectives were minimising average outbreak duration and minimising the probability of an epidemic. Infected premises culling only (with no ring vaccination) being selected as the optimal strategy is denoted by —.

	Seed district		Capacity setting		
Objective	**ID**	**Name**	**Low**	**Medium**	**High**
Average duration	1	Jamalpur	8 km	10 km	9 km
	2	Sherpur	3 km	4 km	9 km
	3	Nasirabad	7 km	10 km	9 km
	4	Netrakona	10 km	10 km	9 km
	5	Tangail	8 km	8 km	10 km
	6	Gazipur	4 km	3 km	4 km
	7	Kishoreganj	9 km	8 km	9 km
	8	Narshingdi	9 km	10 km	6 km
	9	Manikgonj	8 km	4 km	9 km
	10	Dhaka	3 km	8 km	9 km
	11	Naray Angonj	3 km	4 km	10 km
	12	Munshigonj	10 km	5 km	8 km
	13	Rajbari (west)	9 km	10 km	10 km
	14	Rajbari (east)	3 km	6 km	10 km
	15	Faridpur	8 km	6 km	9 km
	16	Gopalgonj	8 km	9 km	9 km
	17	Madaripur	3 km	3 km	9 km
	18	Shariatpur	3 km	9 km	10 km
Epidemic probability	1	Jamalpur	2 km	8 km	8 km
	2	Sherpur	3 km	8 km	9 km
	3	Nasirabad	6 km	6 km	9 km
	4	Netrakona	2 km	4 km	9 km
	5	Tangail	9 km	10 km	5 km
	6	Gazipur	—	3 km	5 km
	7	Kishoreganj	10 km	8 km	9 km
	8	Narshingdi	8 km	8 km	10 km
	9	Manikgonj	3 km	3 km	3 km

Continued on next page

Table B.2: Continued from previous page

	Seed district		Capacity setting		
Objective	**ID**	**Name**	**Low**	**Medium**	**High**
	10	Dhaka	1 km	10 km	8 km
	11	Naray Angonj	2 km	3 km	8 km
	12	Munshigonj	2 km	9 km	5 km
	13	Rajbari (west)	6 km	7 km	7 km
	14	Rajbari (east)	7 km	7 km	10 km
	15	Faridpur	2 km	4 km	9 km
	16	Gopalgonj	—	—	8 km
	17	Madaripur	7 km	3 km	8 km
	18	Shariatpur	2 km	3 km	7 km

Table B.3: Culling and vaccination ring ranges to minimise probability of a widespread outbreak, for wave 5 transmission model with a high control capacity. For each combination of intervention method and district of outbreak origin (seed district) 1,000 simulation runs were performed. We use — to denote when solely culling infected premises was determined to be the optimal strategy.

Seed district		Control method	
ID	**Name**	**Culling**	**Vacc.**
1	Jamalpur	8 km	—
2	Sherpur	8 km	—
3	Nasirabad	8 km	—
4	Netrakona	8 km	—
5	Tangail	1 km	3 km
6	Gazipur	5 km	7 km
7	Kishoreganj	1 km	8 km
8	Narshingdi	2 km	3 km
9	Manikgonj	1 km	—
10	Dhaka	8 km	1 km
11	Naray Angonj	3 km	2 km
12	Munshigonj	8 km	5 km
13	Rajbari (west)	2 km	—
14	Rajbari (east)	10 km	—
15	Faridpur	8 km	—

Continued on next page

Table B.3: Continued from previous page

Seed district		Control method	
ID	**Name**	**Culling**	**Vacc.**
16	Gopalgonj	7 km	—
17	Madaripur	1 km	—
18	Shariatpur	2 km	—

Table B.4: Average outbreak duration for wave 2 type transmission dynamics under different active surveillance strategies, stratified by the district where the outbreak originated. For each combination of capacity setting, district of outbreak origin (seed district) and active surveillance strategy the value displayed was obtained by averaging over 1,000 simulation runs. The strategy that optimises the objective (minisming the expected outbreak duration) is highlighted in bold. All duration values are given to 1 d.p in units of days, with percentages given to 2 s.f.

	Seed district		Active surveillance strategy					
Capacity	**ID**	**Name**	**None**	**A**	**B**	**C**	**D**	**Red.**
Low	1	Jamalpur	14.9	14.8	14.7	**11.4**	14.5	24%
	2	Sherpur	14.5	14.4	14.3	**11.3**	13.5	22%
	3	Nasirabad	17.0	16.8	16.8	**12.3**	16.3	28%
	4	Netrakona	14.6	14.4	14.4	**11.5**	14.2	22%
	5	Tangail	16.8	16.4	16.4	**12.4**	15.7	26%
	6	Gazipur	19.5	19.3	19.3	**13.4**	18.2	31%
	7	Kishoreganj	17.1	16.7	16.7	**12.5**	16.4	27%
	8	Narshingdi	18.7	18.2	18.2	**13.8**	16.4	27%
	9	Manikgonj	16.7	16.3	16.2	**11.8**	15.9	29%
	10	Dhaka	17.6	17.0	16.8	**12.7**	15.7	28%
	11	Naray Angonj	18.1	17.4	17.2	**13.5**	14.1	26%
	12	Munshigonj	16.0	15.6	15.5	**12.0**	15.0	25%
	13	Rajbari (west)	14.3	14.3	14.3	**11.0**	13.9	23%
	14	Rajbari (east)	14.9	14.7	14.7	**11.4**	14.4	24%
	15	Faridpur	14.4	14.2	14.2	**10.9**	13.8	24%
	16	Gopalgonj	13.6	13.5	13.5	**10.6**	13.1	22%
	17	Madaripur	14.3	14.0	14.0	**10.5**	13.7	26%
	18	Shariatpur	14.3	14.1	14.1	**11.1**	13.7	22%
Medium	1	Jamalpur	14.9	14.7	14.7	**10.2**	13.9	32%
	2	Sherpur	14.5	14.3	14.3	**10.5**	12.1	28%

Continued on next page

Table B.4: Continued from previous page

	Seed district		Active surveillance strategy					
Capacity	**ID**	**Name**	**None**	**A**	**B**	**C**	**D**	**Red.**
	3	Nasirabad	17.0	16.7	16.7	**11.4**	15.3	33%
	4	Netrakona	14.6	14.4	14.4	**10.6**	13.7	28%
	5	Tangail	16.8	16.4	16.4	**11.4**	15.1	32%
	6	Gazipur	19.4	19.1	19.1	**12.2**	16.6	37%
	7	Kishoreganj	17.0	16.6	16.6	**11.1**	15.5	35%
	8	Narshingdi	18.7	18.2	18.2	**12.9**	14.0	31%
	9	Manikgonj	16.7	16.2	16.2	**10.5**	14.9	37%
	10	Dhaka	17.5	16.8	16.8	**11.6**	14.5	34%
	11	Naray Angonj	18.0	17.3	17.2	**12.0**	13.2	33%
	12	Munshigonj	15.9	15.6	15.5	**10.3**	14.4	35%
	13	Rajbari (west)	14.3	14.2	14.2	**10.2**	13.4	29%
	14	Rajbari (east)	14.9	14.6	14.6	**10.3**	13.8	31%
	15	Faridpur	14.4	14.1	14.1	**10.2**	13.4	29%
	16	Gopalgonj	13.6	13.4	13.4	**9.9**	12.7	27%
	17	Madaripur	14.2	14.0	14.0	**9.8**	13.3	31%
	18	Shariatpur	14.3	14.1	14.0	**10.3**	13.3	28%
High	1	Jamalpur	14.9	14.7	14.7	**8.4**	11.9	44%
	2	Sherpur	14.5	14.3	14.3	**9.2**	10.4	36%
	3	Nasirabad	16.9	16.6	16.6	**9.9**	12.9	42%
	4	Netrakona	14.6	14.3	14.3	**9.1**	12.4	37%
	5	Tangail	16.8	16.4	16.4	**9.5**	12.9	44%
	6	Gazipur	19.4	19.1	19.1	**9.9**	13.0	49%
	7	Kishoreganj	17.0	16.6	16.6	**8.8**	13.4	48%
	8	Narshingdi	18.7	18.1	18.1	10.5	**10.0**	46%
	9	Manikgonj	16.7	16.2	16.2	**8.2**	12.5	51%
	10	Dhaka	17.5	16.8	16.8	**9.2**	11.8	47%
	11	Naray Angonj	18.0	17.2	17.2	**9.8**	10.7	46%
	12	Munshigonj	15.9	15.5	15.5	**8.1**	13.1	49%
	13	Rajbari (west)	14.3	14.1	14.1	**9.1**	11.9	37%
	14	Rajbari (east)	14.8	14.6	14.6	**9.1**	12.6	39%
	15	Faridpur	14.4	14.1	14.1	**9.0**	12.1	37%
	16	Gopalgonj	13.6	13.4	13.4	**8.8**	11.6	35%
	17	Madaripur	14.2	13.9	13.9	**8.5**	12.0	40%

Continued on next page

Table B.4: Continued from previous page

	Seed district		Active surveillance strategy					
Capacity	**ID**	**Name**	**None**	**A**	**B**	**C**	**D**	**Red.**
	18	Shariatpur	14.2	14.0	14.0	**9.2**	11.3	35%

Active surveillance strategy legend:

A - reactive by distance,

B - reactive by population,

C - proactive by population,

D - proactive by density,

Red. - Percentage decrease in average outbreak duration using preferred active surveillance scheme compared to when no active surveillance is imposed.

Table B.5: Epidemic probabilites for wave 2 type transmission dynamics under different active surveillance strategies, stratified by the district where the outbreak originated. For each combination of capacity setting, district of outbreak origin (seed district) and active surveillance strategy the value displayed was obtained by averaging over 1,000 simulation runs. The strategy that optimises the objective (minisming the probability of an epidemic) is highlighted in bold. All probabilities are given to 3 d.p., with percentages given to 2 s.f.

	Seed district		Active surveillance strategy					
Capacity	**ID**	**Name**	**None**	**A**	**B**	**C**	**D**	**Red.**
Low	1	Jamalpur	0.065	0.062	0.060	**0.012**	0.056	82%
	2	Sherpur	0.055	0.050	0.050	**0.010**	0.046	82%
	3	Nasirabad	0.103	0.098	0.098	**0.023**	0.092	78%
	4	Netrakona	0.065	0.059	0.059	**0.012**	0.060	82%
	5	Tangail	0.099	0.091	0.091	**0.024**	0.081	76%
	6	Gazipur	0.155	0.152	0.151	**0.039**	0.130	75%
	7	Kishoreganj	0.110	0.105	0.104	**0.022**	0.096	80%
	8	Narshingdi	0.138	0.130	0.130	**0.040**	0.108	71%
	9	Manikgonj	0.096	0.086	0.085	**0.018**	0.078	81%
	10	Dhaka	0.111	0.104	0.102	**0.030**	0.082	73%
	11	Naray Angonj	0.122	0.115	0.112	**0.041**	0.073	66%
	12	Munshigonj	0.077	0.072	0.070	**0.018**	0.066	77%
	13	Rajbari (west)	0.058	0.056	0.054	**0.011**	0.051	81%
	14	Rajbari (east)	0.063	0.059	0.057	**0.013**	0.056	79%
	15	Faridpur	0.058	0.054	0.053	**0.009**	0.051	84%

Continued on next page

Table B.5: Continued from previous page

	Seed district		Active surveillance strategy					
Capacity	**ID**	**Name**	**None**	**A**	**B**	**C**	**D**	**Red.**
	16	Gopalgonj	0.047	0.044	0.045	**0.008**	0.041	83%
	17	Madaripur	0.056	0.050	0.050	**0.005**	0.048	91%
	18	Shariatpur	0.056	0.052	0.052	**0.009**	0.051	84%
Medium	1	Jamalpur	0.065	0.059	0.060	**0.007**	0.047	89%
	2	Sherpur	0.055	0.050	0.050	**0.006**	0.035	89%
	3	Nasirabad	0.102	0.096	0.096	**0.011**	0.075	89%
	4	Netrakona	0.065	0.058	0.059	**0.006**	0.049	91%
	5	Tangail	0.099	0.091	0.091	**0.014**	0.067	86%
	6	Gazipur	0.155	0.150	0.150	**0.024**	0.100	85%
	7	Kishoreganj	0.110	0.103	0.103	**0.013**	0.079	88%
	8	Narshingdi	0.138	0.130	0.130	**0.025**	0.080	82%
	9	Manikgonj	0.096	0.085	0.085	**0.011**	0.066	89%
	10	Dhaka	0.110	0.101	0.101	**0.019**	0.066	83%
	11	Naray Angonj	0.122	0.114	0.112	**0.021**	0.058	83%
	12	Munshigonj	0.076	0.069	0.069	**0.008**	0.054	89%
	13	Rajbari (west)	0.058	0.054	0.054	**0.006**	0.043	90%
	14	Rajbari (east)	0.063	0.057	0.057	**0.005**	0.047	92%
	15	Faridpur	0.058	0.052	0.053	**0.005**	0.044	91%
	16	Gopalgonj	0.047	0.043	0.044	**0.003**	0.036	94%
	17	Madaripur	0.056	0.049	0.050	**0.003**	0.040	95%
	18	Shariatpur	0.056	0.052	0.052	**0.004**	0.042	93%
High	1	Jamalpur	0.065	0.059	0.059	**0.001**	0.024	98%
	2	Sherpur	0.055	0.050	0.050	**0.002**	0.018	96%
	3	Nasirabad	0.102	0.094	0.094	**0.002**	0.044	98%
	4	Netrakona	0.065	0.058	0.058	**0.001**	0.028	98%
	5	Tangail	0.099	0.091	0.091	**0.002**	0.036	98%
	6	Gazipur	0.154	0.149	0.149	**0.005**	0.047	97%
	7	Kishoreganj	0.109	0.103	0.103	**0.004**	0.047	96%
	8	Narshingdi	0.137	0.129	0.129	**0.008**	0.035	94%
	9	Manikgonj	0.096	0.085	0.085	**0.001**	0.038	99%
	10	Dhaka	0.110	0.101	0.101	**0.002**	0.037	98%
	11	Naray Angonj	0.121	0.112	0.112	**0.003**	0.026	98%
	12	Munshigonj	0.076	0.069	0.069	**0.001**	0.034	99%

Continued on next page

Table B.5: Continued from previous page

	Seed district		Active surveillance strategy					
Capacity	**ID**	**Name**	**None**	**A**	**B**	**C**	**D**	**Red.**
	13	Rajbari (west)	0.058	0.053	0.053	**0.001**	0.031	98%
	14	Rajbari (east)	0.063	0.057	0.057	**0.001**	0.032	98%
	15	Faridpur	0.058	0.052	0.052	**0.001**	0.029	98%
	16	Gopalgonj	0.047	0.043	0.043	**0.001**	0.024	98%
	17	Madaripur	0.056	0.049	0.050	**0.001**	0.026	98%
	18	Shariatpur	0.056	0.052	0.052	**0.001**	0.025	98%

Active surveillance strategy legend:

A - reactive by distance,

B - reactive by population,

C - proactive by population,

D - proactive by density,

Red. - Percentage decrease in epidemic probability using preferred active surveillance scheme compared to when no active surveillance is imposed.

Table B.6: Average outbreak duration for wave 5 type transmission dynamics under different active surveillance strategies, stratified by the district where the outbreak originated. For each combination of capacity setting, district of outbreak origin (seed district) and active surveillance strategy the value displayed was obtained by averaging over a minimum of 300 simulation runs. The strategy that optimises the objective (minisming the expected outbreak duration) is highlighted in bold. All duration values are given to 1 d.p. in units of days, with percentages given to 2 s.f.

	Seed district		Active surveillance strategy					
Capacity	**ID**	**Name**	**None**	**A**	**B**	**C**	**D**	**Red.**
Low	1	Jamalpur	98.6	95.5	95.4	**61.6**	96.5	37%
	2	Sherpur	51.8	51.6	51.6	**36.2**	49.4	30%
	3	Nasirabad	159.1	156.1	156.1	**99.5**	157.0	37%
	4	Netrakona	59.6	59.5	59.5	**40.1**	58.6	33%
	5	Tangail	138.4	132.5	132.5	**87.2**	131.4	45%
	6	Gazipur	304.1	282.5	282.4	**168.6**	282.2	45%
	7	Kishoreganj	158.9	156.8	156.7	**87.6**	152.8	45%
	8	Narshingdi	135.7	129.7	129.7	**76.9**	121.4	43%
	9	Manikgonj	185.2	177.2	177.2	**77.3**	177.3	58%

Continued on next page

Table B.6: Continued from previous page

	Seed district		Active surveillance strategy					
Capacity	**ID**	**Name**	**None**	**A**	**B**	**C**	**D**	**Red.**
	10	Dhaka	177.2	165.0	165.1	**77.3**	149.5	56%
	11	Naray Angonj	166.0	138.3	136.4	**83.3**	106.0	50%
	12	Munshigonj	87.1	84.1	84.1	**50.4**	84.1	42%
	13	Rajbari (west)	53.9	52.9	52.9	**28.5**	53.9	47%
	14	Rajbari (east)	51.0	50.1	50.0	**32.4**	48.0	37%
	15	Faridpur	62.5	61.6	61.6	**30.4**	61.5	51%
	16	Gopalgonj	23.6	23.5	23.5	**15.8**	23.6	33%
	17	Madaripur	63.6	61.5	61.5	**28.5**	62.5	55%
	18	Shariatpur	26.6	25.6	25.6	**17.8**	26.6	33%
Medium	1	Jamalpur	43.4	41.9	41.9	**20.7**	41.9	52%
	2	Sherpur	24.6	24.3	24.3	**16.0**	22.4	35%
	3	Nasirabad	66.8	64.9	65.9	**38.0**	64.7	43%
	4	Netrakona	28.0	28.0	27.8	**15.6**	27.3	44%
	5	Tangail	58.9	56.3	56.3	**31.2**	54.5	47%
	6	Gazipur	118.9	118.7	117.7	**57.7**	113.5	52%
	7	Kishoreganj	66.6	67.6	67.1	**29.5**	61.0	56%
	8	Narshingdi	58.1	55.4	55.4	**28.8**	42.3	50%
	9	Manikgonj	78.0	74.0	73.7	**25.6**	69.4	67%
	10	Dhaka	74.6	70.9	71.3	**29.5**	58.3	60%
	11	Naray Angonj	70.3	57.6	57.5	**31.3**	43.0	55%
	12	Munshigonj	39.0	37.5	37.5	**18.3**	35.9	53%
	13	Rajbari (west)	25.8	25.4	25.3	**12.4**	25.6	52%
	14	Rajbari (east)	24.7	24.2	24.2	**13.4**	23.4	46%
	15	Faridpur	29.4	28.8	28.8	**15.1**	28.7	49%
	16	Gopalgonj	13.6	13.5	13.5	**9.7**	13.5	28%
	17	Madaripur	29.8	28.8	28.8	**13.5**	29.2	55%
	18	Shariatpur	14.9	14.4	14.4	**9.9**	14.7	34%
High	1	Jamalpur	21.8	20.8	20.8	**9.0**	16.4	59%
	2	Sherpur	14.1	13.8	13.8	**8.8**	11.2	38%
	3	Nasirabad	29.8	28.4	28.4	**12.1**	23.6	59%
	4	Netrakona	15.4	15.1	15.1	**8.5**	14.0	45%
	5	Tangail	27.3	25.8	25.8	**10.9**	20.6	60%
	6	Gazipur	50.5	47.1	46.7	**15.8**	34.2	69%

Continued on next page

Table B.6: Continued from previous page

	Seed district		Active surveillance strategy					
Capacity	**ID**	**Name**	**None**	**A**	**B**	**C**	**D**	**Red.**
	7	Kishoreganj	30.8	29.5	29.5	**10.0**	24.7	68%
	8	Narshingdi	27.7	26.1	26.1	**10.8**	14.4	61%
	9	Manikgonj	35.3	33.0	33.1	**8.8**	27.5	75%
	10	Dhaka	34.2	31.3	31.3	**10.4**	21.9	70%
	11	Naray Angonj	33.2	27.1	27.1	**10.3**	16.2	69%
	12	Munshigonj	20.1	19.2	19.2	**7.9**	17.1	61%
	13	Rajbari (west)	14.7	14.3	14.3	**8.1**	12.7	45%
	14	Rajbari (east)	14.3	13.9	13.9	**8.1**	12.9	43%
	15	Faridpur	16.2	15.7	15.7	**8.3**	14.6	49%
	16	Gopalgonj	9.5	9.4	9.4	**7.4**	9.1	22%
	17	Madaripur	16.3	15.7	15.7	**8.1**	14.6	50%
	18	Shariatpur	10.2	9.9	9.9	**7.6**	8.9	25%

Active surveillance strategy legend:

A - reactive by distance,

B - reactive by population,

C - proactive by population,

D - proactive by density,

Red. - Percentage decrease in average outbreak duration using preferred active surveillance scheme compared to when no active surveillance is imposed.

Table B.7: Epidemic probabilites for wave 5 type transmission dynamics under different active surveillance strategies, stratified by the district where the outbreak originated. For each combination of capacity setting, district of outbreak origin (seed district) and active surveillance strategy the value displayed was obtained by averaging over a minimum of 300 simulation runs. The strategy that optimises the objective (minisming the probability of an epidemic) is highlighted in bold. All probabilities are given to 3 d.p., with percentages given to 2 s.f.

	Seed district		Active surveillance strategy					
Capacity	**ID**	**Name**	**None**	**A**	**B**	**C**	**D**	**Red.**
Low	1	Jamalpur	0.095	0.093	0.093	**0.058**	0.093	39%
	2	Sherpur	0.046	0.046	0.046	**0.030**	0.044	35%
	3	Nasirabad	0.161	0.157	0.157	**0.100**	0.159	38%

Continued on next page

Table B.7: Continued from previous page

	Seed district		Active surveillance strategy					
Capacity	**ID**	**Name**	**None**	**A**	**B**	**C**	**D**	**Red.**
	4	Netrakona	0.055	0.055	0.055	**0.035**	0.054	36%
	5	Tangail	0.138	0.133	0.133	**0.088**	0.131	36%
	6	Gazipur	0.317	0.304	0.295	**0.176**	0.291	44%
	7	Kishoreganj	0.162	0.159	0.159	**0.090**	0.156	44%
	8	Narshingdi	0.135	0.129	0.129	**0.076**	0.121	44%
	9	Manikgonj	0.192	0.181	0.182	**0.080**	0.183	58%
	10	Dhaka	0.179	0.168	0.168	**0.075**	0.151	58%
	11	Naray Angonj	0.172	0.140	0.138	**0.091**	0.107	47%
	12	Munshigonj	0.083	0.081	0.081	**0.047**	0.080	43%
	13	Rajbari (west)	0.049	0.048	0.048	**0.022**	0.049	55%
	14	Rajbari (east)	0.045	0.044	0.044	**0.026**	0.042	42%
	15	Faridpur	0.059	0.058	0.058	**0.025**	0.058	58%
	16	Gopalgonj	0.017	0.017	0.017	**0.009**	0.017	47%
	17	Madaripur	0.059	0.057	0.057	**0.023**	0.058	61%
	18	Shariatpur	0.020	0.019	0.019	**0.011**	0.020	45%
Medium	1	Jamalpur	0.094	0.093	0.093	**0.040**	0.092	58%
	2	Sherpur	0.046	0.046	0.046	**0.026**	0.043	43%
	3	Nasirabad	0.161	0.157	0.159	**0.095**	0.158	41%
	4	Netrakona	0.055	0.055	0.055	**0.026**	0.054	53%
	5	Tangail	0.138	0.133	0.133	**0.073**	0.129	47%
	6	Gazipur	0.305	0.311	0.309	**0.156**	0.291	49%
	7	Kishoreganj	0.162	0.165	0.163	**0.068**	0.150	58%
	8	Narshingdi	0.135	0.129	0.129	**0.061**	0.098	55%
	9	Manikgonj	0.192	0.182	0.181	**0.054**	0.171	72%
	10	Dhaka	0.179	0.173	0.174	**0.065**	0.142	64%
	11	Naray Angonj	0.172	0.137	0.137	**0.073**	0.102	58%
	12	Munshigonj	0.083	0.081	0.081	**0.033**	0.077	60%
	13	Rajbari (west)	0.049	0.048	0.048	**0.015**	0.049	69%
	14	Rajbari (east)	0.045	0.044	0.044	**0.019**	0.042	58%
	15	Faridpur	0.059	0.058	0.058	**0.023**	0.058	61%
	16	Gopalgonj	0.017	0.017	0.017	**0.008**	0.017	53%
	17	Madaripur	0.059	0.057	0.057	**0.020**	0.058	66%
	18	Shariatpur	0.020	0.019	0.019	**0.080**	0.020	60%

Continued on next page

Table B.7: Continued from previous page

	Seed district		Active surveillance strategy					
Capacity	**ID**	**Name**	**None**	**A**	**B**	**C**	**D**	**Red.**
High	1	Jamalpur	0.095	0.093	0.093	**0.029**	0.072	69%
	2	Sherpur	0.046	0.046	0.046	**0.022**	0.039	52%
	3	Nasirabad	0.161	0.157	0.157	**0.069**	0.135	57%
	4	Netrakona	0.055	0.055	0.055	**0.022**	0.052	60%
	5	Tangail	0.138	0.133	0.133	**0.052**	0.107	62%
	6	Gazipur	0.304	0.298	0.292	**0.119**	0.220	61%
	7	Kishoreganj	0.162	0.159	0.159	**0.044**	0.139	73%
	8	Narshingdi	0.135	0.129	0.129	**0.042**	0.070	69%
	9	Manikgonj	0.192	0.180	0.182	**0.031**	0.158	84%
	10	Dhaka	0.179	0.168	0.168	**0.045**	0.120	75%
	11	Naray Angonj	0.172	0.137	0.137	**0.039**	0.078	77%
	12	Munshigonj	0.083	0.081	0.081	**0.024**	0.072	71%
	13	Rajbari (west)	0.049	0.048	0.048	**0.014**	0.042	71%
	14	Rajbari (east)	0.045	0.044	0.044	**0.016**	0.040	64%
	15	Faridpur	0.059	0.058	0.058	**0.017**	0.054	71%
	16	Gopalgonj	0.017	0.017	0.017	**0.007**	0.016	59%
	17	Madaripur	0.059	0.057	0.057	**0.014**	0.053	76%
	18	Shariatpur	0.020	0.019	0.019	**0.007**	0.016	65%

Active surveillance strategy legend:

A - reactive by distance,

B - reactive by population,

C - proactive by population,

D - proactive by density,

Red. - Percentage decrease in epidemic probability using preferred active surveillance scheme compared to when no active surveillance is imposed.

Table B.8: Predicted probability of outbreak duration t being 90 days or less for different active surveillance strategies. For each combination of transmission model, capacity restriction and active surveillance strategy the value displayed was obtained by averaging over a minimum of 300 simulation runs. For each model and control capacity combination the strategy that optimises the objective, having an outbreak duration below 90 days, is highlighted in bold. All values are given to 2 s.f.

		Active surveillance strategy				
Transmission model	**Control capacity**	**None**	**A**	**B**	**C**	**D**
Wave 2	Low	0.51	0.50	0.50	**0.55**	0.53
	Medium	0.51	0.50	0.50	**0.57**	0.55
	High	0.51	0.50	0.50	**0.61**	0.57
Wave 5	Low	0.38	0.40	0.40	**0.46**	0.38
	Medium	0.38	0.40	0.39	**0.48**	0.41
	High	0.38	0.39	0.39	**0.58**	0.47

Active surveillance strategy legend:

A - reactive by distance,

B - reactive by population,

C - proactive by population,

D - proactive by density.

Table B.9: Predicted probability of outbreak size I not exceeding 25 premises for different active surveillance strategies. For each combination of transmission model, capacity restriction and active surveillance strategy the value displayed was obtained by averaging over a minimum of 300 simulation runs. For each model and control capacity combination the strategy that optimises the objective, maximising the likelihood of having an outbreak with 25 premises or less infected, is highlighted in bold. All values are given to 2 s.f.

		Active surveillance strategy				
Transmission model	**Control capacity**	**None**	**A**	**B**	**C**	**D**
Wave 2	Low	0.69	0.70	0.70	**0.79**	0.72
	Medium	0.69	0.70	0.70	**0.82**	0.74
	High	0.69	0.70	0.70	**0.87**	0.81
Wave 5	Low	0.44	0.47	0.47	**0.52**	0.45
	Medium	0.44	0.46	0.46	**0.55**	0.48
	High	0.44	0.45	0.45	**0.64**	0.53

Active surveillance strategy legend:

A - reactive by distance,

B - reactive by population,

C - proactive by population,

D - proactive by density.

Table B.10: Mean number of poultry culled under different active surveillance strategies. For each combination of transmission model, capacity restriction and active surveillance strategy the value displayed was obtained by averaging over a minimum of 300 simulation runs. For each model and control capacity combination the strategy that optimises the objective, minimising the mean number of poultry culled, is highlighted in bold. All values are given to 2 s.f.

		Active surveillance strategy				
	Capacity	**None**	**A**	**B**	**C**	**D**
Wave 2	Low	3.0×10^5	2.8×10^5	2.7×10^5	$\mathbf{1.8 \times 10^5}$	2.6×10^5
	Medium	3.0×10^5	2.8×10^5	2.7×10^5	$\mathbf{1.5 \times 10^5}$	2.3×10^5
	High	3.0×10^5	2.7×10^5	2.7×10^5	$\mathbf{1.1 \times 10^5}$	1.7×10^5
Wave 5	Low	1.1×10^7	1.1×10^7	1.1×10^7	$\mathbf{0.8 \times 10^7}$	1.1×10^7
	Medium	1.0×10^7	1.0×10^7	1.0×10^7	$\mathbf{0.6 \times 10^7}$	0.9×10^7
	High	6.6×10^6	6.2×10^6	6.2×10^6	$\mathbf{1.7 \times 10^6}$	4.8×10^6

Active surveillance strategy legend:

A - reactive by distance,

B - reactive by population,

C - proactive by population,

D - proactive by density.

Table B.11: Normalised proportion of runs with at least one human case, relative to the scenario with unaltered ϵ_h values. For each combination of resource restrictions, transmission model and ϵ_h scaling factor the value displayed was obtained by averaging over 1,000 simulation runs. All values are given to 2 s.f.

		% reduction in ϵ_h		
Control capacity	**Transmission model**	**50**	**75**	**100**
Low	Wave 2	0.81	0.68	0.40
	Wave 5	0.96	0.87	0.75
Medium	Wave 2	0.80	0.67	0.40
	Wave 5	0.94	0.84	0.80
High	Wave 2	0.81	0.67	0.39
	Wave 5	0.95	0.89	0.77

Table B.12: Normalised mean total of days with human cases, relative to the scenario with unaltered ϵ_h values. For each combination of resource restrictions, transmission model and ϵ_h scaling factor the value displayed was obtained by averaging over 1,000 simulation runs. All values are given to 2 s.f.

		% reduction in ϵ_h		
Control capacity	**Transmission model**	**50**	**75**	**100**
Low	Wave 2	0.63	0.47	0.27
	Wave 5	1.01	0.96	0.95
Medium	Wave 2	0.63	0.46	0.27
	Wave 5	0.96	0.90	0.98
High	Wave 2	0.63	0.45	0.26
	Wave 5	0.92	0.99	0.90

B.2 Additional figures

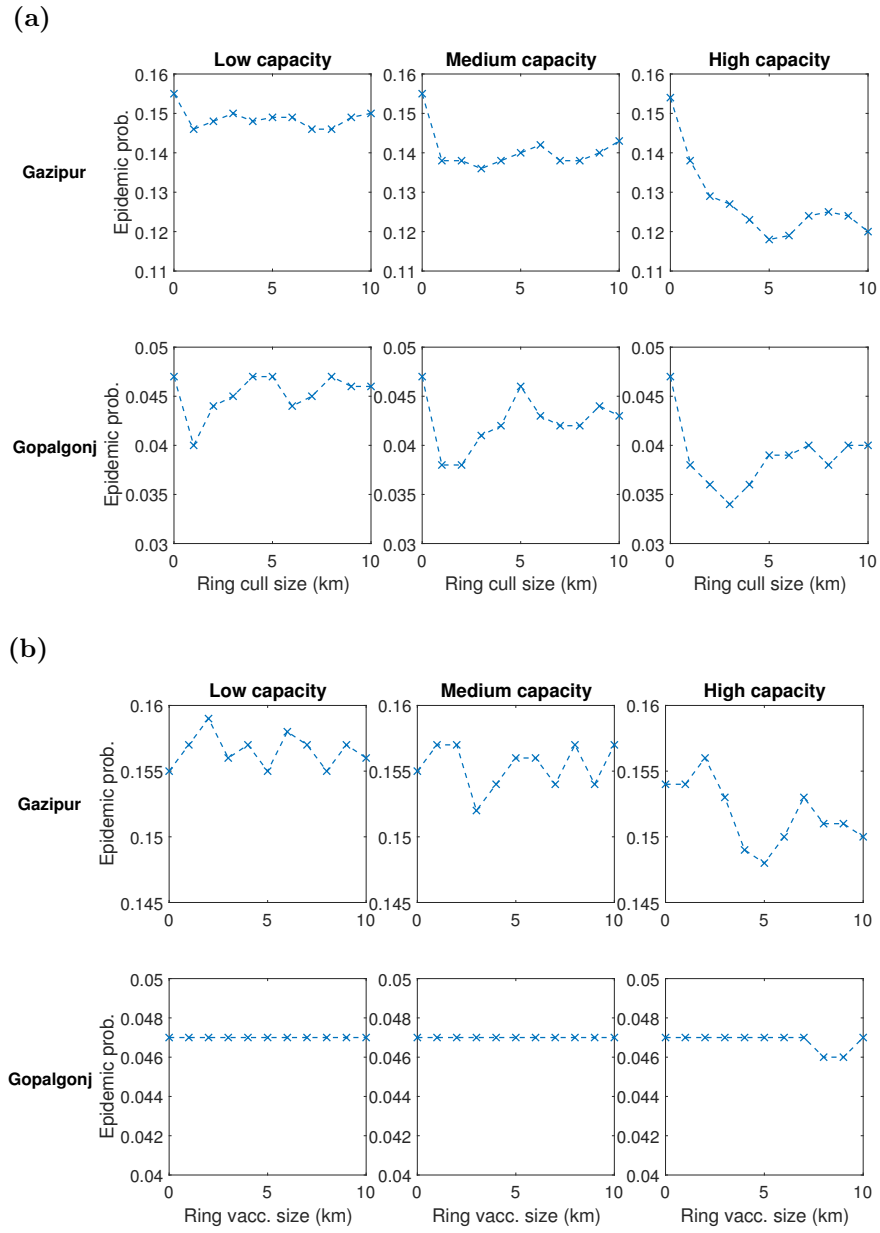

Figure B.1: Sensitivity of epidemic probability to intervention ring size and capacity restrictions, under wave 2 transmission dynamics. For outbreaks originating in the Gazipur and Gopalgonj districts (for a district locator map see figure 4.1), the panels show predicted epidemic probability with the following intervention measure utilised: **(a)** ring culling; **(b)** ring vaccination. Larger variations in this control metric are observed across the range of ring culling sizes tested compared to ring vaccination. Analogous outcomes were found for outbreaks seeded in the remaining districts.

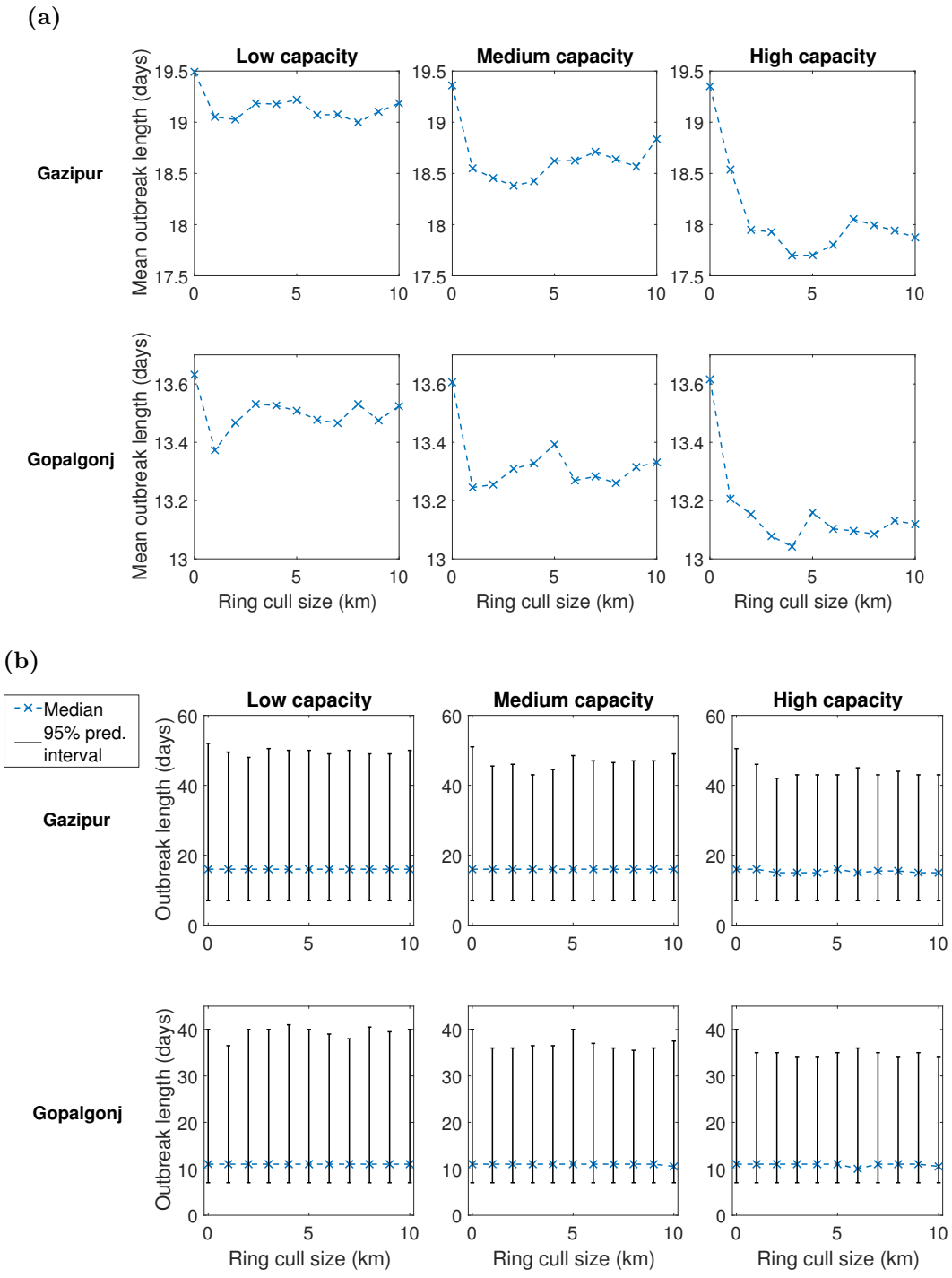

Figure B.2: Sensitivity of outbreak duration to culling ring size and capacity restrictions, under wave 2 transmission dynamics. For outbreaks originating in the Gazipur and Gopalgonj districts (for a district locator map see figure 4.1), the panels are as follows: **(a)** Average outbreak duration against culling ring size under each capacity level, with larger variations evident under less restrictive capacity levels; **(b)** 95% prediction intervals (black bars) for outbreak duration against culling ring size under each capacity level, with median values depicted by the blue markers and dashed line. Similar ranges are obtained, independent of capacity level and culling severity. Analogous outcomes were found for outbreaks seeded in the remaining districts. For each combination of intervention method and district of outbreak origin 1,000 simulation runs were performed.

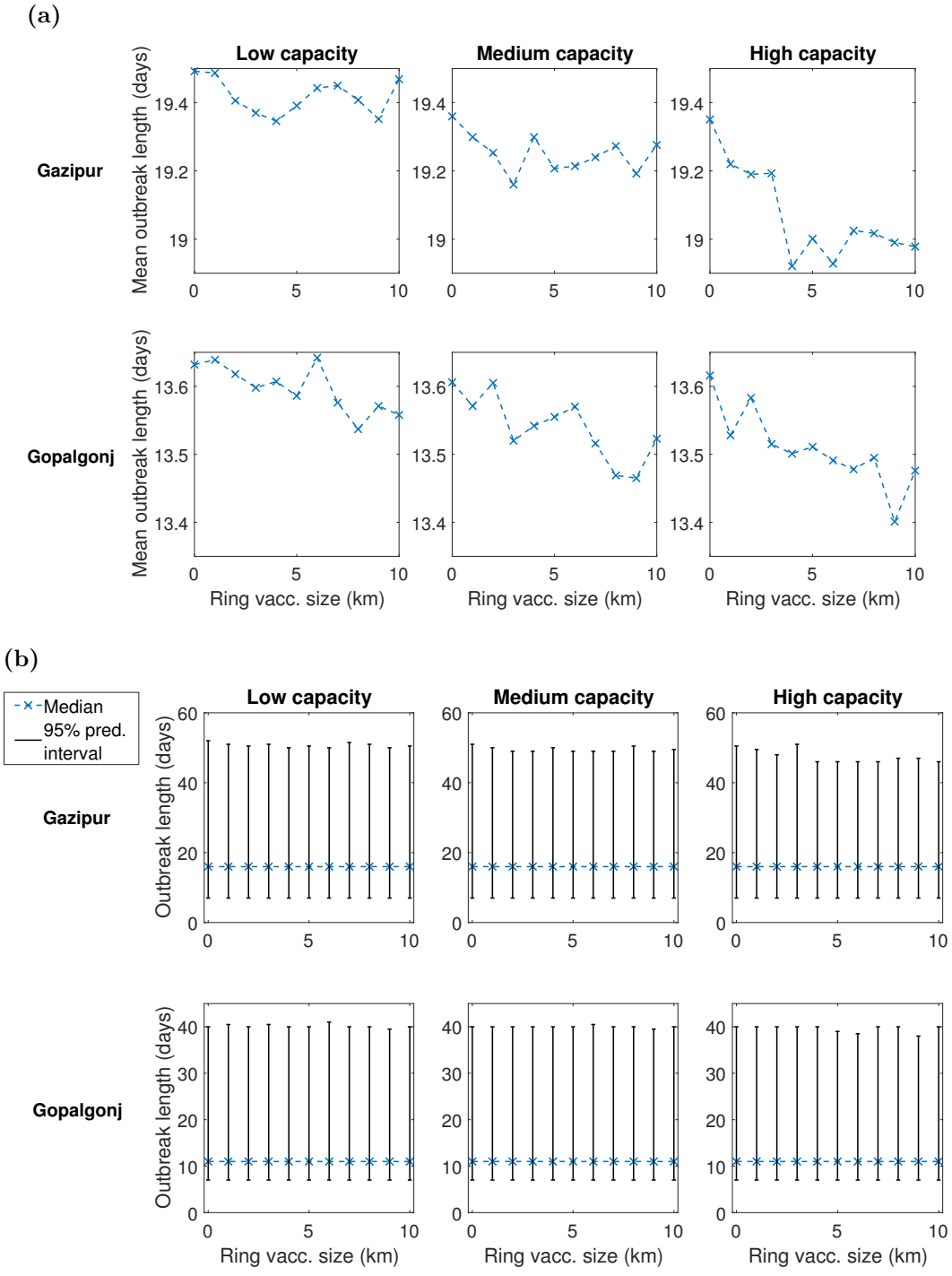

Figure B.3: Sensitivity of outbreak duration to vaccination ring size and capacity restrictions, under wave 2 transmission dynamics. For outbreaks originating in the Gazipur and Gopalgonj districts (for a district locator map see figure 4.1), the panels are as follows: **(a)** Average outbreak duration against vaccination ring size under each capacity level, with larger variations evident under less restrictive capacity levels; **(b)** 95% prediction intervals (black bars) for outbreak duration against vaccination ring size under each capacity level, with median values depicted by the blue markers and dashed line. Similar ranges are obtained, independent of capacity level and vaccination severity. Analogous outcomes were found for outbreaks seeded in the remaining districts. For each combination of intervention method and district of outbreak origin 1,000 simulation runs were performed.

Appendix C

Appendix to chapter 6

C.1 Analysis of confounding

C.1.1 Setup

Our aim here is to state in mathematical language what is meant by transmission of mood, how confounding is possible and not possible. We will do this using pairwise model notation, and will write $[A]$ for the number of nodes of state $[A]$, $[A \to B]$ for the number of individuals in state A naming an individual in state B, at a given time point that we will normally omit; formally

$$[A] = \sum_i \mathrm{I}\left\{X_i^t = A\right\}, \quad [A \to B] = \sum_{i,j} \mathrm{I}\left\{X_i^t = A\right\} \mathrm{I}\left\{G_{ij} = 1\right\} \mathrm{I}\left\{X_j^t = B\right\} \quad \text{(C.1)}$$

We are going to consider how to calculate relevant quantities for both a transmission model and a model with homophily relating to some unobserved property ξ.

C.1.2 Homophily model

Suppose we have a property (or vector of properties) that individuals have, for example age, socio-economic status, or spatial location. We label these properties with ξ and write $[\xi]$ for the number of nodes that are of ξ etc.

Now consider a relatively general model in which the probability of changing state if in state A and with property ξ is ρ_ξ^A. We can then write down equilibrium values

for the expected number of pairs under the stationary distribution π, which are

$$\mathbb{E}_\pi[A \to B] = \sum_{\xi,\xi'} \mathbb{E}_\pi[\xi \to \xi'] \frac{\rho_\xi^{\overline{A}}}{\rho_\xi^{\overline{A}} + \rho_\xi^A} \frac{\rho_{\xi'}^{\overline{B}}}{\rho_{\xi'}^{\overline{B}} + \rho_{\xi'}^B}. \tag{C.2}$$

It is clear that by tuning the propensity of different property types to name each other as friends, and the transition probabilities, arbitrary pair structures can be created. However, for the transitions we have at equilibrium

$$\mathbb{E}_\pi[X_k^{A\to A}] = \sum_\xi \mathbb{E}_\pi[A_\xi](1-\rho_\xi^A), \quad \mathbb{E}_\pi[X_k^{A\to\overline{A}}] = \sum_\xi \mathbb{E}_\pi[A_\xi]\rho_\xi^A. \tag{C.3}$$

These do not depend on k. Overall, therefore, this model cannot be falsified from observations of numbers of pairs $[A \to B]$, but can be falsified from observations of transitions stratified by k, $Y_k^{A\to B}$.

C.1.3 Transmission model

Our model for transmitting state S (taking the value D or N), with k being the number of friends in the transmitting state, took the form

$$\begin{aligned} X_k^{N\to D} &\sim \mathrm{Bin}\left(p_k, \sum_i \mathrm{I}\left\{X_i^t = N\right\} \mathrm{I}\left\{\left(\textstyle\sum_j G_{ij} \mathrm{I}\left\{X_j^t = S\right\}\right) = k\right\}\right), \\ X_k^{D\to N} &\sim \mathrm{Bin}\left(q_k, \sum_i \mathrm{I}\left\{X_i^t = D\right\} \mathrm{I}\left\{\left(\textstyle\sum_j G_{ij} \mathrm{I}\left\{X_j^t = S\right\}\right) = k\right\}\right). \end{aligned} \tag{C.4}$$

This means that given the freedom to choose p_k and q_k for a given network configuration, it is possible to tune the expected values of these transitions to whatever value is required. The probabilities assigned to different network configurations under the invariant distribution π do not in general have an analytic closed form solution. In the event where p_k and q_k do not depend on k, then equations of the form (C.3) will hold where every individual has the same property ξ.

In the event where the population has size n and there are on average m friends per individual, note that basic combinatorial considerations give that

$$m[N] = [N \to N] + [N \to D], \quad m[D] = [D \to N] + [D \to D], \quad n = [N] + [D], \tag{C.5}$$

meaning that there are only three independent parameters: $[N]$; $[N \to D]$; and $[D \to N]$. Now suppose that p_k is monotone decreasing with k and if q_k is monotone

increasing with k, this will lead to fewer $[D \to N]$ pairs than equations of the form (C.3) would suggest due to transmission of N.

C.1.4 Other models

It is possible to combine elements of the transmission and homophily models in various ways. We take the philosophical position that anything more complex than the homophily model above will constitute a *mechanism* for the phenomenon of social contagion rather an alternative to it.

C.2 Additional figures

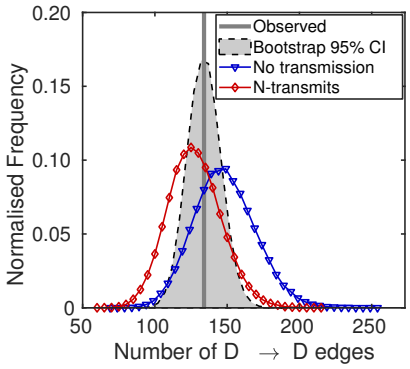

Figure C.1: Number of D $\rightarrow$ D edges for the stationary distributions of the models versus real data. Asterisks above a plot denote a significant statistical difference at the 5% level, corresponding to $p < 0.01$ using the Bonferroni method to account for multiple testing. Observed data could be plausibly generated by both N-transmits ($p = 0.59$) and no-transmission ($p = 0.60$) models.

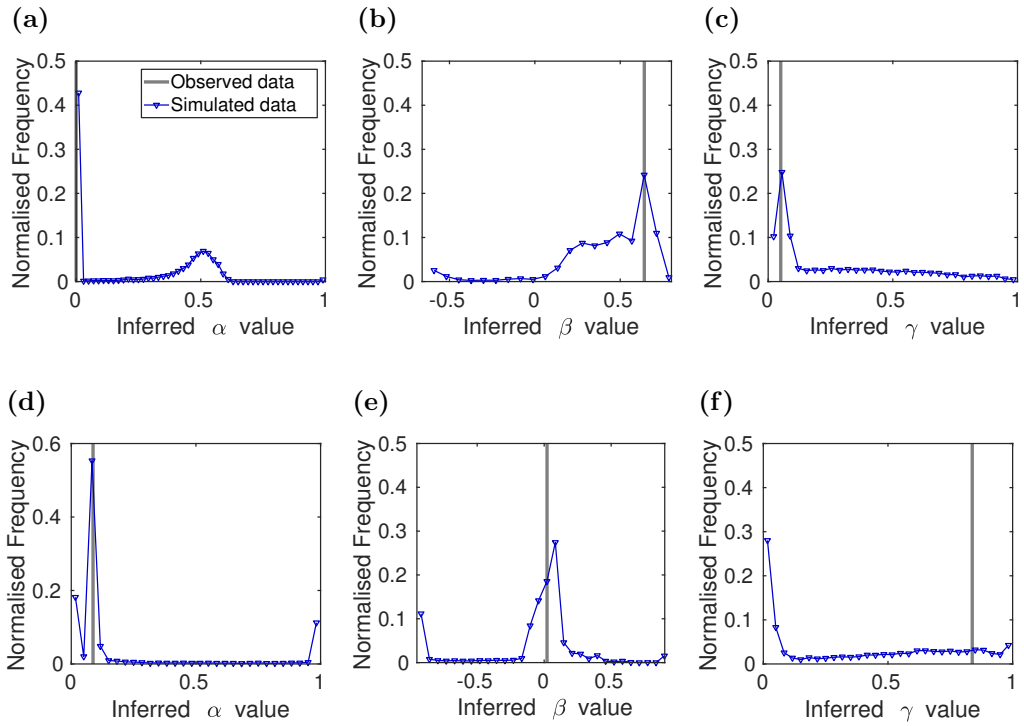

Figure C.2: Normalised frequency of inferred N-transmits model parameters from 10^4 simulated samples. (a,b,c) Transition from N to D and **(d,e,f)** transition from D to N parameters inferred (by MLE) from simulated data versus chosen model values when fitted to the observed data. **(a,d)** α; **(b,e)** β; **(c,f)** γ. The majority of parameter were reliably recovered.

Bibliography

[1] Ross R. An Application of the Theory of Probabilities to the Study of a priori Pathometry. Part I. *Proc. R. Soc. A Math. Phys. Eng. Sci.* **92**(638):204–230 (1916). doi:10.1098/rspa.1916.0007.

[2] Kermack WO, McKendrick AG. A Contribution to the Mathematical Theory of Epidemics. *Proc. R. Soc. London A Math. Phys. Eng. Sci.* **115**(772):700–721 (1927).

[3] Funk S, Gilad E, Watkins C, Jansen VAA. The spread of awareness and its impact on epidemic outbreaks. *Proc. Natl. Acad. Sci.* **106**(16):6872–6877 (2009). doi:10.1073/pnas.0810762106.

[4] Funk S, Salathe M, Jansen VAA. Modelling the influence of human behaviour on the spread of infectious diseases: a review. *J. R. Soc. Interface* **7**(50):1247–1256 (2010). doi:10.1098/rsif.2010.0142.

[5] Manfredi P, D'Onofrio A, editors. *Modeling the Interplay Between Human Behavior and the Spread of Infectious Diseases.* Springer New York, New York, NY (2013). doi:10.1007/978-1-4614-5474-8.

[6] Agaba G, Kyrychko Y, Blyuss K. Mathematical model for the impact of awareness on the dynamics of infectious diseases. *Math. Biosci.* **286**:22–30 (2017). doi:10.1016/j.mbs.2017.01.009.

[7] Hill AL, Rand DG, Nowak MA, Christakis NA. Emotions as infectious diseases in a large social network: the SISa model. *Proc. R. Soc. B Biol. Sci.* **277**(1701):3827–3835 (2010). doi:10.1098/rspb.2010.1217.

[8] Christakis NA, Fowler JH. Social contagion theory: examining dynamic social networks and humanbehavior. *Stat. Med.* **32**(4):556–577 (2013). doi:10.1002/sim.5408.

[9] Earn DJ, Dushoff J, Levin SA. Ecology and evolution of the flu. *Trends Ecol. Evol.* **17**(7):334–340 (2002). doi:10.1016/S0169-5347(02)02502-8.

[10] Hause BM, Collin EA, Liu R, Huang B, Sheng Z, *et al.* Characterization of a Novel Influenza Virus in Cattle and Swine: Proposal for a New Genus in the Orthomyxoviridae Family. *MBio* **5**(2):e00031–14 (2014). doi:10.1128/mBio.00031-14.

[11] Tibayrenc M. *Encyclopedia of Infectious Diseases: Modern Methodologies.* Wiley-Blackwell (2007).

[12] Jennings R, Read RC. *Influenza: Human and Avian in Practice.* Royal Society of Medicine Press (2006).

[13] Kurtz J, Manvell RJ, Banks J. Avian influenza virus isolated from a woman with conjunctivitis. *Lancet* **348**(9031):901–902 (1996). doi:10.1016/S0140-6736(05)64783-6.

[14] Gilsdorf A, Boxall N, Gasimov V, Agayev I, Mammadzade F, *et al.* Two clusters of human infection with influenza A/H5N1 virus in the Republic of Azerbaijan, February-March 2006. *Eurosurveillance* **11**(5):122–126 (2006).

[15] Webster RG. Influenza: an emerging disease. *Emerg. Infect. Dis.* **4**(3):436–441 (1998).

[16] Richard M, Fouchier RA. Influenza A virus transmission via respiratory aerosols or droplets as it relates to pandemic potential. *FEMS Microbiol. Rev.* **40**(1):68–85 (2016). doi:10.1093/femsre/fuv039.

[17] World Health Organisation. Pandemic (H1N1) 2009 - update 112. (2010). URL `http://www.who.int/csr/don/2010_08_06/en/`. [Online] (Accessed: 23 Jan 2017).

[18] World Health Organisation. Pandemic H1N1 - map of affected countries and deaths. (2010). URL `http://www.gamapserver.who.int/h1h1/cases-deaths/h1n1_casesdeaths.html`. [Online] (Accessed: 11 June 2016).

[19] World Organisation for Animal Health (OIE). Avian Influenza Portal. (2017). URL `http://www.oie.int/animal-health-in-the-world/web-portal-on-avian-influenza/`. [Online] (Accessed: 06 Feb 2017).

[20] Sims LD, Domenech J, Benigno C, Kahn S, Kamata A, *et al.* Origin and evolution of highly pathogenic H5N1 avian influenza in Asia. *Vet. Rec.* **157**(6):159–164 (2005). doi:10.1136/vr.157.6.159.

[21] FAO-DAH. H5N1 HPAI Global Overview: April-June 2012. (2012). URL `http://www.fao.org/docrep/016/ap387e/ap387e.pdf`. [Online] (Accessed: 23 Jan 2017).

[22] World Health Organisation. FAQs: H5N1 influenza. (2012). URL `http://www.who.int/influenza/human_animal_interface/avian_influenza/h5n1_research/faqs/en/`. [Online] (Accessed: 06 Feb 2017).

[23] World Health Organisation. Cumulative number of confirmed human cases of avian influenza A(H5N1) reported to WHO: 21 November 2016. (2016). URL `http://www.who.int/influenza/human_animal_interface/2016_11_21_tableH5N1.pdf`. [Online] (Accessed: 23 Jan 2017).

[24] Lai S, Qin Y, Cowling BJ, Ren X, Wardrop NA, *et al.* Global epidemiology of avian influenza A H5N1 virus infection in humans, 19972015: a systematic review of individual case data. *Lancet Infect. Dis.* **16**(7):e108–e118 (2016). doi:10.1016/S1473-3099(16)00153-5.

[25] World Health Organisation. Influenza at the Human-Animal Interface, monthly risk assessment summary: 3 October 2016. (2016). URL `http://www.who.int/influenza/human_animal_interface/Influenza_Summary_IRA_HA_interface_10_03_2016.pdf`. [Online] (Accessed: 24 Apr 2017).

[26] Chen H, Deng G, Li Z, Tian G, Li Y, *et al.* The evolution of H5N1 influenza viruses in ducks in southern China. *Proc. Natl. Acad. Sci. U. S. A.* **101**(28):10452–10457 (2004). doi:10.1073/pnas.0403212101.

[27] Chen H, Smith GJD, Li KS, Wang J, Fan XH, *et al.* Establishment of multiple sublineages of H5N1 influenza virus in Asia: implications for pandemic control. *Proc. Natl. Acad. Sci. U. S. A.* **103**(8):2845–2850 (2006). doi:10.1073/pnas.0511120103.

[28] Jeong OM, Kim MC, Kim MJ, Kang HM, Kim HR, *et al.* Experimental infection of chickens, ducks and quails with the highly pathogenic H5N1 avian influenza virus. *J. Vet. Sci.* **10**(1):53 (2009). doi:10.4142/jvs.2009.10.1.53.

[29] Pfeiffer DU, Minh PQ, Martin V, Epprecht M, Otte MJ. An analysis of the spatial and temporal patterns of highly pathogenic avian influenza occurrence in Vietnam using national surveillance data. *Vet. J.* **174**(2):302–309 (2007). doi:10.1016/j.tvjl.2007.05.010.

[30] Loth L, Gilbert M, Osmani MG, Kalam AM, Xiao X. Risk factors and clusters of Highly Pathogenic Avian Influenza H5N1 outbreaks in Bangladesh. *Prev. Vet. Med.* **96**(1-2):104–113 (2010). doi:10.1016/j.prevetmed.2010.05.013.

[31] Minh PQ, Stevenson MA, Jewell C, French N, Schauer B. Spatio-temporal analyses of highly pathogenic avian influenza H5N1 outbreaks in the Mekong River Delta, Vietnam, 2009. *Spat. Spatiotemporal. Epidemiol.* **2**(1):49–57 (2011). doi:10.1016/j.sste.2010.11.001.

[32] Li R, Jiang Z, Xu B. Global spatiotemporal and genetic footprint of the H5N1 avian influenza virus. *Int. J. Health Geogr.* **13**:14 (2014). doi:10.1186/1476-072X-13-14.

[33] Dhingra MS, Dissanayake R, Negi AB, Oberoi M, Castellan D, *et al.* Spatio-temporal epidemiology of highly pathogenic avian influenza (subtype H5N1) in poultry in eastern India. *Spat. Spatiotemporal. Epidemiol.* **11**:45–57 (2014). doi:10.1016/j.sste.2014.06.003.

[34] Gilbert M, Xiao X, Chaitaweesub P, Kalpravidh W, Premashthira S, *et al.* Avian influenza, domestic ducks and rice agriculture in Thailand. *Agric. Ecosyst. \& Environ.* **119**(3-4):409–415 (2007). doi:10.1016/j.agee.2006.09.001.

[35] Gilbert M, Xiao X, Pfeiffer DU, Epprecht M, Boles S, *et al.* Mapping H5N1 highly pathogenic avian influenza risk in Southeast Asia. *Proc. Natl. Acad. Sci. U. S. A.* **105**(12):4769–4774 (2008). doi:10.1073/pnas.0710581105.

[36] Henning KA, Henning J, Morton J, Long NT, Ha NT, *et al.* Farm- and flock-level risk factors associated with Highly Pathogenic Avian Influenza outbreaks on small holder duck and chicken farms in the Mekong Delta of Viet Nam. *Prev. Vet. Med.* **91**(2-4):179–188 (2009). doi:10.1016/j.prevetmed.2009.05.027.

[37] Martin V, Pfeiffer DU, Zhou X, Xiao X, Prosser DJ, *et al.* Spatial distribution and risk factors of highly pathogenic avian influenza (HPAI) H5N1 in China. *PLoS Pathog.* **7**(3):e1001308 (2011). doi:10.1371/journal.ppat.1001308.

[38] Van Boeckel TP, Thanapongtharm W, Robinson T, Biradar CM, Xiao X, *et al.* Improving risk models for avian influenza: the role of intensive poultry farming and flooded land during the 2004 Thailand epidemic. *PLoS One* **7**(11):e49528 (2012). doi:10.1371/journal.pone.0049528.

[39] Gilbert M, Pfeiffer DU. Risk factor modelling of the spatio-temporal patterns of highly pathogenic avian influenza (HPAIV) H5N1: a review. *Spat. Spatiotemporal. Epidemiol.* **3**(3):173–183 (2012). doi:10.1016/j.sste.2012.01.002.

[40] Tiensin T, Nielen M, Vernooij H, Songserm T, Kalpravidh W, *et al.* Transmission of the highly pathogenic avian influenza virus H5N1 within flocks during the 2004 epidemic in Thailand. *J. Infect. Dis.* **196**(11):1679–1684 (2007). doi:10.1086/522007.

[41] Bouma A, Claassen I, Natih K, Klinkenberg D, Donnelly CA, *et al.* Estimation of transmission parameters of H5N1 avian influenza virus in chickens. *PLoS Pathog.* **5**(1):e1000281 (2009). doi:10.1371/journal.ppat.1000281.

[42] World Health Organisation. Frequently Asked Questions on human infection caused by the avian influenza A(H7N9) virus. (2014). URL `http://www.who.int/influenza/human_animal_interface/faq_H7N9/en/`. [Online] (Accessed: 23 Jan 2017).

[43] Gong Z, Lv H, Ding H, Han J, Sun J, *et al.* Epidemiology of the avian

influenza A (H7N9) outbreak in Zhejiang Province, China. *BMC Infect. Dis.* **14**:244 (2014). doi:10.1186/1471-2334-14-244.

[44] Watanabe T, Watanabe S, Maher EA, Neumann G, Kawaoka Y. Pandemic potential of avian influenza A (H7N9) viruses. *Trends Microbiol.* **22**(11):623–631 (2014). doi:10.1016/j.tim.2014.08.008.

[45] World Health Organisation. Influenza at the Human-Animal Interface, monthly risk assessment summary: 20 April 2017. (2017). URL `http://www.who.int/influenza/human_animal_interface/Influenza_Summary_IRA_HA_interface_04_20_2017.pdf`. [Online] (Accessed: 24 Apr 2017).

[46] Food and Agriculture Organization. H7N9 situation update. (2017). URL `www.fao.org/ag/againfo/programmes/en/empres/H7N9/situation_update.html`. [Online] (Accessed: 27 Jan 2017).

[47] Wu P, Jiang H, Wu JT, Chen E, He J, *et al.* Poultry Market Closures and Human Infection with Influenza A(H7N9) Virus, China, 201314. *Emerg. Infect. Dis.* **20**(11):1891–1894 (2014). doi:10.3201/eid2011.140556.

[48] Centers for Disease Control and Prevention. Images of Avian Influenza A H7N9. (2013). URL `http://www.cdc.gov/flu/avianflu/h7n9-images.html`. [Online] (Accessed: 15 May 2017).

[49] Shen H, Wu B, Li G, Chen F, Luo Q, *et al.* H9N2 Subtype Avian Influenza Viruses in China: Current Advances and Future Perspectives. *Br. J. Virol.* **1**(2):54–63 (2014).

[50] Li X, Shi J, Guo J, Deng G, Zhang Q, *et al.* Genetics, Receptor Binding Property, and Transmissibility in Mammals of Naturally Isolated H9N2 Avian Influenza Viruses. *PLoS Pathog.* **10**(11):e1004508 (2014). doi:10.1371/journal.ppat.1004508.

[51] Food and Agriculture Organization. H5N8 HPAI Global situation update. (2017). URL `http://www.fao.org/ag/againfo/programmes/en/empres/H5N8/situation_update.html`. [Online] (Accessed: 27 Jan 2017).

[52] McCloskey B, Dar O, Zumla A, Heymann DL. Emerging infectious diseases and pandemic potential: status quo and reducing risk of global spread. *Lancet Infect. Dis.* **14**(10):1001–1010 (2014). doi:10.1016/S1473-3099(14)70846-1.

[53] Al-Tawfiq JA, Zumla A, Gautret P, Gray GC, Hui DS, *et al.* Surveillance for emerging respiratory viruses. *Lancet Infect. Dis.* **14**(10):992–1000 (2014). doi:10.1016/S1473-3099(14)70840-0.

[54] Von Dobschuetz S, De Nardi M, Harris KA, Munoz O, Breed AC, *et al.* Influenza surveillance in animals: what is our capacity to detect emerging in-

fluenza viruses with zoonotic potential? *Epidemiol. Infect.* **143**(10):2187–2204 (2015). doi:10.1017/S0950268814002106.

[55] Short KR, Richard M, Verhagen JH, van Riel D, Schrauwen EJ, *et al.* One health, multiple challenges: The inter-species transmission of influenza A virus. *One Heal.* **1**:1–13 (2015). doi:10.1016/j.onehlt.2015.03.001.

[56] Ludwig S, Zell R, Schwemmle M, Herold S. Influenza, a One Health paradigm-Novel therapeutic strategies to fight a zoonotic pathogen with pandemic potential. *Int. J. Med. Microbiol.* **304**(7):894–901 (2014). doi:10.1016/j.ijmm.2014.08.016.

[57] Russell CA, Kasson PM, Donis RO, Riley S, Dunbar J, *et al.* Improving pandemic influenza risk assessment. *Elife* **3**:e03883 (2014). doi:10.7554/eLife.03883.

[58] Trock SC, Burke SA, Cox NJ. Development of Framework for Assessing Influenza Virus Pandemic Risk. *Emerg. Infect. Dis.* **21**(8):1372–8 (2015). doi:10.3201/eid2108.141086.

[59] World Health Organisation. Noncommunicable diseases - Fact Sheet No.355. (2015). URL `http://www.who.int/mediacentre/factsheets/fs355/en/`. [Online] (Accessed: 27 Jan 2017).

[60] World Health Organisation. Global health risks: Mortality and burden of disease attributable to selected major risks. (2009). URL `http://www.who.int/healthinfo/global_burden_disease/GlobalHealthRisks_report_full.pdf`. [Online] (Accessed: 27 Jan 2017).

[61] World Health Organisation. Obesity and overweight - Fact Sheet No.311. (2016). URL `http://www.who.int/mediacentre/factsheets/fs311/en/`. [Online] (Accessed: 27 Jan 2017).

[62] Moussavi S, Chatterji S, Verdes E, Tandon A, Patel V, *et al.* Depression, chronic diseases, and decrements in health: results from the World Health Surveys. *Lancet* **370**(9590):851–858 (2007). doi:10.1016/S0140-6736(07)61415-9.

[63] Vos T, Allen C, Arora M, Barber RM, Bhutta ZA, *et al.* Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990-2015: a systematic analysis for the Global Burden of Disease Study 2015. *Lancet* **388**(10053):1545–1602 (2016). doi:10.1016/S0140-6736(16)31678-6.

[64] World Health Organisation. Depression and Other Common Mental Disorders: Global Health Estimates. (2017). URL `http://apps.who.int/iris/bitstream/10665/254610/1/WHO-MSD-MER-2017.2-eng.pdf`. [Online] (Ac-

cessed: 16 May 2017).

[65] World Health Organisation. Mental health - Suicide data. (2017). URL `http://www.who.int/mental_health/prevention/suicide/suicideprevent/en/`. [Online] (Accessed: 16 May 2017).

[66] Office for National Statistics. Social and Vital Statistics Division, Office for National Statistics. Health and Care Division, Ford, T., Goodman, R. Mental Health of Children and Young People in Great Britain, 2004. (2015). doi: 10.5255/UKDA-SN-5269-1. [data collection]. UK Data Service. SN: 5269.

[67] National Institute of Mental Health. Major Depression Among Adolescents. (2016). URL `https://www.nimh.nih.gov/health/statistics/prevalence/major-depression-among-adolescents.shtml`. [Online] (Accessed: 16 May 2017).

[68] World Health Organisation. Mental disorders - Fact Sheet No.396. (2016). URL `http://www.who.int/mediacentre/factsheets/fs396/en/`. [Online] (Accessed: 27 Jan 2017).

[69] Bartlett MS. *Stochastic Population Models in Ecology and Epidemiology*. Methuen's monographs on applied probability and statistics. Methuen (1960).

[70] Anderson RM, May RM. *Infectious Diseases of Humans: Dynamics and Control*. Dynamics and Control. OUP Oxford (1991).

[71] Diekmann O, Verduyn Lunel SM, van Gils SA, Walther HO. *Delay Equations*, volume 110 of *Applied Mathematical Sciences*. Springer New York, New York, NY (1995). doi:10.1007/978-1-4612-4206-2.

[72] Keeling M, Rohani P. *Modeling infectious diseases in humans and animals*. Princeton University Press (2008).

[73] Diekmann O, Heesterbeek JAP, Metz JAJ. On the definition and the computation of the basic reproduction ratio R 0 in models for infectious diseases in heterogeneous populations. *J. Math. Biol.* **28**(4):365–382 (1990). doi: 10.1007/BF00178324.

[74] De Jong M, Diekmann O, Heesterbeek J. The computation of R0 for discrete-time epidemic models with dynamic heterogeneity. *Math. Biosci.* **119**(1):97–114 (1994). doi:10.1016/0025-5564(94)90006-X.

[75] Fine PE. Herd Immunity: History, Theory, Practice. *Epidemiol. Rev.* **15**(2):265–302 (1993). doi:10.1093/oxfordjournals.epirev.a036121.

[76] Keeling MJ, Woolhouse ME, Shaw DJ, Matthews L, Chase-Topping M, *et al.* Dynamics of the 2001 UK foot and mouth epidemic: stochastic dispersal in a heterogeneous landscape. *Science* **294**(5543):813–817 (2001). doi:10.1126/science.1065973.

[77] Danon L, Ford AP, House T, Jewell CP, Keeling MJ, *et al.* Networks and the Epidemiology of Infectious Disease. *Interdiscip. Perspect. Infect. Dis.* **2011**:1–28 (2011). doi:10.1155/2011/284909.

[78] Christakis NA, Fowler JH. The Spread of Obesity in a Large Social Network over 32 Years. *N. Engl. J. Med.* **357**(4):370–379 (2007). doi:10.1056/NEJMsa066082.

[79] Christakis NA, Fowler JH. The Collective Dynamics of Smoking in a Large Social Network. *N. Engl. J. Med.* **358**(21):2249–2258 (2008). doi:10.1056/NEJMsa0706154.

[80] Bartlett MS. Deterministic and Stochastic Models for Recurrent Epidemics. In *Proc. Third Berkeley Symp. Math. Stat. Probab. Vol. 4 Contrib. to Biol. Probl. Heal.*, Third Berkeley Symposium on Mathematical Statistics and Probability, pages 81–109. University of California Press, Berkeley, Calif. (1956). ISBN 0097-0433.

[81] Olsen L, Truty G, Schaffer W. Oscillations and chaos in epidemics: A nonlinear dynamic study of six childhood diseases in Copenhagen, Denmark. *Theor. Popul. Biol.* **33**(3):344–370 (1988). doi:10.1016/0040-5809(88)90019-6.

[82] Keeling MJ, Woolhouse MEJ, May RM, Davies G, Grenfell BT. Modelling vaccination strategies against foot-and-mouth disease. *Nature* **421**(6919):136–142 (2003). doi:10.1038/nature01343.

[83] Tildesley MJ, Savill NJ, Shaw DJ, Deardon R, Brooks SP, *et al.* Optimal reactive vaccination strategies for a foot-and-mouth outbreak in the UK. *Nature* **440**(7080):83–86 (2006). doi:10.1038/nature04324.

[84] Jones KE, Patel NG, Levy MA, Storeygard A, Balk D, *et al.* Global trends in emerging infectious diseases. *Nature* **451**(7181):990–993 (2008). doi:10.1038/nature06536.

[85] Lloyd-Smith JO, George D, Pepin KM, Pitzer VE, Pulliam JRC, *et al.* Epidemic Dynamics at the Human-Animal Interface. *Science* **326**(5958):1362–1367 (2009). doi:10.1126/science.1177345.

[86] Hsieh Yh, Wu J, Fang J, Yang Y, Lou J. Quantification of Bird-to-Bird and Bird-to-Human Infections during 2013 Novel H7N9 Avian Influenza Outbreak in China. *PLoS One* **9**(12):e111834 (2014). doi:10.1371/journal.pone.0111834.

[87] Li R, Bai Y, Heaney A, Kandula S, Cai J, *et al.* Inference and forecast of H7N9 influenza in China, 2013 to 2015. *Eurosurveillance* **22**(7):30462 (2017). doi:10.2807/1560-7917.ES.2017.22.7.30462.

[88] Lloyd-Smith JO, Funk S, McLean AR, Riley S, Wood JL. Nine challenges in modelling the emergence of novel pathogens. *Epidemics* **10**:35–39 (2015).

doi:10.1016/j.epidem.2014.09.002.

[89] Klous G, Huss A, Heederik DJ, Coutinho RA. Human-livestock contacts and their relationship to transmission of zoonotic pathogens, a systematic review of literature. *One Heal.* (2016). doi:10.1016/j.onehlt.2016.03.001.

[90] Centola D. The Spread of Behavior in an Online Social Network Experiment. *Science* **329**(5996):1194–1197 (2010). doi:10.1126/science.1185231.

[91] Ali MM, Amialchuk A, Gao S, Heiland F. Adolescent weight gain and social networks: is there a contagion effect? *Appl. Econ.* **44**(23):2969–2983 (2012). doi:10.1080/00036846.2011.568408.

[92] Cacioppo JT, Fowler JH, Christakis NA. Alone in the crowd: The structure and spread of loneliness in a large social network. *J. Pers. Soc. Psychol.* **97**(6):977–991 (2009). doi:10.1037/a0016076.

[93] Balbo N, Barban N. Does Fertility Behavior Spread among Friends? *Am. Sociol. Rev.* **79**(3):412–431 (2014). doi:10.1177/0003122414531596.

[94] Fowler JH, Christakis NA. Dynamic spread of happiness in a large social network: longitudinal analysis over 20 years in the Framingham Heart Study. *BMJ* **337**:a2338–a2338 (2008). doi:10.1136/bmj.a2338.

[95] Cohen-Cole E, Fletcher JM. Detecting implausible social network effects in acne, height, and headaches: longitudinal analysis. *BMJ* **337**:a2533 (2008). doi:10.1136/bmj.a2533.

[96] Aral S, Muchnik L, Sundararajan A. Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. *Proc. Natl. Acad. Sci.* **106**(51):21544–21549 (2009). doi:10.1073/pnas.0908800106.

[97] Lyons R. The Spread of Evidence-Poor Medicine via Flawed Social-Network Analysis. *Stat. Polit. Policy* **2**(1) (2011). doi:10.2202/2151-7509.1024.

[98] Thomas AC. The social contagion hypothesis: comment on 'Social contagion theory: examining dynamic social networks and human behavior'. *Stat. Med.* **32**(4):581–590 (2013). doi:10.1002/sim.5551.

[99] Fisher RA. On the Mathematical Foundations of Theoretical Statistics. *Philos. Trans. R. Soc. London A Math. Phys. Eng. Sci.* **222**(594-604):309–368 (1922). doi:10.1098/rsta.1922.0009.

[100] Aldrich J. R.A. Fisher and the making of maximum likelihood 1912-1922. *Stat. Sci.* **12**(3):162–176 (1997). doi:10.1214/ss/1030037906.

[101] O'Neill PD, Roberts GO. Bayesian inference for partially observed stochastic epidemics. *J. R. Stat. Soc. Ser. A (Statistics Soc.)* **162**(1):121–129 (1999). doi:10.1111/1467-985X.00125.

[102] Neal PJ, Roberts GO. Statistical inference and model selection for the 1861

Hagelloch measles epidemic. *Biostatistics* **5**(2):249–261 (2004).

[103] Deardon R, Brooks SP, Grenfell BT, Keeling MJ, Tildesley MJ, *et al.* Inference for individual-level models of infectious diseases in large populations. *Stat. Sin.* **20**(1):239–261 (2010).

[104] Walker DM, Allingham D, Lee HWJ, Small M. Parameter inference in small world network disease models with approximate Bayesian Computational methods. *Phys. A Stat. Mech. its Appl.* **389**(3):540–548 (2010). doi: 10.1016/j.physa.2009.09.053.

[105] McKinley TJ, Ross JV, Deardon R, Cook AR. Simulation-based Bayesian inference for epidemic models. *Comput. Stat. Data Anal.* **71**:434–447 (2014). doi:10.1016/j.csda.2012.12.012.

[106] Gilks W, Richardson S, Spiegelhalter D. *Markov Chain Monte Carlo in Practice*. Chapman and Hall (1996).

[107] Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E. Equation of State Calculations by Fast Computing Machines. *J. Chem. Phys.* **21**(6):1087–1092 (1953). doi:10.1063/1.1699114.

[108] Hastings WK. Monte Carlo Sampling Methods Using Markov Chains and Their Applications. *Biometrika* **57**:97–109 (1970). doi:10.1093/biomet/57.1.97.

[109] Neal RM. Slice sampling. *Ann. Stat.* **31**(3):705–767 (2003). doi:10.1214/aos/1056562461.

[110] Haario H, Saksman E, Tamminen J. An adaptive Metropolis algorithm. *Bernoulli* **7**(2):223–242 (2001).

[111] Duane S, Kennedy AD, Pendleton BJ, Roweth D. Hybrid Monte Carlo. *Phys. Lett. B* **195**(2):216–222 (1987). doi:10.1016/0370-2693(87)91197-X.

[112] Homan MD, Gelman A. The No-U-turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo. *J. Mach. Learn. Res.* **15**(1):1593–1623 (2014).

[113] Green PJ. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* **82**(4):711–732 (1995). doi: 10.1093/biomet/82.4.711.

[114] Tildesley MJ, Bessell PR, Keeling MJ, Woolhouse MEJ. The role of pre-emptive culling in the control of foot-and-mouth disease. *Proc. R. Soc. B Biol. Sci.* **276**(1671):3239–3248 (2009). doi:10.1098/rspb.2009.0427.

[115] Rohani P, Zhong X, King AA. Contact Network Structure Explains the Changing Epidemiology of Pertussis. *Science* **330**(6006):982–985 (2010). doi: 10.1126/science.1194134.

[116] Bartlett MS. Measles Periodicity and Community Size. *J. R. Stat. Soc. Ser. A* **120**(1):48–70 (1957). doi:10.2307/2342553.

[117] Renshaw E. *Modelling Biological Populations in Space and Time*. Cambridge Studies in Mathemati. Cambridge University Press (1993).

[118] Gillespie DT. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J. Comput. Phys.* **22**(4):403–434 (1976). doi:10.1016/0021-9991(76)90041-3.

[119] Gillespie DT. Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.* **81**(25):2340–2361 (1977). doi:10.1021/j100540a008.

[120] Gillespie DT. Approximate accelerated stochastic simulation of chemically reacting systems. *J. Chem. Phys.* **115**(4):1716–1733 (2001). doi:10.1063/1.1378322.

[121] Sellke T. On the asymptotic distribution of the size of a stochastic epidemic. *J. Appl. Probab.* **20**(2):390–394 (1983).

[122] Alexander DJ. Summary of Avian Influenza Activity in Europe, Asia, Africa, and Australasia, 20022006. *Avian Dis.* **51**(s1):161–166 (2007). doi:10.1637/7602-041306R.1.

[123] Ahmed SSU, Ersbøll AK, Biswas PK, Christensen JP. The spacetime clustering of highly pathogenic avian influenza (HPAI) H5N1 outbreaks in Bangladesh. *Epidemiol. Infect.* **138**(06):843 (2010). doi:10.1017/S0950268810000178.

[124] Ahmed SSU, Ersbøll AK, Biswas PK, Christensen JP, Toft N. Spatio-Temporal Magnitude and Direction of Highly Pathogenic Avian Influenza (H5N1) Outbreaks in Bangladesh. *PLoS One* **6**(9):e24324 (2011). doi:10.1371/journal.pone.0024324.

[125] United Nations: Department of Economic and Social Affairs: Population Division. World Population Prospects, the 2015 Revision. (2015). URL `http://www.un.org/en/development/desa/population/theme/trends/index.shtml`. [Online] (Accessed: 23 Jan 2017).

[126] The World Bank. Implementation completion and results report (IDA-43400 TF- 90662) on a credit in the amount of SDR 10.5 million (US$16.0 million equivalent) to the People's Republic of Bangladesh for an avian influenza preparedness and response project under the global. (2013). URL `http://documents.worldbank.org/curated/en/651381468210870082/pdf/ICR21770ICR0Av0Box0377341B00PUBLIC0.pdf`. [Online] (Accessed: 23 Jan 2017).

[127] Biswas PK, Christensen JP, Ahmed SSU, Barua H, Das A, *et al.* Risk factors

for infection with highly pathogenic influenza A virus (H5N1) in commercial chickens in Bangladesh. *Vet. Rec.* **164**(24):743–746 (2009). doi:10.1136/vr.164.24.743.

[128] Osmani MG, Thornton RN, Dhand NK, Hoque MA, Milon SMA, *et al.* Risk Factors for Highly Pathogenic Avian Influenza in Commercial Layer Chicken Farms in Bangladesh During 2011. *Transbound. Emerg. Dis.* **61**(6):e44–e51 (2014). doi:10.1111/tbed.12071.

[129] Gilbert M, Newman SH, Takekawa JY, Loth L, Biradar C, *et al.* Flying Over an Infected Landscape: Distribution of Highly Pathogenic Avian Influenza H5N1 Risk in South Asia and Satellite Tracking of Wild Waterfowl. *Ecohealth* **7**(4):448–458 (2010). doi:10.1007/s10393-010-0672-8.

[130] Ahmed SSU, Ersbøll AK, Biswas PK, Christensen JP, Hannan ASMA, *et al.* Ecological Determinants of Highly Pathogenic Avian Influenza (H5N1) Outbreaks in Bangladesh. *PLoS One* **7**(3):e33938 (2012). doi:10.1371/journal.pone.0033938.

[131] Biswas PK, Christensen JP, Ahmed SS, Das A, Rahman MH, *et al.* Risk for Infection with Highly Pathogenic Avian Influenza Virus (H5N1) in Backyard Chickens, Bangladesh. *Emerg. Infect. Dis.* **15**(12):1931–1936 (2009). doi:10.3201/eid1512.090643.

[132] Osmani MG, Ward MP, Giasuddin M, Islam MR, Kalam A. The spread of highly pathogenic avian influenza (subtype H5N1) clades in Bangladesh, 2010 and 2011. *Prev. Vet. Med.* **114**(1):21–27 (2014). doi:10.1016/j.prevetmed.2014.01.010.

[133] Dejpichai R, Laosiritaworn Y, Phuthavathana P, Uyeki TM, O'Reilly M, *et al.* Seroprevalence of antibodies to avian influenza virus A (H5N1) among residents of villages with human cases, Thailand, 2005. *Emerg. Infect. Dis.* **15**(5):756–760 (2009). doi:10.3201/eid1505.080316.

[134] Nasreen S, Khan SU, Luby SP, Gurley ES, Abedin J, *et al.* Highly Pathogenic Avian Influenza A(H5N1) Virus Infection among Workers at Live Bird Markets, Bangladesh, 20092010. *Emerg. Infect. Dis.* **21**(4):629–637 (2015). doi:10.3201/eid2104.141281.

[135] OIE. Highly pathogenic avian influenza, Bangladesh, (Follow-up report No. 43). *OIE, WAHID(World Anim. Heal. Inf. Database), Wkly. Dis. Inf.* **26**(52) (2013).

[136] Robinson T, Wint G, Conchedda G, Van Boeckel T, Ercoli V, *et al.* Mapping the global distribution of livestock. *PLoS One* **9**(5):e96084 (2014). doi:10.1371/journal.pone.0096084.

[137] Arino O, Ramos Perez JJ, Kalogirou V, Bontemps S, Defourny P, *et al.* Global Land Cover Map for 2009 (GlobCover 2009) (2012). doi:10.1594/PANGAEA.787668.

[138] Xiao X, Boles S, Liu J, Zhuang D, Frolking S, *et al.* Mapping paddy rice agriculture in southern China using multi-temporal MODIS images. *Remote Sens. Environ.* **95**(4):480–492 (2005). doi:10.1016/j.rse.2004.12.009.

[139] Xiao X, Boles S, Frolking S, Li C, Babu JY, *et al.* Mapping paddy rice agriculture in South and Southeast Asia using multi-temporal MODIS images. *Remote Sens. Environ.* **100**(1):95–113 (2006). doi:10.1016/j.rse.2005.10.004.

[140] Zhang G, Xiao X, Dong J, Kou W, Jin C, *et al.* Mapping paddy rice planting areas through time series analysis of MODIS land surface temperature and vegetation index data. *ISPRS J. Photogramm. Remote Sens.* **106**:157–171 (2015). doi:10.1016/j.isprsjprs.2015.05.011.

[141] FAO. EMPRES-i. (n.d.). URL `http://empres-i.fao.org/eipws3g/`. [Online] (Accessed: 23 Jan 2017).

[142] Biswas PK, Christensen JP, Ahmed SSU, Barua H, Das A, *et al.* Mortality rate and clinical features of highly pathogenic avian influenza in naturally infected chickens in Bangladesh. *Rev. Sci. Tech.* **30**(3):871–878 (2011).

[143] Food and Agriculture Organization of the United Nations. Approaches to controlling, preventing and eliminating H5N1 Highly Pathogenic Avian Influenza in endemic countries. *FAO Anim. Prod. Heal. Pap.* (No. 171) (2011).

[144] Hancock PA, Rehman Y, Hall IM, Edeghere O, Danon L, *et al.* Strategies for Controlling Non-Transmissible Infection Outbreaks Using a Large Human Movement Data Set. *PLoS Comput. Biol.* **10**(9):e1003809 (2014). doi:10.1371/journal.pcbi.1003809.

[145] Farnsworth ML, Hamilton-West C, Fitchett S, Newman SH, de La Rocque S, *et al.* Comparing national and global data collection systems for reporting, outbreaks of H5N1 HPAI. *Prev. Vet. Med.* **95**(3-4):175–185 (2010). doi:10.1016/j.prevetmed.2010.03.012.

[146] Lindström T, Sisson SA, Nöremark M, Jonsson A, Wennergren U. Estimation of distance related probability of animal movements between holdings and implications for disease spread modeling. *Prev. Vet. Med.* **91**(2-4):85–94 (2009). doi:10.1016/j.prevetmed.2009.05.022.

[147] Tildesley MJ, Deardon R, Savill NJ, Bessell PR, Brooks SP, *et al.* Accuracy of models for the 2001 foot-and-mouth epidemic. *Proc. R. Soc. London B Biol. Sci.* **275**(1641):1459–1468 (2008). doi:10.1098/rspb.2008.0006.

[148] Spiegelhalter DJ, Best NG, Carlin BP, van der Linde A. Bayesian measures

of model complexity and fit. *J. R. Stat. Soc. Ser. B (Statistical Methodol.)* **64**(4):583–639 (2002). doi:10.1111/1467-9868.00353.

[149] Gelman A, Carlin J, Stern H, Rubin D. *Bayesian Data Analysis, Second Edition*. Chapman and Hall/CRC (2003).

[150] Rabinowitz PM, Galusha D, Vegso S, Michalove J, Rinne S, *et al.* Comparison of Human and Animal Surveillance Data for H5N1 Influenza A in Egypt 20062011. *PLoS One* **7**(9):e43851 (2012). doi:10.1371/journal.pone.0043851.

[151] Ripley BD. The Second-Order Analysis of Stationary Point Processes. *J. Appl. Probab.* **13**(2):255–266 (1976). doi:10.2307/3212829.

[152] Ripley BD. Modelling Spatial Patterns. *J. R. Stat. Soc. Ser. B* **39**(2):172–212 (1977).

[153] Tildesley MJ, Keeling MJ. Is R_0 a good predictor of final epidemic size: Foot-and-Mouth Disease in the UK. *J. Theor. Biol.* **258**(4):623–629 (2009). doi:10.1016/j.jtbi.2009.02.019.

[154] Ansari WK, Parvej MS, El Zowalaty ME, Jackson S, Bustin SA, *et al.* Surveillance, epidemiological, and virological detection of highly pathogenic H5N1 avian influenza viruses in duck and poultry from Bangladesh. *Vet. Microbiol.* **193**:49–59 (2016). doi:10.1016/j.vetmic.2016.07.025.

[155] Biswas PK, Giasuddin M, Nath BK, Islam MZ, Debnath NC, *et al.* Biosecurity and Circulation of Influenza A (H5N1) Virus in Live-Bird Markets in Bangladesh, 2012. *Transbound. Emerg. Dis.* **64**(3):883–891 (2017). doi:10.1111/tbed.12454.

[156] Islam MR, Haque ME, Giasuddin M, Chowdhury EH, Samad MA, *et al.* New Introduction of Clade 2.3.2.1 Avian Influenza Virus (H5N1) into Bangladesh. *Transbound. Emerg. Dis.* **59**(5):460–463 (2012). doi:10.1111/j.1865-1682.2011.01297.x.

[157] World Health Organization/World Organisation for Animal Health/Food and Agriculture Organization (WHO/OIE/FAO) H5N1 Evolution Working Group. Revised and updated nomenclature for highly pathogenic avian influenza A (H5N1) viruses. *Influenza Other Respi. Viruses* **8**(3):384–388 (2014). doi:10.1111/irv.12230.

[158] Food and Agriculture Organization of the United Nations. Poultry Sector Country Review Bangladesh. *FAO Anim. Prod. Heal. Livest. Ctry. Rev.* (2008).

[159] World Health Organisation. Influenza at the Human-Animal Interface: Summary and assessment as of 5 March 2012. (2012). URL `http://www.who.int/influenza/human_animal_interface/Influenza_`

Summary_IRA_HA_interface_05March12.pdf?ua=1. [Online] (Accessed: 23 Jan 2017).

[160] Jewell CP, Kypraios T, Neal P, Roberts GO. Bayesian analysis for emerging infectious diseases. *Bayesian Anal.* **4**(3):465–496 (2009). doi:10.1214/09-BA417.

[161] Jewell C, Kypraios T, Christley R, Roberts G. A novel approach to real-time risk prediction for emerging infectious diseases: A case study in Avian Influenza H5N1. *Prev. Vet. Med.* **91**(1):19–28 (2009). doi:10.1016/j.prevetmed.2009.05.019.

[162] World Organisation for Animal Health (OIE). Terrestrial Animal Health Code, 24th Ed. OIE, Paris. (2015). URL http://www.oie.int/international-standard-setting/terrestrial-code/access-online/. [Online] (Accessed: 23 Jan 2017).

[163] Food and Agriculture Organization of the United Nations. Biosecurity for Highly Pathogenic Avian Influenza. *FAO Anim. Prod. Heal. Pap.* (No. 165) (2008).

[164] Soares Magalhaes RJ, Pfeiffer DU, Otte J. Evaluating the control of HPAIV H5N1 in Vietnam: virus transmission within infected flocks reported before and after vaccination. *BMC Vet. Res.* **6**:31 (2010). doi:10.1186/1746-6148-6-31.

[165] Liu Q, Mena I, Ma J, Bawa B, Krammer F, *et al.* Newcastle Disease Virus-Vectored H7 and H5 Live Vaccines Protect Chickens from Challenge with H7N9 or H5N1 Avian Influenza Viruses. *J. Virol.* **89**(14):7401–7408 (2015). doi:10.1128/JVI.00031-15.

[166] Nelson MI, Vincent AL. Reverse zoonosis of influenza to swine: new perspectives on the humananimal interface. *Trends Microbiol.* **23**(3):142–153 (2015). doi:10.1016/j.tim.2014.12.002.

[167] Nelson MI, Schaefer R, Gava D, Cantão ME, Ciacci-Zanella JR. Influenza A Viruses of Human Origin in Swine, Brazil. *Emerg. Infect. Dis.* **21**(8):1339–1347 (2015). doi:10.3201/eid2108.141891.

[168] Ren H, Jin Y, Hu M, Zhou J, Song T, *et al.* Ecological dynamics of influenza A viruses: cross-species transmission and global migration. *Sci. Rep.* **6**:36839 (2016). doi:10.1038/srep36839.

[169] Food and Agriculture Organization of the United Nations. *Manual on livestock disease surveillance and information systems.* FAO animal health manual. Food and Agriculture Organization of the United Nations (1999).

[170] Probert WJ, Shea K, Fonnesbeck CJ, Runge MC, Carpenter TE, *et al.* Decision-making for foot-and-mouth disease control: Objectives matter. *Epi-*

demics **15**:10–19 (2016). doi:10.1016/j.epidem.2015.11.002.

[171] Machalaba CC, Elwood SE, Forcella S, Smith KM, Hamilton K, *et al.* Global Avian Influenza Surveillance in Wild Birds: A Strategy to Capture Viral Diversity. *Emerg. Infect. Dis.* **21**(4) (2015). doi:10.3201/eid2104.141415.

[172] Zhang WD, Zu ZH, Xu Q, Xu ZJ, Liu JJ, *et al.* Optimized Strategy for the Control and Prevention of Newly Emerging Influenza Revealed by the Spread Dynamics Model. *PLoS One* **9**(1):e84694 (2014). doi:10.1371/journal.pone.0084694.

[173] Stephenson I, Nicholson KG, Colegate A, Podda A, Wood J, *et al.* Boosting immunity to influenza H5N1 with MF59-adjuvanted H5N3 A/Duck/Singapore/97 vaccine in a primed human population. *Vaccine* **21**(15):1687–1693 (2003). doi:10.1016/S0264-410X(02)00632-1.

[174] Goji NA, Nolan C, Hill H, Wolff M, Noah DL, *et al.* Immune Responses of Healthy Subjects to a Single Dose of Intramuscular Inactivated Influenza A/Vietnam/1203/2004 (H5N1) Vaccine after Priming with an Antigenic Variant. *J. Infect. Dis.* **198**(5):635–641 (2008). doi:10.1086/590916.

[175] Ledgerwood JE, Zephir K, Hu Z, Wei CJ, Chang L, *et al.* Prime-Boost Interval Matters: A Randomized Phase 1 Study to Identify the Minimum Interval Necessary to Observe the H5 DNA Influenza Vaccine Priming Effect. *J. Infect. Dis.* **208**(3):418–422 (2013). doi:10.1093/infdis/jit180.

[176] Babu TM, Levine M, Fitzgerald T, Luke C, Sangster MY, *et al.* Live attenuated H7N7 influenza vaccine primes for a vigorous antibody response to inactivated H7N7 influenza vaccine. *Vaccine* **32**(50):6798–6804 (2014). doi:10.1016/j.vaccine.2014.09.070.

[177] Talaat KR, Luke CJ, Khurana S, Manischewitz J, King LR, *et al.* A Live Attenuated Influenza A(H5N1) Vaccine Induces Long-Term Immunity in the Absence of a Primary Antibody Response. *J. Infect. Dis.* **209**(12):1860–1869 (2014). doi:10.1093/infdis/jiu123.

[178] Ledgerwood JE, Wei CJ, Hu Z, Gordon IJ, Enama ME, *et al.* DNA priming and influenza vaccine immunogenicity: two phase 1 open label randomised clinical trials. *Lancet Infect. Dis.* **11**(12):916–924 (2015). doi:10.1016/S1473-3099(11)70240-7.

[179] Offeddu V, Cowling BJ, Peiris JM. Interventions in live poultry markets for the control of avian influenza: A systematic review. *One Heal.* (2016). doi:10.1016/j.onehlt.2016.03.002.

[180] Hosseini PR, Fuller T, Harrigan R, Zhao D, Arriola CS, *et al.* Metapopulation Dynamics Enable Persistence of Influenza A, Including A/H5N1, in Poultry.

PLoS One **8**(12):e80091 (2013). doi:10.1371/journal.pone.0080091.

[181] Shea K, Tildesley MJ, Runge MC, Fonnesbeck CJ, Ferrari MJ. Adaptive Management and the Value of Information: Learning Via Intervention in Epidemiology. *PLoS Biol.* **12**(10):e1001970 (2014). doi:10.1371/journal.pbio.1001970.

[182] Reid AH, Fanning TG, Hultin JV, Taubenberger JK. Origin and evolution of the 1918 "Spanish" influenza virus hemagglutinin gene. *Proc. Natl. Acad. Sci.* **96**(4):1651–1656 (1999). doi:10.1073/pnas.96.4.1651.

[183] Taubenberger JK, Reid AH, Lourens RM, Wang R, Jin G, *et al.* Characterization of the 1918 influenza virus polymerase genes. *Nature* **437**(7060):889–893 (2005). doi:10.1038/nature04230.

[184] Imai M, Watanabe T, Hatta M, Das SC, Ozawa M, *et al.* Experimental adaptation of an influenza H5 HA confers respiratory droplet transmission to a reassortant H5 HA/H1N1 virus in ferrets. *Nature* **486**(7403):420–428 (2012). doi:10.1038/nature10831.

[185] Herfst S, Schrauwen EJA, Linster M, Chutinimitkul S, de Wit E, *et al.* Airborne transmission of influenza A/H5N1 virus between ferrets. *Science* **336**(6088):1534–1541 (2012). doi:10.1126/science.1213362.

[186] Buhnerkempe MG, Gostic K, Park M, Ahsan P, Belser JA, *et al.* Mapping influenza transmission in the ferret model to transmission in humans. *Elife* **4**:e07969 (2015). doi:10.7554/eLife.07969.

[187] Lipsitch M. Can Limited Scientific Value of Potential Pandemic Pathogen Experiments Justify the Risks? *MBio* **5**(5):e02008–14 (2014). doi:10.1128/mBio.02008-14.

[188] World Health Organisation. Influenza virus infections in humans (February 2014). (2014). URL `http://www.who.int/influenza/human_animal_interface/virology_laboratories_and_vaccines/influenza_virus_infections_humans_feb14.pdf`. [Online] (Accessed: 20 Mar 2017).

[189] Patterson KD. *Pandemic Influenza 1700-1900.* Rowman and Littlefield, New Jersey (1986).

[190] Beveridge W. The chronicle of influenza epidemics. *Hist. Philos. Life Sci.* **13**(2):223–234 (1991).

[191] Potter CW. A history of influenza. *J. Appl. Microbiol.* **91**(4):572–579 (2001). doi:10.1046/j.1365-2672.2001.01492.x.

[192] Taubenberger JK, Morens DM. Pandemic influenza including a risk assessment of H5N1. *Rev. Sci. Tech.* **28**(1):187–202 (2009).

[193] House T. Epidemiological dynamics of Ebola outbreaks. *Elife* **3**:e03908 (2014). doi:10.7554/eLife.03908.

[194] Schoenbaum SC, Coleman MT, Dowdle WR, Mostow SR. Epidemiology of influenza in the elderly: evidence of virus recycling. *Am. J. Epidemiol.* **103**(2):166–173 (1976).

[195] Dowdle WR. Influenza A virus recycling revisited. *Bull. World Health Organ.* **77**(10):820–828 (1999).

[196] Nakajima K, Desselberger U, Palese P. Recent human influenza A (H1N1) viruses are closely related genetically to strains isolated in 1950. *Nature* **274**(5669):334–339 (1978). doi:10.1038/274334a0.

[197] Kilbourne ED. Influenza Pandemics of the 20th Century. *Emerg. Infect. Dis. J.* **12**(1):9 (2006). doi:10.3201/eid1201.051254.

[198] Neuts MF. *Probability distributions of phase type*. Purdue University. Department of Statistics (1974).

[199] Asmussen S. *Applied Probability and Queues*. Stochastic Modelling and Applied Probability. Springer New York, New York, NY (2003). doi:10.1007/b97236.

[200] Bar-Hillel M, Wagenaar WA. The perception of randomness. *Adv. Appl. Math.* **12**(4):428–454 (1991). doi:10.1016/0196-8858(91)90029-I.

[201] Akaike H. A new look at the statistical model identification. *IEEE Trans. Automat. Contr.* **19**(6):716–723 (1974). doi:10.1109/TAC.1974.1100705.

[202] Hurvich CM, Tsai CL. Regression and Time Series Model Selection in Small Samples. *Biometrika* **76**(2):297–307 (1989). doi:10.2307/2336663.

[203] Burnham KP, Anderson DR, editors. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. Springer New York, New York, NY (2004). doi:10.1007/b97636.

[204] Kass RE, Raftery AE. Bayes Factors. *J. Am. Stat. Assoc.* **90**(430):773–795 (1995). doi:10.1080/01621459.1995.10476572.

[205] Centers for Disease Control and Prevention. Outbreaks Chronology: Ebola Virus Disease. (2016). URL `http://www.cdc.gov/vhf/ebola/outbreaks/history/chronology.html`. [Online] (Accessed: 23 Jan 2017).

[206] Bos KI, Schuenemann VJ, Golding GB, Burbano HA, Waglechner N, *et al.* A draft genome of Yersinia pestis from victims of the Black Death. *Nature* **478**(7370):506–510 (2011). doi:10.1038/nature10549.

[207] Reichert T, Chowell G, McCullers JA. The age distribution of mortality due to influenza: pandemic and peri-pandemic. *BMC Med.* **10**(1):162 (2012). doi:10.1186/1741-7015-10-162.

[208] Reichert T, Chowell G, Nishiura H, Christensen RA, McCullers JA. Does Glycosylation as a modifier of Original Antigenic Sin explain the case age

distribution and unusual toxicity in pandemic novel H1N1 influenza? *BMC Infect. Dis.* **10**(1):5 (2010). doi:10.1186/1471-2334-10-5.

[209] Lessler J, Cummings DA, Read JM, Wang S, Zhu H, *et al.* Location-specific patterns of exposure to recent pre-pandemic strains of influenza A in southern China. *Nat. Commun.* **2**:423 (2011). doi:10.1038/ncomms1432.

[210] Lessler J, Riley S, Read JM, Wang S, Zhu H, *et al.* Evidence for Antigenic Seniority in Influenza A (H3N2) Antibody Responses in Southern China. *PLoS Pathog.* **8**(7):e1002802 (2012). doi:10.1371/journal.ppat.1002802.

[211] Gerhard W, Mozdzanowska K, Zharikova D. Prospects for Universal Influenza Virus Vaccine. *Emerg. Infect. Dis.* **12**(4):569–574 (2006). doi:10.3201/eid1204.051020.

[212] Kanekiyo M, Wei CJ, Yassine HM, McTamney PM, Boyington JC, *et al.* Self-assembling influenza nanoparticle vaccines elicit broadly neutralizing H1N1 antibodies. *Nature* **499**(7456):102–106 (2013). doi:10.1038/nature12202.

[213] World Health Organisation. Depression Fact Sheet No.369. (2012). URL `http://www.who.int/mediacentre/factsheets/fs369/en/`. [Online] (Accessed: 27 Jan 2017).

[214] Rueger SY, Malecki CK, Demaray MK. Relationship Between Multiple Sources of Perceived Social Support and Psychological and Academic Adjustment in Early Adolescence: Comparisons Across Gender. *J. Youth Adolesc.* **39**(1):47–61 (2010). doi:10.1007/s10964-008-9368-6.

[215] Mead N, Lester H, Chew-Graham C, Gask L, Bower P. Effects of befriending on depressive symptoms and distress: systematic review and meta-analysis. *Br. J. Psychiatry* **196**(2):96–101 (2010). doi:10.1192/bjp.bp.109.064089.

[216] Kramer ADI, Guillory JE, Hancock JT. Experimental evidence of massive-scale emotional contagion through social networks. *Proc. Natl. Acad. Sci.* **111**(24):8788–8790 (2014). doi:10.1073/pnas.1320040111.

[217] Joiner TE, Katz J. Contagion of Depressive Symptoms and Mood: Meta-analytic Review and Explanations From Cognitive, Behavioral, and Interpersonal Viewpoints. *Clin. Psychol. Sci. Pract.* **6**(2):149–164 (1999). doi:10.1093/clipsy.6.2.149.

[218] Bentley RA, Ormerod P. A rapid method for assessing social versus independent interest in health issues: A case study of 'bird flu' and 'swine flu'. *Soc. Sci. Med.* **71**(3):482–485 (2010). doi:10.1016/j.socscimed.2010.03.042.

[219] Harris K, Halpern C, Whitsel E, Hussey J, Tabor J, *et al.* The National Longitudinal Study of Adolescent to Adult Health: Research Design. (2009). URL `http://www.cpc.unc.edu/projects/addhealth/design`. [Online] (Ac-

cessed: 23 Jan 2017).

[220] Radloff LS. The CES-D Scale: A Self-Report Depression Scale for Research in the General Population. *Appl. Psychol. Meas.* **1**(3):385–401 (1977). doi: 10.1177/014662167700100306.

[221] Roberts RE, Lewinsohn PM, Seeley JR. Screening for Adolescent Depression: A Comparison of Depression Scales. *J. Am. Acad. Child Adolesc. Psychiatry* **30**(1):58–66 (1991). doi:10.1097/00004583-199101000-00009.

[222] Centola D, Macy M. Complex Contagions and the Weakness of Long Ties. *Am. J. Sociol.* **113**(3):702–734 (2007). doi:10.1086/521848.

[223] Grimmett G, Stirzaker D. *Probability and Random Processes*. OUP, Oxford (2001).

[224] Bland JM, Altman DG. Multiple significance tests: the Bonferroni method. *BMJ* **310**(6973):170 (1995). doi:10.1136/bmj.310.6973.170.

[225] Hosmer DW, Lemeshow S. *Applied Logistic Regression*. Wiley series in probability and statistics. John Wiley \& Sons, Inc., New York, Chichester, Weinheim (2005).

[226] Brown LD, Cai TT, DasGupta A. Confidence Intervals for a binomial proportion and asymptotic expansions. *Ann. Stat.* **30**(1):160–201 (2002). doi: 10.1214/aos/1015362189.

[227] Schaefer DR, Kornienko O, Fox AM. Misery Does Not Love Company: Network Selection Mechanisms and Depression Homophily. *Am. Sociol. Rev.* **76**(5):764–785 (2011). doi:10.1177/0003122411420813.

[228] Neumann R, Strack F. "Mood contagion": The automatic transfer of mood between persons. *J. Pers. Soc. Psychol.* **79**(2):211–223 (2000). doi:10.1037/0022-3514.79.2.211.

[229] Lakin JL, Chartrand TL. Using Nonconscious Behavioral Mimicry to Create Affiliation and Rapport. *Psychol. Sci.* **14**(4):334–339 (2003). doi:10.1111/1467-9280.14481.

[230] Leighton J, Bird G, Orsini C, Heyes C. Social attitudes modulate automatic imitation. *J. Exp. Soc. Psychol.* **46**(6):905–910 (2010). doi:10.1016/j.jesp.2010.07.001.

[231] Höhn P, Menne-Lothmann C, Peeters F, Nicolson NA, Jacobs N, *et al.* Moment-to-Moment Transfer of Positive Emotions in Daily Life Predicts Future Course of Depression in Both General Population and Patient Samples. *PLoS One* **8**(9):e75655 (2013). doi:10.1371/journal.pone.0075655.