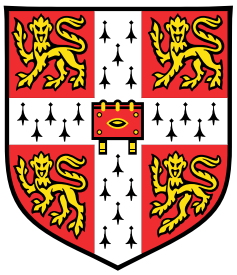
Modelling the Epidemiological Dynamics of Seasonal Influenza Viruses at Local Scales



Edward Kong Seng Lam

Supervisor: Prof. Colin Russell.
Supervisor
Dr Olivier Restif.
Supervisor

Department of Veterinary Medicine
University of Cambridge

This dissertation is submitted for the degree of
Doctor of Philosophy

Trinity College

September 2020

Declaration

This thesis is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the preface and specified in the text. It is not substantially the same as any work that has already been submitted before for any degree or other qualification except as declared in the preface and specified in the text. It does not exceed the prescribed word limit for the Department of Veterinary Medicine Degree Committee.

Edward Kong Seng Lam
September 2020

Modelling the Epidemiological Dynamics of Seasonal Influenza Viruses at Local Scales

Edward Kong Seng Lam

Abstract

Seasonal influenza viruses are a substantial source of disease burden globally, causing epidemics across all climatic regions. Through error-prone RNA replication, influenza viruses can escape pre-existing humoral immunity and reinfect humans, resulting in recurrent epidemics within populations. From year to year, individual epidemics differ substantially in timing, duration and size. Despite intensive study, characterising the spatiotemporal patterns of virus circulation and identifying the underlying sources of this variability at global, regional and local scales remain as ongoing challenges. There is a need to reconcile environmental, virological and host drivers of virus epidemiological dynamics across diverse contexts. Such insights can only be generated through a holistic approach that integrates observational, ecological, experimental and modelling studies: this would enable more accurate and timely epidemiological forecasts and more efficient allocation of public health resources.

In this thesis, I investigate the phylodynamical interactions between the seasonal influenza virus, environment and human host population, integrating analyses from observational study and theoretical modelling approaches. The current knowledge gap on the drivers of local city-level epidemics is identified in Chapter 2 and subsequently addressed over 4 research chapters. In Chapter 3, I review existing epidemic detection algorithms and present a novel statistical model that I developed for use with noisy disease surveillance data and is optimised for the context of seasonal influenza. In Chapter 4, I apply this novel algorithm and analyse a 15-year dataset of 18,250 typed, subtyped, and antigenically characterised seasonal influenza viruses from the five most populous cities in Australia. With the necessary geographical and virus resolution, I quantify the effects of previously hypothesised environmental and virological factors. Most surprisingly, despite an apparent lack of marked change in virus antigenicity, individual antigenic variants are capable of reinvading the same population over consecutive seasons, which runs contrary to predictions made by existing mathematical models.

In Chapters 5 and 6, I investigate how antigenic variants are capable of causing recurrent epidemics at local scales by building upon previous theoretical modelling studies and developing a modelling framework to investigate the interactions between and joint effects exerted by the topology of cross-immunity and host contact structure within a population. In

Chapter 5, I investigate the effects of correlations between network structure and individual susceptibility. In Chapter 6, I examine the population-level significance of age-specific changes to an individual's immune response. In Chapter 7, I review my findings and discuss how these new insights into virus ecology can open new avenues for better influenza control and future research.

Acknowledgements

I would like to thank my supervisor Colin Russell for his counsel, insightful comments and encouragement, which have enabled me to complete this body of work. I am also very grateful to Olivier Restif for facilitating the continuity of supervision and support within the Disease Dynamics Unit at the Department of Veterinary Medicine here in Cambridge.

I would also like to thank Aeron Hurt previously at the University of Melbourne and Ian Barr at Federation University for their dedication towards accurate influenza virus surveillance over past decades and graciously providing access to their carefully curated data set. Dylan Morris at Princeton University also deserves my gratitude for his technical assistance and stimulating conversations. Additionally, my thanks to my fellow inhabitants of Office 19: particularly, the other members of the original trio, Edyth Parker and Laura Cooper.

I would like to acknowledge the support provided by the MBPhD Programme and the Clinical School, University of Cambridge in enabling early career clinician scientists like myself to pursue their passions.

Last but not least, I would like to thank my parents and brother wholeheartedly for their unconditonal love and faith. Hope is a commodity that is in short supply these days and I wish that this treatise will bear testament to the importance of conviction and that we can all make incremental contributions to mankind, no matter how insignificant.

Table of contents

List of figures

List of tables

Chapter 1

Aims and Rationale

Seasonal influenza viruses are a substantial source of disease burden and result in 650,000 deaths per year[269]. There are currently four subtypes/lineages of influenza viruses within the human population, which co-circulate across tropical, subtropical and temperate climatic regions. Through error-prone RNA replication, influenza viruses are capable of escaping pre-existing humoral immunity[58] and reinfecting humans, resulting in recurrent epidemics. From year to year, individual epidemics differ substantially in timing, duration and size[63,79]. Despite intensive study over past decades, characterising the spatiotemporal patterns of virus circulation and identifying the underlying sources of this variability at global, regional and local scales remain as ongoing challenges. Insights into these processes would enable more accurate and timely epidemiological forecasts[86] and more efficient allocation of public health resources[203].

To date, our understanding of climatic drivers of influenza virus epidemics has been primarily informed by the distinct patterns observed in temperate regions of the Northern and Southern Hemispheres[239]. In these climes, influenza virus activity is most common in winter, suggesting that cool and dry conditions are conducive to or potentially sufficient for driving wintertime epidemics. In particular, fluctuations in absolute humidity have been hypothesised to trigger epidemic activity[222]. Experimental studies have demonstrated that reductions in temperature and absolute humidity enhance viral stability and aerosol transmission[161,162,221], forming a putative link between observed seasonality and underlying climatic mechanisms. However, more recent epidemiological studies in tropical and subtropical regions have shown that increased virus activity often coincides with the rainy season[224,240], during periods of high temperature and humidity. This apparent quandary illustrates the need to move beyond the simplistic dichotomy of temperate vs subtropical/tropical regions, by further characterising the spectrum of virus seasonality and roles of climatic drivers[239].

The accumulation of genetic mutations in the globular head of haemagglutinin protein periodically results in the emergence of virus variants that partially escape antibodies raised during prior influenza virus infections[231]. At a global level, this process of antigenic drift and antibody-mediated selection eventually result in the domination of a single strain and a "cactus-like" structure of the phylogenetic tree[25,278], which is particularly prominent for A/H3N2 viruses. Of course, evolution does not occur in isolation and modelling studies have theorised how such global fitness advantages would manifest themselves at local scales and affect the dynamics of individual epidemics. By partially escaping pre-existing immunity induced by prior infections and vaccinations, epidemiological theory predicts that new major antigenic variants should cause larger epidemics than previously circulated variants[130,131,275].

Evidently, the phylodynamics of influenza viruses involves complex, mutual and reciprocal interactions between virus, host and the environment[104,260]. Experimental and theoretical studies may uncover potential mechanistic drivers for influenza dynamics but quantifying the effects of these environmental and virological drivers on virus transmission requires surveillance data that has sufficient geographical resolution and virus characterisation, in order to carefully dissect apart the overlapping local-level, subtype- and antigenic variant-specific patterns.

Without accurate virus- and antigenic variant-specific time series data, the development of hypotheses and mathematical models is hampered: simply put, what are the phenomena of interest that model frameworks are attempting to recapitulate? In order to strike a fair balance between capturing (biological) realism and mathematical/computational tractability[141,199], it is of paramount importance to carefully select the most important underlying factors to model and how best to represent them mechanistically, both of which are challenges in their own rights. Capturing the dynamics of multiple competing strains has been a prime focus[141] but the nature of cross-immunity between strains and host immune history remains poorly elucidated. Similarly, models incorporating population structure have been developed, yet these are commonly plagued by a lack of empirically driven and robust parameterisation[267].

In this thesis, I investigate the phylodynamical interactions between seasonal influenza viruses, climate, and human host populations, integrating analyses from observational study and theoretical modelling approaches. In Chapter 2, I first review the current literature and identify a knowledge gap surrounding the drivers of local city-level epidemics. This gap is then addressed over 4 research chapters. In Chapter 3, I review existing epidemic detection algorithms and present a novel statistical model that I developed for use with noisy disease surveillance data, which is optimised for the context of seasonal influenza. In Chapter 4, I apply this novel framework and analyse a 15-year dataset of 18,250 typed, subtyped, and antigenically characterised seasonal influenza viruses from the five most populous

cities in Australia. With the necessary geographical and virus resolution, I critically assess previously hypothesised environmental and virological factors and quantify their effects on epidemiological dynamics. Most surprisingly, individual antigenic variants are capable of reinvading the same population over consecutive seasons, without necessitating marked phenotypic changes, which runs contrary to predictions made by existing mathematical models.

Host population structure and partial cross-immunity shape epidemiological dynamics but have often been evaluated by previous modelling studies in isolation. In Chapter 5 and Chapter 6, I build upon these previous theoretical studies and develop a modelling framework to investigate the how these factors can interact and through their joint effects, underpin the empirical observation that antigenic variants are capable of causing recurrent epidemics. Additionally, I instil more biological detail and realism into the mechanistic implementation of these factors, demonstrating the importance of including oft-overlooked nuances. In Chapter 5, I investigate how epidemiological dynamics are affected by the underlying contact and immune structure within the host population. I expand upon this framework in Chapter 6 and examine whether previously hypothesised immune mechanisms could have further impacts on the likelihood of reinvasion. In Chapter 7, I review my findings and discuss how these new insights into virus phylodynamics can open new avenues for better influenza control and future research.

Chapter 2

Introduction

In this chapter, I review the current literature on the phylodynamics of seasonal influenza viruses, in order to identify knowledge gaps and limitations of current modelling approaches, both of which I address over subsequent research chapters. Firstly, I examine how the interactions between virus antigenicity and host immunity, alongside the effects of climatic factors, have been hypothesised to drive epidemiological and evolutionary dynamics, across scales. This evaluation identified a key knowledge gap surrounding the drivers for the dynamics of local city-level epidemics, which I quantitatively assess in Chapter 4, using highly resolved surveillance data from Australia.

Secondly, I evaluate existing mathematical models that have attempt to characterise and quantify these non-linear and reciprocal interactions. In particular, I assess the challenges posed by model complexity, tractability and parameterisation: such modelling considerations, alongside the identification of key phylodynamical factors from the literature, guide the methodological development of my network models in Chapters 5 and 6.

2.1 Influenza virus biology

Influenza viruses are negative-sense, single-stranded RNA viruses of the Orthomyxoviridae family. Whilst there are three types of influenza viruses, the burden of seasonal virus disease can overwhelmingly be attributed to just influenza types A and B; type C viruses have historically been considered to be a minor respiratory pathogen, responsible for less severe paediatric disease[9,218]. The current seasonal influenza A/H3N2 (A/H3) and A/H1N1 (currently A/H1pdm09; previously A/H1sea) subtypes are zoonotic in origin, having adapted to continued circulation within the human population, after initially causing a pandemic. In contrast, influenza B viruses do not appear to have animal reservoirs and currently con-

sist of two co-circulating, antigenically distinct lineages: B/Victoria/2/87-like (B/Vic), and B/Yamagata/16/88-like (B/Yam) viruses.

Influenza A and B viruses have similar genomes that consist of 8 RNA segments, which encode functionally homologous proteins that are essential to its life cycle. Structurally, the virus comprises an envelope, a layer of virus matrix protein and the encapsulated ribonucleoprotein (RNP) complex. Within the lipid membrane of the envelope are the primary surface glycoproteins, haemagglutinin (HA) and neuraminidase (NA), which play key roles in virus ingress and egress. HA binds onto sialic acids on the glycan receptors of epithelial cells in the upper respiratory tract of humans and mediates subsequent membrane fusion. However, prior to cell entry, invading virions must first transverse the viscous mucus that is rich in heavily sialylated mucins: the sialidase activity of NA is critical to overcoming this physical barrier by cleaving these decoy receptors[45]. At the end of viral replication, sialidase activity of NA enables the release of nascent virions from the surface of the host cell by cleaving the very same receptors used for virus entry[45]. Owing to their location embedded on the surface of the viral envelope, HA and NA are subject to neutralising antibodies; HA[257], in particular its globular HA1 head domain[129,265,266], is the primary target of this humoral response.

Due to a lack of proofreading activity, the viral RNA-dependent RNA polymerase (RdRP) complex frequently incorporates nucleotide mismatches during genome replication; it has been shown that the RdRP of type B viruses is less error-prone than type A viruses[195]. Over time, the accumulation of nucleotide mutations and amino acid substitutions within the HA and NA glycoproteins can produce mutant viruses that are capable of escaping the humoral immunity conferred by prior infection and vaccination. These new mutants can eventually come to dominance, replacing the previous antigenic variant. Virus antigenicity can be quantified using the haemagglutination inhibition (HI) assay or neuraminidase inhibition (NAI) assay and its punctuated evolution over time can be visualised using antigenic cartography[231].

This process of antigenic drift, particularly in HA, profoundly impacts vaccine effectiveness necessitating the need to periodically update the strain composition of the seasonal vaccine. In order to allow sufficient time for large-scale vaccine production, hemisphere-specific strain selection occurs nine months prior to the upcoming winter[212]. Inevitably, this task is fraught with uncertainty but the prospect that virus evolution could be somewhat predictable provides glimmers of hope. Seminal studies in the 1980s identified five antigenic sites in HA[228,265]: viruses responsible for large epidemics in the 1970s, when major antigenic change was suspected, had at least one amino acid substitution in each of the five broadly defined sites. Following on from the original HA antigenic maps made by Smith et al.[231], systematic phenotypic chararcterisation of all of the amino acid substitutions

between representative viruses of each antigenic cluster showed that in most instances, single amino acid changes in one of seven positions around the HA receptor binding site were sufficient to recapitulate changes in antigenicity associated with historically observed cluster transitions[129]. Subsequently, it has been shown that the magnitude of immune escape from specific neutralising antibodies was dependent on both the specific amino acid position and substitution[66].

The net effect of immune selection pressures on antigenic evolution, i.e. the continual extinction of old and emergence of new variants, is the culmination of countless instances of within-host selection, each of which are uniquely shaped by an individual's infection and vaccination history. This selection pressure is exerted by components of the innate and adaptive immune system. Naïve individuals, by definition, have no prior exposure to influenza virus antigens and rely on innate mechanisms: the physical mucosal barrier, activation of the antiviral state by infected cells and recruitment of innate effector cells. These non-specific mechanisms act independent of antigenic phenotype, and play important roles in limiting the size of the initial virus inoculum and early containment of virus replication. Owing to the acute timescale of infection, viruses are able to elude the adaptive immunity, completing subsequent replication within host cells before the host is able to mount a *de novo* humoral response. Viral titres and shedding typically peak at around 24-72 hours[147], whilst the generation of specific antibodies in response to primary infection takes 7-10 days, due to the slower process of recruiting and activation of naïve B cells via antigen presentation and clonal selection.

Primary infection primes the adaptive immune system. The mucosal barrier is augmented by secretory immunoglobulin A (sIgA), which is highly specific towards previously encountered virus strains and can neutralise or slow down virus invasion. Secondary infection and viral replication within epithelial cells triggers a memory B cell recall response, which enables the rapid production of antigen-specific antibodies within 3-5 days. Even with a more timely humoral response, the peak of viral replication still precedes the onset of the recall response during secondary infection. It is thus, at the point of inoculation, in the presence of these cross-reactive sIgAs, that virions are subjected to strong selection pressures that favour variants with large antigenic effects[185,200]. This strong initial bottleneck and weak selection pressures during replication limit within-host selection and may serve to limit the global rate at which new antigenic variants emerge. However, immunocompromised hosts are unable to clear the virus so infections can last for weeks or months[176]: prolonged virus replication can now be affected by immune selection so these hosts may act as a hot bed for generating antigenic diversity and contribute to global virus evolution[185,272].

The pattern of an individual's serological response to virus exposure varies with their age and is shaped by their immune history, with most responses containing a degree of immunological backboosting. It has long been observed that titre responses are often strongest towards antigenic variants encountered in the first decade of life[57,58,59,60]: the first strain may hold a privileged position in the immune response, potentially resulting in the phenomenon of the "original antigenic sin"[83] and the detrimental suppression of novel and more specific responses. More recent longitudinal serological studies have found that the relative strength of backboosting may in fact be determined by a hierarchy of antigenic seniority, where early strains take "senior" positions and subsequently encountered strains take on progressively more "junior" ones[81,140,151]. Childhood exposure may therefore render older individuals perpetually susceptible towards current and future viruses by detrimentally biasing their immune response towards the production of non-neutralising antibodies[100].

Competition for antigens is a likely mechanistic explanation antigenic seniority, where common epitopes shared between current and historic viruses preferentially activates the memory response. Predominantly, these immmunodominant epitopes are located within the globular head of HA[129,265,266] but over a lifetime's worth of exposure, some older individuals may develop broadly neutralising antibodies that target the poorly accessible and subdominant epitopes on the HA stalk domain[7,71]. Due to their functional importance in mediating membrane fusion, these stalk epitopes are highly conserved between antigenic variants and even subtypes[134,237], suggestive of low mutational tolerance. Unlike the variable head epitopes targeted by the current seasonal vaccines, the prospect of lower plasticity in the stalk domain have made them prime candidates for the development of universal vaccines, which would not require periodic reformulation and could potentially offer protection against novel pandemic subtypes[3,134,136].

However, the development of universal vaccines has proven to be considerably difficult. Unlike murine and ferret models, where stalk-reactive memory B cells can be readily enriched through sequential infection or vaccination[135,189,235], the natural frequency of these cells and broadly neutralising antibodies are relatively low in the human population[7,71]. Even if vaccination is able to elicit anti-stalk immunity, these stalk epitopes are likely to be more plastic than previously assumed and they may exhibit antigenic drift once subjected to immune selection. Indeed, positive selection has been observed in the HA stalk domain of A/H1pdm09 viruses[3,236]. Escape mutants have also been generated experimentally, during *in vitro* culturing of A/H1pdm09 viruses in the presence of stalk-specific monoclonal antibodies[37,244]. Crucially, some of escape mutations did not appear to be associated with high fitness costs; mutants retained their replicative fitness and virulence in mice[3,271].

Any discussion of influenza immunity would not be complete without acknowledging the limitations in identifying correlates or surrogates of protection, which can meaningfully associate immune responses with protection from disease in individuals. An individual's mucosal sIgA response might provide an accurate representation for their protection but sample collection via nasal wash would often be impractical[169] so most studies utilise serum immunoglobulin G (IgG). However, IgG does not readily translocate across the lung epithelium and is likely a poor proxy for immune activity at the location of infection[200]. The ability to quantify immune protection is further hampered by the use of HI assays, which measure the ability of sera to inhibit the virus-induced agglutination of red blood cells. These assays are unable to characterise antibodies that recognise epitopes distant from the receptor binding site. Other approaches, such as microneutralisation (MN) assays, may be more sensitive than HA assays and better detect antibodies that neutralise *in vitro* virus growth. These other antibodies may act through cellular cytotoxicity, phagocytosis and complement activation, which may have important *in vivo* roles to control infection[145,156].

Many studies show that serum HI titres are correlated with protection: individuals with higher titres experience a reduced likelihood of subsequent infection[69,82,201]. However, it is difficult to accurately determine the relationship between HI titre and protection. Whilst a titre greater than 1:40 is widely considered a protective response[115], the absolute value for the cutoff threshold can vary substantially between adults and children[26] and depend on virus subtype/type[49,169] and strain[69,82,201] tested against. MN titres are even more difficult to interpret as correlates of protection, since such titre-protection relationships have yet to be formally established[70,169].

As part of vaccine strain selection, it is important to characterise the immune selection pressures exerted at the population level by accounting for the heterogeneity in strain-specific immunity to influenza viruses between individuals. Indeed, a recent study found substantial individual-to-individual variation in the antigenicity of amino acid escape mutations for influenza HA[149], which likely result from their complex exposure histories to different influenza viruses. This important heterogeneity cannot be recapitulated in HI assays, where the reference anti-sera are produced by infecting naïve ferrets: primary infection elicits a relatively homogenous response towards a limited number of immunodominant epitopes on the HA globular head. By treating population level immunity as a monolithic response, our understanding of virus antigenic evolution may be biased towards mutations that cause large antigenic changes[129,231], overlooking important potential evolutionary trajectories that involve subtle changes in antigenicity[111,156].

The focus on the antigenic evolution of HA has overshadowed the contributions of the non-antigenic internal gene segments towards the overall adaptive evolution of seasonal

influenza viruses[204]. Consistent with strong humoral selection, internal segments tend to have lower rates of nonsynonymous substitutions than HA[205]. However, the evolution of internal segments could be compensatory and necessary to offset any potential loss in receptor binding affinity arising from immune escape[105]. Indeed, the A/H3N2/California/7/2004 antigenic variant possessed substitutions in the polymerase acidic (PA) segment, which may have accounted for its greater replicative fitness and virulence over its predecessor, A/H3N2/Fujian/411/2002[175]. *In vitro* and *in vivo* experiments have shown that the observed substitutions in the nucleoprotein (NP) gene of A/H3N2 viruses may have had similar impacts in promoting viral RNA replication[120]. These results illustrate the need for whole genome sequencing, in order to develop a more comprehensive understanding of existing phenotypic variation, fitness constraints and trade-offs between different facets of virus fitness.

Whilst the components of the antibody-mediated protection require further elucidation, it is evident that the mechanistic representation of cross-immunity within mathematical models needs further development, in order to better account for the varied exposure histories of hosts and associate it to the diverse spectrum of *in vivo* immune response between individuals.

2.2 Global evolutionary dynamics

Given the acute nature of infection and distinct wintertime seasonality of epidemics in temperate regions, virus transmission must somehow be maintained between seasons, whether it be through low-level persistence within local populations[117] or externally with reintroductions every winter[2,256]. If the virus population is maintained externally, there is a need to further delineate between a strict source-sink model, where virus activity occurs year-round in individual tropical countries[80,226,256], or as a wave of epidemics that moves sequentially through susceptible populations[212]. Elucidating between these putative explanations is critical, since this not only shapes the global evolutionary dynamics by linking the within- and between- host processes that give rise to new antigenic variants, but also has profound implications for optimal disease control strategies at local scales[43,212].

Early phylogenetic analyses demonstrated regular, cross-hemisphere viral migration between seasons and a lack of local persistence in temperate countries between seasons[190]. With improved virus surveillance coverage and the development of more sophisticated phylogenetic models, it has become possible to infer, with increasing geographical resolution, the patterns in migratory dynamics from the virus genealogy. Subsequent studies identified East and Southeast Asia as an important source of viruses, where a circulation network sustains A/H3 viruses, which moves through a series of temporally overlapping epidemics across the region[22,205,212]. Indeed, the importance of E-SE Asia is unsurprising, given that

this highly urbanised region is home to approximately a third of the world's population. Additionally, these populations are highly connected and mobile: internally within their countries, regionally and globally. More recently, India has been identified as another member of this A/H3 virus circulation network[25], demonstrating the need to address historical and ongoing surveillance biases in particular regions of the world.

In addition to geographical coverage, virus surveillance and sequencing of A/H1 and influenza B viruses have improved substantially. Unlike A/H3 viruses, A/H1 and type B viruses have on occasion, persisted locally between epidemics in regions outside of E-SE Asia and India, producing co-circulating genetic lineages and even antigenic variants[25]. Again, these results highlight our understanding of global evolutionary dynamics may be limited by our incomplete and patchy sampling of global virus diversity. These subtype- and type-specific differences in migratory rates are likely to stem from differences in the rates of virus evolution and age distributions of infected individuals[25]. With the highest rate of antigenic drift, A/H3 viruses rapidly evade population immunity, causing frequent epidemics and infecting individuals across all age groups. Conversely, A/H1sea and influenza B viruses evolve at a slower rate, causing less frequent epidemics that predominantly affect children[82,95,150]. By infecting more adults, who account for the majority of air travellers, A/H3 viruses have more opportunities to move internationally and initiate new chains of transmission, whilst A/H1sea and influenza B viruses have longer durations of regional persistence[25].

A comprehensive understanding of global influenza virus evolution and migratory patterns requires the elucidation of the processes underlying the emergence of new antigenic variants. At local scales, circulating viruses are subjected to selection pressures from host immunity within discrete populations. It is therefore crucial that the dynamics of city-level epidemics are further characterised.

2.3 Local influenza virus epidemiology

Seasonal influenza viruses can only be imported from the global circulation network if local conditions are conducive to epidemic initiation and sustained transmission; otherwise, individual seeding events would by and large, result in dead-end transmission chains. There is thus great impetus to further elucidate the epidemiological dynamics of influenza virus activity and its drivers at local scales. The factors that influence and contribute to the variability in the timing and magnitude of local epidemics are of particular interest, especially in the context of disease control.

Historically, hypotheses surrounding influenza seasonality have primarily been informed by the distinct wintertime seasonality experienced by countries in temperate regions. Whilst the exact timing and magnitude of local epidemics differ from year to year, virus activity increases rapidly and entire influenza virus epidemics last only approximately three months[33,160,255], with an absence of sustained activity over summer months. Three classes of mechanisms have been proposed as a basis of this wintertime seasonality[158,160]: fluctuations in virus survival, fluctuations in host immunity, and changes in host behaviour.

In temperate countries, it has been long-speculated that wintertime epidemic activity might be causally linked to seasonal changes in climatic factors[61,109]. Experimental studies using animal models have shown that low temperature and relative humidity (*RH*) conditions enhance aerosol virus survival and transmission[161,162,221], as well as persistence on surfaces[133,170]. Since the relative contributions of each route of transmission are unknown[142,242], the sensitivity of overall *in vivo* transmission between humans towards climatic conditions remains poorly elucidated[163]. Furthermore, attempts at establishing and quantifying causal links between climatic factors and disease are confounded by the fact that the oft-used atmospheric conditions can differ quite substantially from indoor conditions[107], where interactions between people are most likely to occur.

A small number of experimental studies have suggested that the relationship between *RH* and virus survival may in fact be bimodal, with moderate survival possible at low and high *RH* values[216,223]. A bimodal relationship, rather than a negative monotonic one, could potentially explain the disparate observation that epidemics in the tropics and subtropics often occur during the humid and hot rainy season[224,240]. Conflicting findings surrounding virus stability at high *RH* values may stem from differences in the salt and protein composition of the media solution used to create virus droplets[133,180,273].

Supplementing experimental approaches, population-level observation studies have generated additional evidence for the importance of climatic factors as drivers of influenza seasonality: a seminal study of state-level epidemiological data from the United States found that epidemics were sometimes preceded by periods of anomalously low absolute humidity[222]. Subsequent epidemiological studies have found similar patterns using prefecture-level data from Japan[213], city-level data from the New York Metropolitan Area[62] and region-level data from France[219]. In addition to providing a crucial link between enhanced virus survival, transmission between individuals and overall epidemic activity, these studies have identified an environmental signal with predictive power that could be potentially leveraged for short-term forecasting.

Besides affecting virus viability, seasonal variations in meterological factors could influence host susceptibility towards infection, directly at the critical immune milieu of the lungs.

Dry atmospheric conditions can cause the all-important mucosal barrier to lose moisture and slow down mucociliary clearance[12,214]. The recruitment and activity of downstream immune effector cells are also less efficient at low temperatures, due to vasoconstriction and reduced blood supply[68]. Whilst low temperature did not diminish the innate immune response and activation of the antiviral state in guinea pigs[162], the duration of peak viral shedding was substantially prolonged.

Host behaviour and temporal variations in contact rates could also drive disease seasonality. Due to inclement weather during winter and the rainy season for temperate and tropical regions respectively, individuals spend more time in close proximity indoors[101,160]. Curiously, in the arid Southern States of the United States, epidemics still exhibit wintertime seasonality, despite people spending more time indoors in air-conditioned environments to avoid the scorching summer sun[239]. Given the burden of disease bore by children and their importance in driving community transmission[183], it is plausible that school openings may have triggered the fall wave of pandemic A/H1N1[38] and conversely, reductions in transmission during seasonal epidemics may be associated with school holidays[36]. However, such variations in contact rates are unlikely to be the sole or main drivers of epidemic activity: epidemics do not occur during school term time outside winter, nor are associated with events involving large gatherings of people.

As with our understanding of global dynamics, improvements in surveillance coverage of the subtropics and tropics have uncovered a diverse spectrum of influenza activity: disease patterns and strength of seasonality are highly variable between different countries[64,114,240]. Some locations, such as Singapore and Hong Kong, exhibit semi-annual peaks[41,42] whilst others display irregular year-round activity that lack well-defined seasons[2,144,187]. Hypotheses based on factors that potentially drive disease seasonality in temperate regions appear to be highly context dependent and cannot parsimoniously explain the patterns observed in the tropics[158]. It remains to be seen whether the varied patterns of disease across such diverse climatic backgrounds can truly be reconciled by a set of shared epidemiological drivers or not, if the primary drivers differ between local contexts[240]. The effect and strength of identical seasonal drivers could differ, modulated by the underlying spatiotemporal organisation in population density[54,55]: densely populated cities could potentially incubate limited chains of transmission during the summer, partially depleting the pool of susceptibles and reducing the size and intensity of wintertime epidemics.

The magnitude of epidemics is another key epidemiological quantity of interest. The final size of an epidemic is intuitively positively associated with how transmissible a pathogen is, although the exact form of this non-linear relationship mathematically depends on the complexity of the compartmental models considered[4,112,166,167,211]. For seasonal viruses,

where the potential for further intra-host adaptation is limited, it has been theorised that virus antigenicity and availability of susceptible individuals are key determinants of transmissibility. Different strains of the same antigenic variant are almost fully cross-reactive[94]. Due to imperfect cross-immunity between successive antigenic variants[58], newly arisen variants are able to partially escape immunity induced by prior infections and vaccinations. With a higher fraction of the population susceptible to infection, epidemiological theory predicts that new antigenic variants should cause larger epidemics than previously-circulated ones, resulting in stereotyped boom-and-bust cycles[130,131,275]. In support of this hypothesis is the observation that heightened mortality rates from pneumonia and influenza coincide with cluster transitions[102,130,265].

Antigenic change could also influence the exact timing of epidemics during the influenza season. With a greater proportion of the population susceptible, each individual introduction of a virus into a population is more likely to initiate a transmission chain that is less prone to stochastic termination. This means that new antigenic variants could potentially cause earlier local epidemics, as was the case in Israel[29]. Within a country, successful transmission events between different cities could become more frequent, resulting in greater spatio-temporally synchrony across epidemics, which was observed in United States[39,102,130], Japan[213] and Australia[92].

In addition to homotypic competition between antigenic variants, interactions between viruses of different types and subtypes have long been hypothesised[76], whereby prior infection by a virus of one subtype could reduce the likelihood of subsequent infection by viruses of complement subtypes[52,73,232]. However, the time scale and mechanistic basis for this partial competitive exclusion remain poorly established. Some studies have suggested that T cell immunity[171] or non-specific infection-induced interferon responses[56] may underpin short-term heterotypic competition, which would be limited to within the duration of a single influenza season. Indeed, this would be consistent with evidence from observational studies, which have found that the subtype or type that initiates above baseline levels of activity first is most likely to have the largest epidemic of that season[99,277]. Curiously, the ability to interfere with competitors appeared to vary between subtypes[99,275].

By its very nature, interrogating the phylodynamics of influenza is difficult and previous attempts at quantifying the effects of environmental and virological drivers of influenza epidemiology have been limited by three factors: 1. the reliance on influenza-like illness (ILI) data; 2. the aggregation of ILI or virologically confirmed data over large geographical scales (state/province/country); 3. where virologically confirmed data are available, the use of data without subtype and antigenic variant-level resolution.

ILI is diagnosed clinically, based on a set of symptoms and frequently includes a wide variety of non-influenza respiratory infections[217]. Occasionally, there is further laboratory characterisation but this is insufficient to delineate clearly between influenza virus type/subtype- and antigenic variant-specific patterns. Regardless, any potential patterns of interest are lost since data collection by disease surveillance networks often aggregate at ecological scales (state/province/country) that sum over multiple local (county/city/town-level) epidemics), which can individually vary substantially in timing, magnitude and influenza virus composition. All together, these sources of obfuscation make it difficult to disentangle local-level, antigenic variant-specific patterns and critically investigate the impact of putative drivers of epidemiological dynamics.

There is a clear need to robustly assess and quantify the impact of putative drivers on the dynamics of local city-level epidemics, which has not been feasible due to a lack of sufficient virus-specific and geographical resolution. Subsequently, the development of mathematical models has been hampered; I address this knowledge gap in Chapter 4, using highly resolved surveillance data from Australia.

2.4 Epidemiological modelling for influenza viruses

Epidemic processes are inherently difficult to study using conventional statistical analyses, due to their sensitivity to underlying noise from demographic and environmental processes[267]. Mechanistic models can partially account for and mitigate such challenges; despite the simplicity of these models, analytical insights into transmission dynamics and stationary states have proven invaluable in informing disease control[211]. Over the past three decades, mechanistic approaches have been further refined, leveraging on their ability to intuitively represent a system of diverse processes and capture the nonlinear effects of epidemiological drivers. This flexibility in being able to formulate models to investigate specific questions of public health interest can however lead to issues of mathematical and computational tractability: robust parametrisation can be hindered by a lack of suitable empirical data and inferential methods[15,199].

For seasonal influenza, model development has focused on creating frameworks to characterise the homosubtypic interactions between strains, in order to investigate how population disease dynamics are shaped by individual infection histories[141]. One approach to account for individual differences in immunity is to divide the host population into a small number of compartments, members of which all are identical in their status and susceptibility towards the disease. In the classical *SIR* model[128], individuals are split into three compartments, Susceptible, Infectious and Recovered, with a system of differential

equations to define the rate at which individuals transition between compartments. Whilst this framework could be an approximation of dynamics within a single season, the assumption that individuals are either fully susceptible or fully immune means that the continued evolution of the virus and the recurrent nature of epidemics cannot be recapitulated.

SIRS models attempt to better capture the process of antigenic drift, which is captured phenomenologically by the addition of a waning immunity rate, where individuals can return from the Recovered to the Susceptible compartment[198]. This however assumes that the loss of immunity by an individual is predicated on the time since their last infection. Overall, this produces a stereotyped pattern of gradual and continuous antigenic drift, rather than the empirically observed punctuated evolution, where the rapid emergence of a new antigenic variant is accompanied by the abrupt disruption of herd immunity[44,131].

By explicitly incorporating distinct viruses, multiple-strain models can be used to in investigate the interactions between different "strains". Here, the modelling term "strain" is used to categorise viruses into groups: different strains in the biological sense based on their genetic identities or different antigenic variants based on their broad phenotypical characteristics. The scale at which competition occurs at is often chosen to maintain analytical tractability: some models opt for a linear[98] or a relatively limited dimensionality[106] strain-space. Another approach is to make the tacit assumption that within-antigenic cluster evolution is neutral[251] and only consider the emergence of new major antigenic variants[130,131].

The strength of interactions between different "strains" is dependent on their cross-immunity, reflecting their relatedness in either the genetic or antigenic domains. In what amounts to a serial *SIR* framework, multiple-strain models record the levels of susceptibility and immunity towards each "strain". However, capturing the effects of past exposure is complicated by the fact that for influenza, the immune response is comprised of distinct contributions derived from prior infection events, where each encounter may or may not confer partial cross-immunity.

Multiple-strain models thus have two components to track: infection history (the set of antigenic variants exposed to) and immune status (the degree of cross-immunity conferred given exposure to a particular strain). To avoid issues with intractability, compartmental models often fall into a dichotomy, omitting either of the two components[5,35,97,98,131,207]. History-based models[5,35] explicitly track the number of individuals with each possible combination of variants exposed to. It is assumed that all hosts with a given infection history $\{i_1, i_2, \ldots, i_n\}$ will possess identical levels of partial cross-immunity against variant j[137]. On the other hand, status-based models record the levels of polarised immunity at the population level towards each variant[97,98,131,207]. In other words, upon infection by antigenic variant i, the host is immediately rendered immunised or not towards variant j. By avoiding explicit

tracking of the combinations of past infections, the number of state variables for status-based models increase linearly and not exponentially with the number of variants.

Whilst the convenience of numerical analyses is desirable, these approaches in dimension reduction can prove to be problematic. Status-based models do not preclude co-infection by two different antigenic variants[16], which could reduce power of new "strains" to exclude older ones. Additionally, model output could vary substantially depending on whether cross-immunity is assumed to reduce susceptibility and/or transmission, both of which reflect elements of the immune response[141]. This sensitivity towards underlying assumptions highlights the need for a deeper understanding of the inner workings of mechanistic models[16], since one would expect biologically grounded mechanisms to produce empirically plausible dynamics.

On the other end of the model complexity spectrum, individual-based models[23,76] keep track of each individual in the population, their infection history and immune status, with additional opportunities to incorporate salient host attributes that affect their disease susceptibility or transmission. Rather than fractions of the population transitioning between compartments through the law of mass action, the ability to trace the sequence of infections within a transmission chain enables the random process of genetic mutation and the emergence of new strains and eventually antigenic variants to be modelled explicitly, allowing for a more detailed representation of cross-immunity.

Regardless of one's choice of simplifying assumptions, cross-immunity should logically be associated with some metric of genetic and/or antigenic relatedness between "strains"; the method for deriving this distance is dependent on the dimensionality of strain-space within the model[179]. In the abstract case of linear strain-space[98], distance can be readily calculated through simple subtraction. When complexity is increased by representing genomes as a sequence of loci or epitopes, relatedness can be quantified using the Hamming distance[5,76], ie the number of loci occupied by different alleles, or the Euclidean distance[230]. Genetic distance can be transformed into a value of cross-immunity through the use of genotype-phenotype maps. Often, linear or exponential functions are used to represent how cross-reactivity decreases monotonically as the degree of kinship decreases[140]. However, direct transformations between genetic distance and cross-immunity gloss over the highly degenerate and non-linear relationship between genetic and antigenic distance, since a small number of genetic changes can result in large antigenic changes[231] due to the importance of a limited number of epitopes in the globular head[129]. This may inadvertently inflate the strain-space accessible to influenza viruses and necessitate the addition of further mechanisms to constrain virus diversity[76].

In addition to investigating the competitive interactions between "strains" within a closed system, some studies have investigated the process of evolution and ad hoc emergence of new "strains" over multiple seasons. The short duration of epidemics and clear wintertime seasonality of influenza in temperate regions means that it is possible to assume that new "strains" arise externally and are only introduced at the beginning of each discrete season[6]. In models with low dimension strain-space, the process of introducing new "strains" into the system can be modelled phenomenologically with a rate that is governed by a step or hazard function that takes into account the duration and/or extent of circulation of current viruses[130,132,207]. For models that represent "strains" as a sequence of loci or epitopes, the process of genetic or antigenic mutation[5,23,76,230] can occur with estimated rates, which reflect the diversity generated over the course of virus replication within an individual or at the population-level over time.

Many of the aforementioned models assume homogenous mixing between hosts, who differ in their immunity but are otherwise homogenous. These approximations, derived at the limit of large and well-mixed populations, fail to reflect the substantial heterogeneity in an individual's risk of infection and transmission potential[139,159]. Such heterogeneities can have substantial impacts on the final size and epidemic threshold, which would differ greatly from the values derived from average-based approaches[85]. The importance of host population structure has been demonstrated widely in the context of sexually transmitted[155] and livestock disease[30], no doubt aided by the relative richness and availability of data on sexual contact networks and animal movements between farms respectively. In the context of seasonal influenza, there is substantial evidence implicating the role of schools in facilitating transmission between children[123], who are key in further dissemination in the community[183] and households[255]. There is thus a need to use empirical data to identify the critical social structures underpinning disease transmission and encode them into the modelling of epidemics, to optimise targeted control strategies, such as vaccination[78] or school closures[173].

Similar to expanding the dimensionality of strain-space, social structures can be encoded in compartmental models by further dividing hosts into subpopulations. In these multi-group models, it is assumed that hosts mix homogenously within their subgroup; typically, contacts between individuals of the same subgroups will occur at a higher rate than those between different subgroups[154]. Depending on the scope of the research question, the relative weighting of within- and between-subgroup contact rates can be altered, giving contextual meaning and display social and/or geographical segregation of individuals. In metapopulation models, the subgroups could represent spatial patches across the range of geographical scales: patches could represent households, schools and workplaces within a

community[13,14,36], communities within a city[108], municipalities within a country[46,164] or countries across the world. Subgroups could also constitute different age groups, reflecting the highly age assortative contact patterns observed in large-scale empirical studies[27,85,186]. Since these multi-type models can be expressed as a system of ordinary differential equations and are fundamentally extensions of the basic *SIR* model, they remain readily amenable to analyses[14,165,166,167], using algorithmic approaches built upon classical results[112].

The issues surrounding individuality and stochasticity in transmission have long been acknowledged to be epidemiologically important[209]; in metapopulation models, their effects become further amplified as the subgroups become smaller and smaller in size[127]. When transmission between subgroups is relatively low, whether through weak coupling between subgroups or the presence of pre-existing immunity, omitting individual identities and taking a well-mixed average-based approach can lead to a substantial over-estimation of the epidemic growth rate and spatial spread[127].

Owing to the flexibility of individual-based models, structure in the host population can be rendered as a network graph consisting of nodes and edges. Nodes in a contact network represent individual hosts, whereas an edge between two nodes represents an interaction that could allow for disease transmission. Heterogeneity in an individual's contact patterns is reflected by a node's degree, the number of edges attached to it. The topology of the overall network can thus be summarised by its degree distribution, which is the key determinant of epidemiological dynamics. Both the reproduction number and final size of an epidemic are affected by the degree distribution, with the former being particularly sensitive to the degree variance[18,127,199].

For some pathogens, arbitrary degree distributions may prove to be suitable approximations for the patterns in disease-causing contacts: survey studies have suggested that sexual[155] and more broadly speaking, social networks[168,192] are highly clustered and similar to scale-free networks, whose degree distribution follows that of the power-law. Using the principles of bond percolation originating from statistical physics, exact solutions have been derived for the classical results, including the distribution of epidemic sizes and the transmissibility threshold for epidemics[192,193].

The spread of epidemics is shaped by but also reshapes network geometry: highly connected individuals and clusters of individuals are preferentially depleted leaving a residual network that is more fragmented than the original[78,194]. In the case of pathogens that cause repeated epidemics, network geometry has to not only account for contact between hosts but also the cross-immunity that they possess. These two attributes are likely to correlated and act to facilitate transmission[229]: the tendency of age groups to segregate can create clusters

of susceptible children, who are less likely to have pre-existing immunity, amplifying and facilitating community transmission.

The importance of the interaction between epidemics and geometry of immunity becomes more apparent when age structure is added to networks. With this added realism, network models predict that as a novel pandemic influenza virus sweeps through a naïve population, the burden of disease would initially be borne by children, who typically have the highest contact rates within the population[18,19]. The dramatic pruning and reshaping of underlying network structure would result in a shift in attack rates towards the adults in subsequent seasons. This changing hierarchy of individual risk has been found by some epidemiological studies of past pandemics[19,84,93], suggesting that seasonal vaccination strategies should be targeted and adjusted accordingly between seasons, in order to maximise the levels of herd immunity achieved.

In addition to changes in the structure of immunity, there is a need to further elucidate how the underlying contact network changes over time. As a consequence of changes in population demography or host behaviour, there is a need to rewire the network, through the addition and removal of nodes and edges. The impact of rewiring is dependent on the timescale it occurs over, relative to the life course of the pathogen: demographic turnover is unlikely to have much impact on the transient epidemics caused by acute infections[89], whilst endemic sexually transmitted diseases are greatly affected by the turnover in sexual partnerships. Unlike static networks, changes in the local topology of dynamic networks can have dramatic effects on global network properties. As an example, school closures would cause a global reduction in edge density, with high contact children losing their contacts whilst concurrently increasing the relative clustering of other individuals[17]. Such scenarios cannot be tackled using existing analytical frameworks, which model partnership turnover as a process of neighbour-exchange: it is assumed that the number of concurrent partnerships for any given node and hence global network properties both remain constant[259].

Given the sensitivity of model outputs towards network representation and parameterisation, there has been a shift towards constructing more realistic, data-driven networks, rather than relying on arbitrary degree distributions[199]. Owing to numerous studies employing direct observation, contact diaries[27,186] and electronic sensors[215], the availability of high resolution social contact data has increased over the past decade. Although parameterising and validating network models, in addition to their epidemiological outputs, remain far from trivial[199], Exponential random graph models have been developed to handle egocentrically sampled network data[32,250]. Within this framework, observed data is used to estimate the likelihood of an edge between two nodes, based on their nodal attributes, as well as the presence of other edges or structures within the network. Networks, each with the properties

of the original observed network, can then be independently sampled from this probability distribution.

Theoretical studies have explored the implications of explicitly modelling population structures and have compared them against traditional compartmental models. In contrast to well-mixed *SIR* models, where the critical depletion of susceptibles prevents the possibility of subsequent reinvasion, residual networks are potentially still vulnerable: the properties and robustness of the residual network towards subsequent reinvasion is dependent on both the transmissibility of the pathogen [194] and degree distribution of the original network [78]. The topology of networks has also been shown to impose further constraints to competition dynamics [17,125,194] between strains. In the context of pandemic influenza, it has been suggested that network structure is critical in enabling recurrent epidemics by limiting the spread of epidemics and extent of population immunity [124]. These results could have similar implications for the dynamics of seasonal influenza, although there has yet been any investigations into the joint effects of realistic network structures and cross-immunity, which are intrinsically correlated with each other.

2.5 Summary

From reviewing the existing literature, it is evident that many fundamental questions surrounding the biology and epidemiology of seasonal influenza viruses remain unanswered. With the substantial advances in information technology over the past decade, we have been able to collect, process and store data sets of ever-increasing size. Alongside this, technical development of experimental techniques have generated insights into facets of the immune response in individuals, as well as global evolutionary patterns for the viruses. In particular, elucidating how immune selection can bridge across these two extremes of phylodynamical scale will require the epidemiological dynamics and its drivers at local scales within discrete populations to be precisely characterised.

An abundance of data can itself be a hinderance to effective model development, especially if it lacks sufficient resolution, as is the case with ILI data. Only with high quality surveillance data, with the necessary virus-specificity and local scale resolution, can we accurately identify the most pertinent aspects of individual, populational immunity, host contact structure and other drivers. This will in turn facilitate the development of new modelling frameworks, by reducing their dimensionality and enabling their robust parameterisation. Validating these tools is crucial in ensuring that they produce meaningful outputs: firstly, quantifying the effect of these factors on epidemiological dynamics and secondly, generating better epidemiological forecasts and implementing more targeted control measures.

In the next chapter, I first evaluate existing and develop my own statistical model for epidemic detetion, to enable me to carefully characterise the city-level dynamics of seasonal influenza virus epidemics in Australia (Chapter 4). I then develop a modelling framework to investigate how population cross-immunity and host contact structure can interact to facilitate recurrent epidemics.

Chapter 3

Outbreak Detection Algorithms

This chapter details the development of a novel data-driven and statistically principled approach for the identification of epidemics in noisy and data sparse settings. I first review existing outbreak detection algorithms, their implementations and limitations, using examples from empirical influenza surveillance data. Based on these findings, I developed a more refined statistical framework, which I subsequently used in Chapter 4 to characterise and analyse the patterns of seasonal influenza activity in Australia. Whilst outside the scope and intent of my research, my framework is simple to implement and can handle real-time case count data, highlighting its potential for future use in routine influenza surveillance.

Parts of this chapter, notably the description of my statistical model for epidemic detection, have been previously published in Nature Communications (see Appendix A). The content of this manuscript, specific to model conceptualisation, development and evaluation, was led by myself; all authors (myself; Dylan Morris at Princeton University; Aeron Hurt at the University of Melbourne; Ian Barr at Federation University; Colin Russell at the University of Amsterdam) contributed to editing the final manuscript.

3.1 Introduction

The impetus behind disease surveillance is the timely detection of outbreaks so that public health authorities can intervene with control measures and potentially reduce the extent of morbidity and mortality. To this end, surveillance systems must be able to rapidly detect outbreaks as they unfold in real-time, with only limited access to case notifications as they accumulate[75]. This emphasis on situation awareness and the speed of detection over the accuracy of disease diagnosis has shaped existing approaches taken towards data collection and analysis[31,261].

Even with the advent of more sophisticated laboratory characterisation, routine syndromic surveillance remains a mainstay, due to its speed, flexibility and need for minimal reporting infrastructure. Indeed, the lack of specificity has its advantages in pandemic situations or outbreaks, when the aetiology has yet to be formally identified[196]. In order to remain agnostic towards the causative pathogen and mechanism of transmission, statistical methods have been employed in the field of prospective detection[247,261]. These approaches have their mathematical foundations in signal detection theory, where the challenge is to define a threshold level that can be used to discern an information-bearing signal from a background of noise[126,261].

In the context of epidemics, the aim is to define a threshold level, which can separate between baseline and heightened levels of disease activity. From a purely epidemic detection standpoint, the value is chosen to reflect the transition between spontaneous dead-end and sustained chains of transmission. For pragmatic applications, values can be chosen as warnings or triggers for public health interventions; in such instances, the choice of threshold level determines the sensitivity and specificity of the detection algorithm[261]. Other than the rather crude fixed thresholds[50,264], most frameworks attempt to infer this critical threshold value from the time series data[75,116,220]. However, there is no consensus on a gold standard[241]. Broadly speaking, frameworks fall into two broad categories, short-term and long-term methods. Short-term methods only require a limited set of reference values from within the focal season. They require little domain knowledge on the causative pathogen, making them highly generalisable and suited for rapid deployment during outbreaks of syndromic disease. In contrast, long-term methods aim to identify historical trends in activity and are thus more suited for recurrent seasonal diseases that exhibit strong periodicity.

However, without modelling the underlying environmental noise and disease transmission processes, which can only be possible if they were thoroughly elucidated, what constitutes an outbreak is a statistical criterion, since the choice of sensitivity and threshold values can be somewhat arbitrary[75]. Whilst not readily amenable to meaningful validation or sensitivity analyses with empirical time series, what matters is the utility of the tool: a statistically significant aberrant count is not necessarily of practical importance. The fine-tuning of outbreak detection algorithms is thus context dependent, requiring empirical judgement and domain knowledge[65]: for instance, sensitive and early detection may be beneficial from a public health perspective but as an automated detection system, it needs to be countenanced against false positives and the potential costs of economic disruption. For my intended application, I required a tool that is sensitive yet sufficiently specific to correctly identify epidemics from empirical surveillance data. Ideally, the chosen framework would rely upon

a minimal set of assumptions on the properties of epidemics, such that they can be readily validated, assessing the robustness of their results to model parameterisation.

Despite the apparent successes of many existing algorithms in detecting the onset of annual epidemics for seasonal influenza[252,255], they have inadvertently been assisted by the particular properties of influenza-like illness (ILI) syndromic surveillance data. By the very merit of being nonspecific and capturing the upper tract infections caused by a plethora of pathogens[217], ILI data streams are voluminous and display periodicity that closely follows the natural cycle of seasons[68]. This provides a distinct epidemic signal with a large amplitude and clear periodicity. Furthermore, the effects of noise may have been artificially suppressed, resulting in even more pronounced epidemics: an artefact of surveillance systems with limited sampling outside of winter months. Overall, these algorithms are unlikely to perform as well with data sets that emphasise quality over quantity.

The aim of the work described in this chapter was to develop a practical tool that can be applied to an antigenically characterised city-level influenza virus surveillance dataset from Australia and more robustly infer the timings for the onset and end of epidemics, which are necessary for downstream analysis of the effects of climatic factors and antigenic change. The smaller data set that comes with increased city-level resolution and antigenic characterisation presents unique problems for epidemic detection; here I use examples from empirical data to highlight the limitations of existing methods and demonstrate how my proposed method can overcome them.

3.2 Methods

3.2.1 Australian surveillance data

Influenza viruses from Australia were collected by the WHO Collaborating Centre (WHOCC) for Reference and Research on Influenza in Melbourne, Australia. The Melbourne WHOCC receives a subset of influenza-positive clinical samples collected by various sentinel surveillance systems across Australia throughout the year. The samples in this study were typed, subtyped and antigenically characterised by haemagglutination inhibition assay to the vaccine reference strain in use at the time of sample collection.

The dataset consists of 18,250 influenza positive cases, collected between 2000 and 2015 in the city of Brisbane, the city of Perth, the state of South Australia, the city of Sydney and the state of Victoria. The breakdown at the subtype/lineage level is as follows: A/H3 (7,661), A/H1sea (1,410), A/H1pdm09 (3,987), B/Vic (3,021) and B/Yam (2,171). All of these correspond specifically to individual cities except for the data from Victoria and South

Australia. As of June 2015, 75% and 78% of the inhabitants of the states of Victoria and South Australia resided in the cities of Melbourne and Adelaide, respectively. I therefore treated the Victoria and South Australia data as representative of city-level patterns in those two major cities, which is a fair assumption given the strong metropolitan bias in surveillance coverage[263].

Aggregation of cases by two-week periods was deemed necessary, to smoothen the time series in light of the relatively low number of cases within the data set; this relatively long timescale could however potentially obscure transient fluctuations in weather of interest in downstream analyses. Whilst weekly time series were appreciably noisier, I found a high degree of correspondence in estimated epidemic onset and end timings with values calculated using the Poisson Count Detection Method from data aggregated by two-week periods.

3.2.2 Fixed threshold

The simplest yet most subjective approach is to define a fixed threshold value, which remains constant within the season and between years[241,264]. Often, threshold levels are determined rather subjectively through visual inspection of virus activity across seasons. For instance, based on their real-world experience, the public health authorities for the Australian state of Victoria set the threshold level at 0.25 ILI cases per 100 patients seen by sentinel practices[264]. Crossing this threshold level for community activity serves as a trigger for initiating elements of their winter preparedness strategy, since it typically heralds a concurrent rise in influenza positive hospitalisations and the onset of an influenza epidemic. Unfortunately, since it was not possible to find out the proportion of ILI cases that were influenza positive and also subsequently characterised, I could not work out an equivalent fixed threshold level that corresponded to this empirically determined 0.25 ILI cases per 100 patients; instead, I determined the level by visual inspection. Initially, I set it at a relatively conservative value of 5 cases per fortnight across all cities.

3.2.3 Detection methods using short-term data

Segmented linear regression models

Rather than using a threshold value of counts to denote the transition between baseline and epidemic levels of activity, the objective of segmented linear regression models is to find the breakpoint or the timing marking an abrupt change in the growth rate in case counts[39,92]. This reflects the underlying disease process as the epidemic becomes established, going from limited dead-end chains of transmission to sustained exponential growth.

Only a subset of the time series running from the first week of winter to the week of peak incidence are considered; weekly ILI counts Y_t are log-transformed. For each season and location separately, a segmented linear model is fitted to identify the timing of the breakpoint $t_{breakpoint}$, which corresponds to the onset of the epidemic (Eq. (3.1)).

$$Y_t = \begin{cases} \beta_0 + \beta_1 t + \varepsilon_t & \text{if t} \leq t_{breakpoint} \\ \beta_0 + \beta_1 t + \beta_2 (t - t_{breakpoint}) + \varepsilon_t & \text{if t} > t_{breakpoint} \end{cases} \tag{3.1}$$

$$\varepsilon_t \sim Normal(0, \theta^2) \tag{3.2}$$

By definition, the rate of growth after the onset of an epidemic β_2 is constrained to be positive. Model error ε_t is assumed to be normally distributed with a mean of zero and variance of θ^2 (Eq. (3.2)). Identifying the breakpoint is an optimisation problem that can be solved by maximising the likelihood function $\mathscr{L}(\beta_0, \beta_1, \beta_2, t_{breakpoint})$. This problem is non-differentiable and can be solved iteratively using the Nelder-Mead simplex search algorithm[268], which is implemented in the segmented package in R[188].

Early Abberation Detection System C1 method

In response to the potential pandemics and the threat of bioterrorism in the early 2000s, the CDC proposed the Early Abberation Detection System (EARS): many statistical detection methods could not be applied to such scenarios due to an absence of pre-existing historical data for background information[122]. The most basic variant within this family of methods is C1, which compares the number of case counts y_{t_0} at time t_0 against a moving sample of the most recent n timepoints, $\{y_{t_0-n}, \ldots, y_{t_0-1}\}$. C1 is often used with daily syndromic data, with n set at 7 days[121]. The $C_1(t_0)$ statistic is defined in Eq. (3.3a), where $\bar{y}_{t_0}$ and s_{t_0} are the moving sample mean and sample standard deviation respectively Eqs. (3.3b)–(3.3c).

$$C_1(t_0) = \frac{y_{t_0} - \bar{y}_{t_0}}{s_{t_0}} \tag{3.3a}$$

$$\bar{y}_{t_0} = \frac{1}{n} \sum_{i=t_0-n}^{t_0-1} y_i \tag{3.3b}$$

$$s_{t_0}^2 = \frac{1}{n-1} \sum_{i=t_0-n}^{t_0-1} (y_i - \bar{y}_{t_0})^2 \tag{3.3c}$$

Under the null hypothesis of no outbreak, it is assumed that $C_1(t_0) \sim N(0,1)$; the threshold is typically set at three sample standard deviations above the sample mean, ie an outbreak is declared when $C_1(t_0) > 3$.

Moving Epidemic Method

The Moving Epidemic Method[252,253] is typically used for retrospective detection of epidemics. Rather than using a threshold level, it aims to split individual season into three periods: pre-epidemic, epidemic and post-epidemic. Under the premise that the baseline activity during the pre- and post-epidemic periods is minimal compared to activity during the epidemic period, the epidemic period (t_{onset} to t_{end}) can thus be expressed as the minimum number of consecutive weeks r^* with the maximum accumulated rates percentage (MAP).

For a time window of length r, the maximum accumulated rate MA^r and MAP^r for a given season are defined as follows Eqs. (3.4)–(3.5):

$$MA^r = \max_{i\in\{1,\ldots,S-r+1\}} \left\{ \sum_{t=i}^{i+r-1} y_t \right\}, \forall r \in \{1,\ldots,S\} \tag{3.4}$$

$$MAP^r = \frac{MA^r}{\sum_{t=1}^{S} y_t} \tag{3.5}$$

Where S is the number of weeks in the season and y_t are the number of observed counts in week t. A MAP curve can be drawn and smoothed by regressing MAP^r over r. The incremental change Δ^r in the smoothed values $\widetilde{MAP}^r$ between consecutive weeks is given as Eq. (3.6):

$$\Delta^r = \widetilde{MAP}^{r+1} - \widetilde{MAP}^r \tag{3.6}$$

The optimal epidemic period r^* is the value at which any further increases in r results in the inclusion of pre- or post-pandemic time points, ie a modest incremental increase in Δ^r that is lower than a predefined value δ Eq. (3.7).

$$r^* = \min_{r\in\{1,\ldots,S-1\}} \{r : \Delta^r < \delta\} \tag{3.7}$$

Previous studies using ILI deemed that sensitivity and specificity are maximised at values of δ in the range of 0.02 to 0.04[208,252]. Once the optimal length of the epidemic period r^* has been identified, the onset t_{onset} and end t_{end} timings can be found by revisiting Eq. (3.4) (Eq. (3.8)).

$$MA^{r^*} = \sum_{t=t_{onset}}^{t_{end}} y_t = \max_{i\in\{1,\ldots,S-r^*+1\}} \left\{ \sum_{t=t_i}^{i+r^*-1} y_t \right\} \tag{3.8}$$

By collating the epidemic periods across individual seasons, it can be possible to determine a threshold level, based on the case counts observed at epidemic onsets[253]. This

extension to MEM enables its use for prospective detection and has been implemented by the European Influenza Surveillance Networks.

3.2.4 Detection methods using long-term historical data

Serfling method

The Serfling Method[220] was a seminal development in time series analyses and remains widely used in the US[227,255], UK[74], France[48] and China[262]. It was designed for use with pneumonia and influenza (P&I) mortality data: similar to ILI data, P&I captures the activity from a multitude of respiratory pathogens. The Serfling approach assumes that seasonal influenza activity is superimposed upon a background of baseline activity Y_t, which takes the form of r sinusoidal seasonal component(s) (Eq. (3.9)). In addtion, there is a linear time trend across years to account for potential systematic changes in disease activity or surveillance between seasons. Fitting the baseline requires user input to identify summertime periods, when epidemic activity from seasonal influenza viruses is all but absent.

$$Y_t = \beta_0 + \beta_1 t + \sum_{i=1}^{r} \gamma_i \sin(\omega_i t) + \delta_i \cos(\omega_i t) + \varepsilon_t \tag{3.9}$$

Model error ε_t is assumed to be normally distributed with a mean of zero and variance of θ^2. From the baseline, a time-varying threshold with similar sinusoidal periodicity can be calculated by taking an upper percentile for the prediction distribution (typically 95%).

For the purposes of analysing my antigenically characterised and highly specific data set, I assume that the sinusoidal baseline ($r = 1$) instead represents sporadic background activity. Case counts in excess of the threshold would thus be attributable to true seasonal epidemic activity.

Farrington method

Instead of imposing a specific functional form for seasonality, the Farrington Method[74] uses an overdispersed Poisson generalised linear model: the count data distribution Y_t can be used to estimate the baseline counts y_{t_0} for week t_0 (Eqs. (3.10a)–(3.10b)) and compute a threshold by taking an upper percentile for the prediction distribution (typically 95%).

$$E(Y_t) = \mu_t \tag{3.10a}$$

$$Var(Y_t) = \phi \mu_t, \quad where\ \phi \geq 1 \tag{3.10b}$$

The count data distribution Y_t is estimated from the set of reference values s is formed from a window of $2w+1$ weeks, taken from each of the b prior seasons (Eq. (3.11a)). Seasonal patterns and potential trends across years are accounted for by the log-linear predictor (Eq. (3.11b)).

$$s = \{\bigcup_{i=1}^{b} \bigcup_{j=-w}^{w} y_{t_0 - i \cdot 52 + j}\} \tag{3.11a}$$

$$\log(\mu_t) = \beta_0 + \beta_1 s \tag{3.11b}$$

The Poisson count detection method

For each individual antigenic-variant specific time series, I used a Poisson count detection algorithm implemented in the Surveillance package in R [116] to distinguish periods of sustained epidemic activity from a background of sporadic interseasonal activity. I assume that the start of the calendar year falls sometime within the interseasonal period, which is justified by the scarce number of cases observed during this time of the year and the fact that it is summertime in Australia. Making no further assumptions on the exact duration and timing for the interseasonal period or epidemic onset, starting at the beginning of the year, successive fortnights y_t are evaluated using the number of cases in each of the n preceding fortnights $\{y_{t-n}, y_{t-n+1}, \ldots, y_{t-2}, y_{t-1}\}$ as reference values for sporadic activity. These reference values are used to predict a threshold value y_α: if the observed number of cases y_t exceeds the threshold y_α, the focal fortnight is marked as the two-week period of epidemic onset.

In the model notation that follows, "~" is a "sampling statement"; it denotes that a random variable is distributed according to the given distribution. Poisson distributions are parameterised as *Po*(mean), Gamma distributions are parametrised as *Ga*(shape, rate), Poisson-Gamma distributions are parameterised as *PoGa*(shape, rate) and Negative Binomial distributions are parameterised as *NegBin*(number of successes, probability of success).

The Poisson count model assumes that the reference values y_i are identically and independently Poisson distributed with a mean of λ (Eq. (3.12a)). λ itself has a Gamma-distribution as a prior (Eq. (3.12b)).

$$y_i \sim Po(\lambda), i = t-n, t-n+1, \ldots, t-2, t-1 \tag{3.12a}$$

$$\lambda \sim Ga(\alpha, \beta) \tag{3.12b}$$

From Eqs. (3.12a)–(3.12b), the posterior distribution for λ is given by Eq. (3.13).

$$\lambda \mid y_{t-n}, y_{t-n+1}, \ldots, y_{t-2}, y_{t-1} \sim Ga\left(\alpha + \sum_{i=t-n}^{t-1} y_i, \beta + n\right) \tag{3.13}$$

From the predictive distribution (Eq. (3.14)), the posterior predictive distribution can be expressed as a Poisson-Gamma (Eq. (3.15a)), or equivalently as a Negative Binomial distribution (Eq. (3.15b)).

$$f(y_t \mid y_{t-n}, y_{t-n+1}, \ldots, y_{t-2}, y_{t-1}) = \int_0^\infty f(y_t \mid \lambda) f(\lambda \mid y_{t-n}, y_{t-n+1}, \ldots, y_{t-2}, y_{t-1}) d\lambda \tag{3.14}$$

$$y_t \mid y_{t-n}, y_{t-n+1}, \ldots, y_{t-2}, y_{t-1} \sim PoGa\left(\alpha + \sum_{i=t-n}^{t-1} y_i, \beta + n\right) \tag{3.15a}$$

$$y_t \mid y_{t-n}, y_{t-n+1}, \ldots, y_{t-2}, y_{t-1} \sim NegBin\left(\alpha + \sum_{i=t-n}^{t-1} y_i, \frac{\beta + n}{\beta + n + 1}\right) \tag{3.15b}$$

The threshold value y_α can then be calculated using quantile parameter α, where y_α is the smallest value that satisfies (Eq. (3.16)).

$$p(y \leq y_{alpha}) \geq 1 - \alpha \tag{3.16}$$

During interseasonal periods, where there were often many fortnights reporting no cases, an isolated fortnight with sporadic activity can be misconstrued as the onset of an epidemic. To reduce the impact of outliers in the time series and increase specificity of the detection algorithm, I first applied the $4253H$ twice nonlinear data smoothing algorithm[254], which is a compound smoother consisting of multiple running medians.

I tested a variety of n and α parameter values and chose $n = 3$ and $\alpha = 0.12$ for the analyses presented in Chapter 4 as a good compromise between sensitivity and specificity in the identification of all of the epidemics within the time series and their individual onset and end timings, which was confirmed by visual inspection.

3.3 Results

3.3.1 Determining the onset timing for stereotypical epidemics

For seasons with well-defined epidemics, all existing short-term data detection methods perform well. The 2002 epidemic in Brisbane caused by A/H3/Moscow/10/99-like viruses is one such example (Fig. 3.1). In such stereotyped situations with large epidemics and limited interseasonal activity, the biases arising from an operator-selected fixed threshold value (Fig. 3.1a) would be partially limited: for instance, selecting a higher threshold value would mean that onset and end timings across epidemics within the same city would be consistently detected later and earlier respectively.

Segmented regression models (Fig. 3.1b) tend to estimate later timings for epidemic onset. By fitting and constraining the model to two secular linear trends, the breakpoint represents the point at which an epidemic has become established and within a regime of sustained exponential growth, in contrast to threshold based approaches that aim to identify the earliest timepoint at which counts exceed some pre-defined background baseline.

Whilst the EARS C1 algorithm can readily differentiate large epidemics from low-level background noise, it is severely impacted by zero-inflated data, since its threshold level (red line in Fig. 3.1c) is based upon the sample standard deviation from a set of recent historical values. During interseasonal periods when detected activity can be sparse, this inherent weakness is highlighted by the algorithm aberrantly detecting isolated case counts as epidemics (Fig. 3.1c).

When the distribution of case counts across the year exhibit substantial temporal clustering (Fig. 3.1d), the Moving Epidemic Method can readily segregate the season into interseasonal and epidemic periods (Fig. 3.2). With a large number of case counts and a smooth *MAP* curve, identifying the duration of an epidemic is straightforward since it is clear when an increase in window length fails to contribute to substantial increases in the number of counts captured (Eq. (3.7) and Fig. 3.2b).

Similarly, the Poisson count detection framework performs well on the control benchmark scenario. The initial median smoothing (blue line in Fig. 3.1e) reduces the effects of aberrant and isolated activity during interseasonal periods. Here, I have tuned the parameters for the Poisson model, such that the threshold level (red line in Fig. 3.1e) is sensitive enough to detect epidemics, as soon as it exceeds background activity.

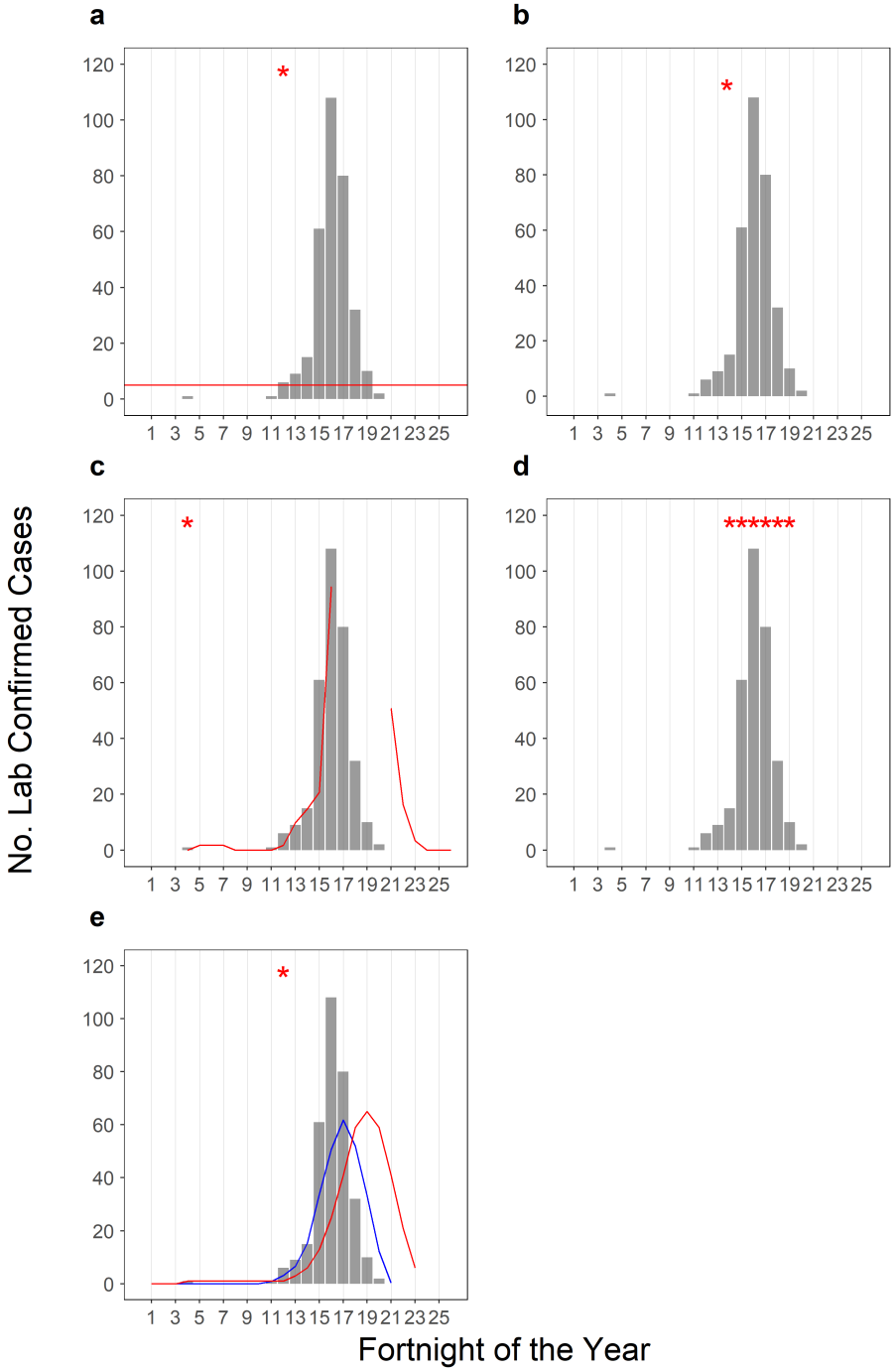


Fig. 3.1 Determining the onset timing for the 2002 A/H3/Moscow/10/99-like epidemic in Brisbane. The estimated fortnight of epidemic onsets are denoted by red stars and were determined using (a) Fixed Threshold of 5 case counts per fortnight. (b) Segmented Linear Regression Model with the red star showing the intermediate breakpoint. (c) EARS C1 Method determines the threshold level, denoted by red line, based on the distribution of case counts in previous time points. (d) Moving Epidemic Method, which identifies the overall epidemic period. (e) Poisson Count Detection Method with the blue and red lines denoting the smoothed case counts and threshold level respectively.

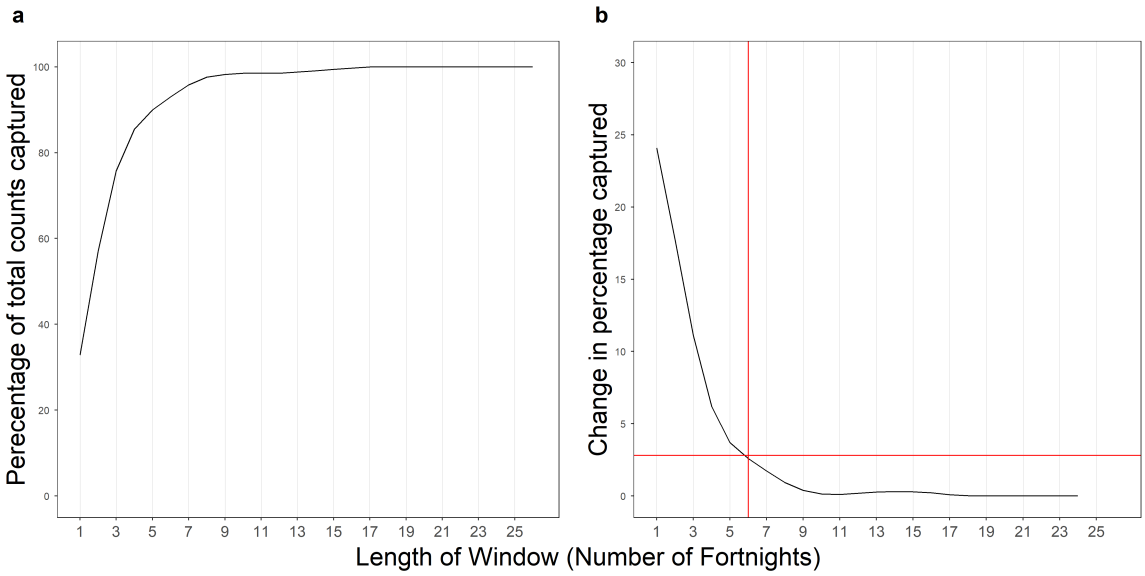


Fig. 3.2 Identifying the optimal window length for the Moving Epidemic Method. (a) The MAP curve shows that the percentage of total counts captured increases with the length of the window considered (Eqs. (3.4)–(3.5)). (b) The Δ^r curve shows that as the length of the window increases, the incremental increase in *MAP* reduces (Eq. (3.6)). Based on previous studies using ILI surveillance data, by default, a value of 2.8% for Δ^r is used. This is denoted by the horizontal red line and corresponds to the optimal window length and epidemic duration of 6 fortnights.

3.3.2 Determining the absence of epidemic activity

With virological and antigenic characterisation, it is evident that some subtypes and antigenic variants often fail to cause epidemics. In one such instance, over the entirety of the 2012 season in Melbourne, only 9 cases of B/Brisbane/60/2008-like viruses were recorded (Fig. 3.3); this sporadic activity cannot be attributed to a lack of surveillance since a total of 569 influenza positive cases underwent antigenic characterisation. Whilst it is evident from visual inspection alone that these 9 cases should not be classified as an epidemic, this example highlights the limitations of the different detection frameworks.

Since fixed thresholds are often subjectively determined via visual inspection, it is unsurprising that the level set mirrors our intuition: the same fixed threshold of 5 counts per fortnight is able to successfully filter out such cases of isolated activity (Fig. 3.3a).

The Segmented Linear Regression model and Moving Epidemic Method presupposes the existence of an epidemic in each season, resulting in low specificity since epidemic activity is always detected (Fig. 3.3b, Fig. 3.3d); such methods would require the imposition of

additional subjective criteria, ie each antigenic variant needs to exceed a minimum number of counts or proportion of total counts detected within a season.

Again, zero-inflation causes the EARS C1 algorithm to detect the onset of an epidemic at the first time point with any counts (Fig. 3.3c). To mitigate this issue, additional operator input would be necessary to scrutinise the time series; fully resolving it would be difficult, requiring the statistical criterion for the threshold level (Eq. (3.3a)) to be fundamentally redefined.

The specificity of my detection framework is enhanced by the initial smoothing (blue line in Fig. 3.3e) and modulated by subsequent parameterisation of the Poisson count model, thus preventing accidental identification of an epidemic of spurious background activity (Fig. 3.3e).

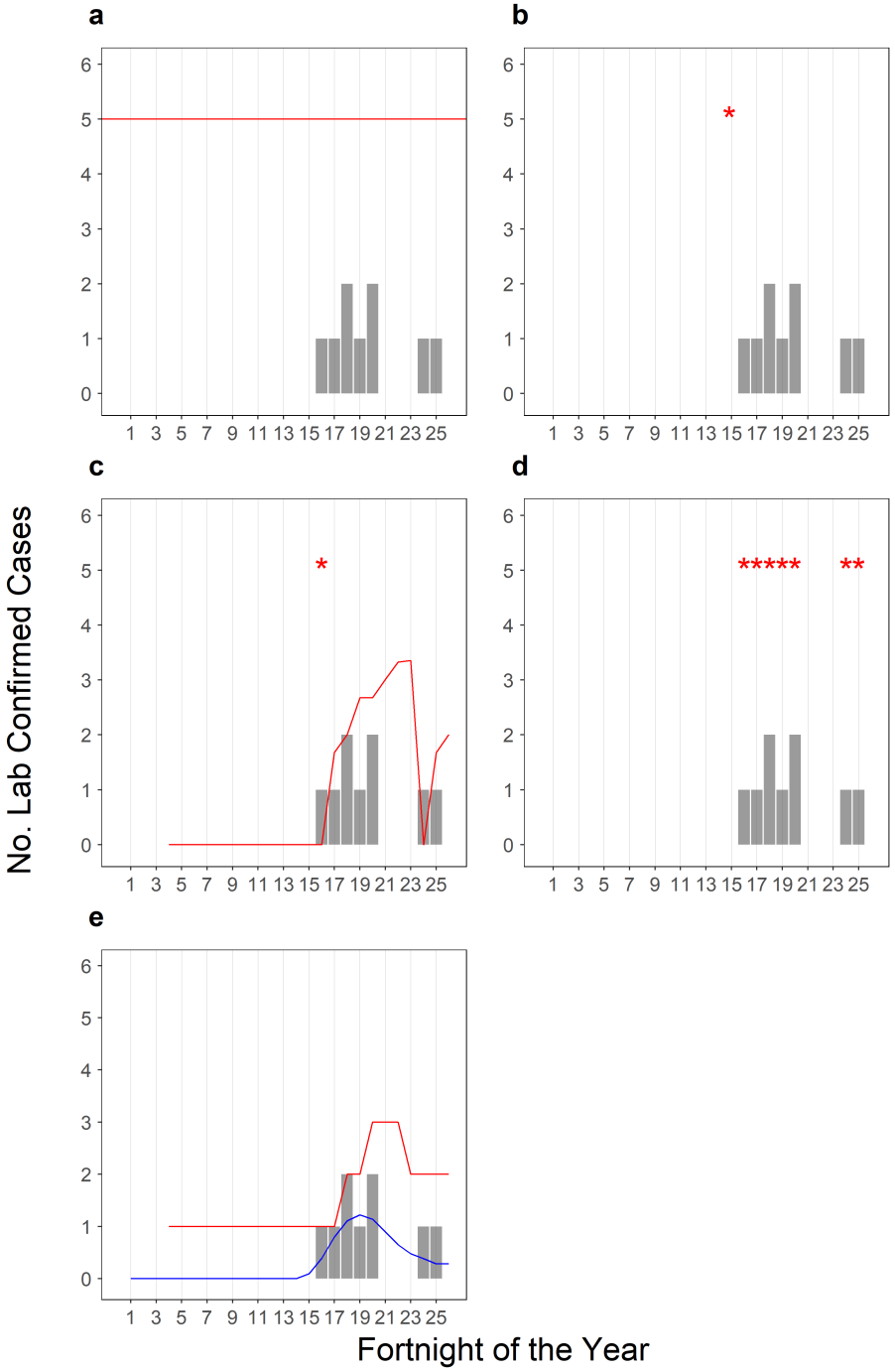


Fig. 3.3 Determining the absence of epidemic activity by B/Brisbane/60/2008-like viruses in Melbourne, 2010. Epidemic onsets are denoted by red stars and were determined using (a) Fixed Threshold of 5 case counts. (b) Segmented Linear Regression Model with the red star showing the intermediate breakpoint. (c) EARS C1 Method determines the threshold level, denoted by red line, based on the distribution of case counts in previous time points. (d) Moving Epidemic Method, which identifies the overall epidemic period. (e) Poisson Count Detection Method with the blue and red lines denoting the smoothed case counts and threshold level respectively.

3.3.3 Accounting for interseasonal activity

Whilst there is no evidence from the data to suggest any systematic changes or enhancement in out-of-season surveillance after the 2009 A/H1N1 pandemic, it is prudent to account for such potential variation, as well as more generally, the intrinsic variability in virus activity between seasons. Relative to the peak number of case counts per fortnight recorded, a rather substantial amount of interseasonal activity was recorded for A/H3/Switzerland/9715293/2013-like viruses in Adelaide, 2015 (Fig. 3.4). Fixed thresholds are by their very definition dependent on chosen values, can result in the erroneous determination of an epidemic by early sporadic activity (Fig. 3.4a): increasing th.

Segmented linear models (Fig. 3.4b) tend to perform well, as long as incidence during the epidemic period exceeds that of interseasonal activity. However, the pre-epidemic line segment is particularly sensitive to the presence of stochasticity in case counts, which can affect the inferred position of the breakpoint and epidemic onset timing. By increasing the number of counts in fortnight 1 of the year, the estimated breakpoint was shifted earlier by approximately a fortnight. Intuitively, more interseasonal background noise should result in a higher threshold and consequently a later estimated onset timing.

The Moving Epidemic Method fails to identify a meaningful epidemic period (Fig. 3.4d), as a result of its underlying approach of defining the epidemic period based on maximising the proportion of counts captured.

Both the EARS C1 (Fig. 3.4c, Fig. 3.5c) and Poisson count (Fig. 3.4e, Fig. 3.5e) methods account for baseline activity during interseasonal periods and define dynamic threshold levels. Consequently, their estimate epidemic onset timings are mostly consistent with each other. A key difference is that for the Poisson framework, a smoothing algorithm is first applied: sudden increases in activity are dampened, with the intention of preventing the aberrant detection of spontaneous activity during interseasonal periods (Fig. 3.5c,e). However, in some instances, this initial smoothing step can contribute to a delay in detecting an initially sudden but subsequently sustained increase in epidemic activity. In Fig. 3.4, the EARS C1 method (Fig. 3.4c) has determined an onset timing at fortnight 13, which is one fortnight earlier and appears to be more accurate than the timing estimated by the Poisson count method (Fig. 3.4e).

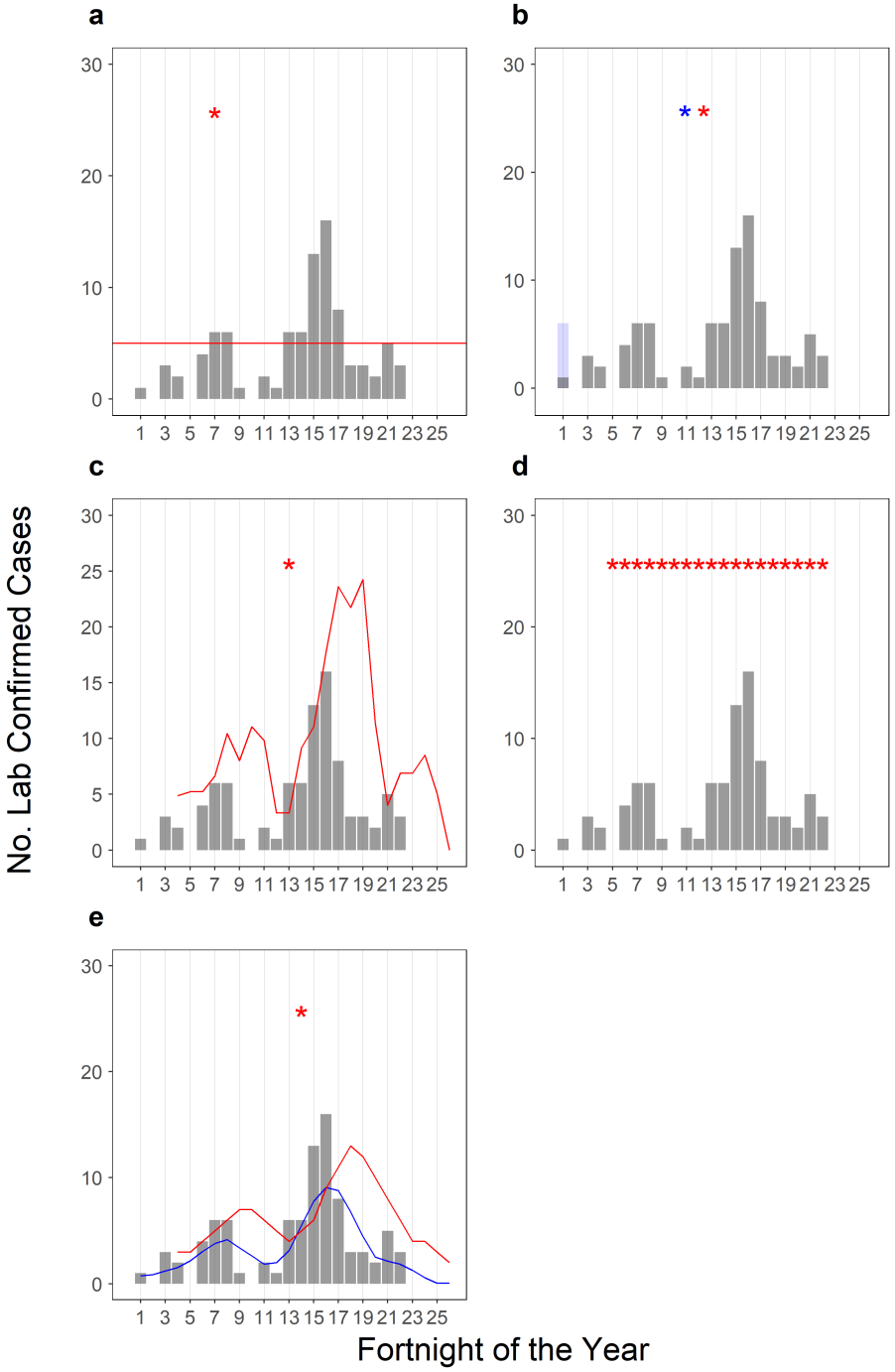


Fig. 3.4 Accouting for interseasonal activity by A/H3/Switzerland/9715293/2013-like viruses in Adelaide, 2015. Epidemic onsets are denoted by red stars and were determined using (a) Fixed Threshold of 5 case counts. (b) Segmented Linear Regression Model with the red star showing the intermediate breakpoint; a counterfactual example with increased number of counts in fortnight 1 (blue column) results in an earlier breakpoint (blue star). (c) EARS C1 Method determines the threshold level, denoted by red line, based on the distribution of case counts in previous time points. (d) Moving Epidemic Method, which identifies the overall epidemic period. (e) Poisson Count Detection Method with the blue and red lines denoting the smoothed case counts and threshold level respectively.

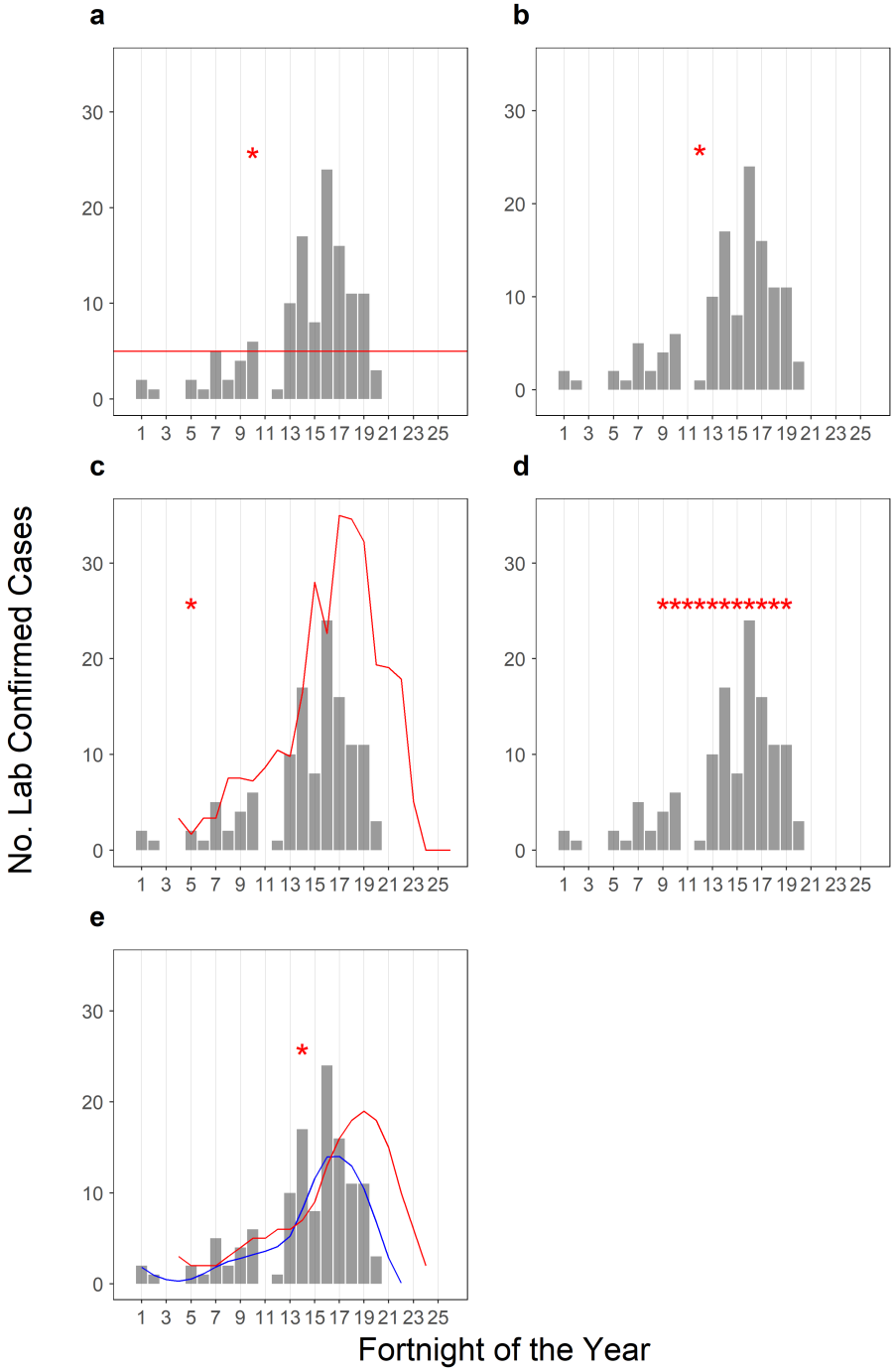


Fig. 3.5 Accouting for interseasonal activity by A/H3/Switzerland/9715293/2013-like viruses in Melbourne, 2015. Epidemic onsets are denoted by red stars and were determined using (a) Fixed Threshold of 5 case counts. (b) Segmented Linear Regression Model with the red star showing the intermediate breakpoint. (c) EARS C1 Method determines the threshold level, denoted by red line, based on the distribution of case counts in previous time points. (d) Moving Epidemic Method, which identifies the overall epidemic period. (e) Poisson Count Detection Method with the blue and red lines denoting the smoothed case counts and threshold level respectively.

3.3.4 Limitations of using historical data to infer and predict seasonal patterns and trends

The Serfling and Farrington methods attempt to derive time-varying threshold levels, by inferring seasonal patterns and trends from historical data. In contrast to the pronounced summertime troughs and wintertime peaks observed in more commonly used country-level ILI or P&I data sets, local epidemics of individual antigenic variants and subtypes fail to exhibit such predictable patterns and trends.

I attempted to fit the Serfling model to the A/H3 time series for Brisbane from 2000 to 2015 but the empirical data does not support a sinusoidal component for baseline activity (Table 3.1). In essence, the Serfling model predicted a fixed threshold that could potentially increase very modestly towards later years, resulting in performance and limitations similar to that of the basic fixed value approach discussed above.

Coefficient	Value	SE	p value
β_0	0.51	0.39	0.19
β_1	0.005	0.0021	0.02
γ_i	-0.31	0.24	0.19
δ_i	-0.62	0.4	0.12

Table 3.1 Results from Serfling's seasonal regression method (Eq. (3.3a)). The model was fitted to A/H3 case counts for Brisbane from 2000-2015. The linear trend was statistically significant ($p<0.05$), although its effect β_1 is very modest, representing a difference of approximately one case per interseasonal fortnights between 2000 and 2015.

The Farrington approach is confounded by the lack of underlying periodicity in the data and highly variable epidemic sizes: the estimation of the mean and variance of the count data distribution Y_t for fortnight t of the year is thus highly sensitive to year-to-year variation in y_t (Eqs. (3.10a)–(3.10b)). This variation negatively impacts the specificity of the model, due to its recency bias that is inherent to using the most recent b seasons as reference values. In Brisbane, the 2002 A/H3 epidemic happened to be preceded by two seasons with minimal activity and was thus readily identified (Fig. 3.6). However, despite being of similar magnitude, the subsequent 2003 season was erroneously missed.

The detrimental effect of this recency bias can be more clearly illustrated by counterfactual scenarios (Fig. 3.7). For both (a) and (b), the training data consists of A/H3 epidemics in Brisbane from the years of 2001, 2002 and 2005; the fitted models are then tested on the same 2003 season. In (a), the training seasons are ordered such that there is a trend of

increasingly early and large epidemics: consequently, the relatively late yet comparably large 2003 epidemic fails to be detected in a timely fashion. In (b), the training order is reversed so there is a trend of increasingly late and small epidemics. In this case, the 2003 epidemic is detected since the threshold level is smaller and peaks later within the year.

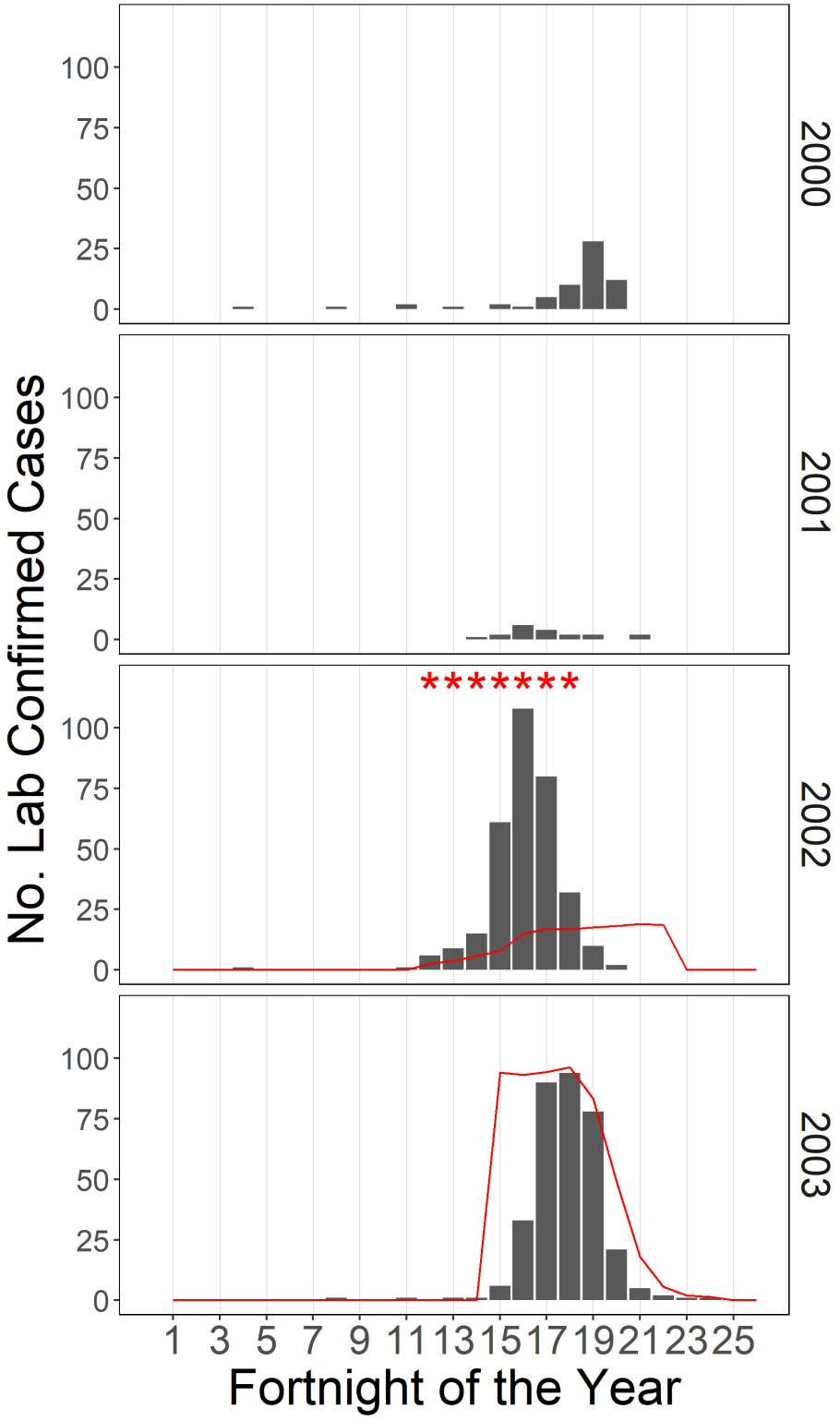


Fig. 3.6 Determining epidemic onset timings using the Farrington Method. Windows of 7 timepoints were taken from the prior two seasons as training data ($w = 3$; $b = 2$; Eq. (3.11a)), ie 2000 and 2001 were used to define the threshold level (red line) in 2002; fortnights with counts exceeding the threshold are marked with red stars.

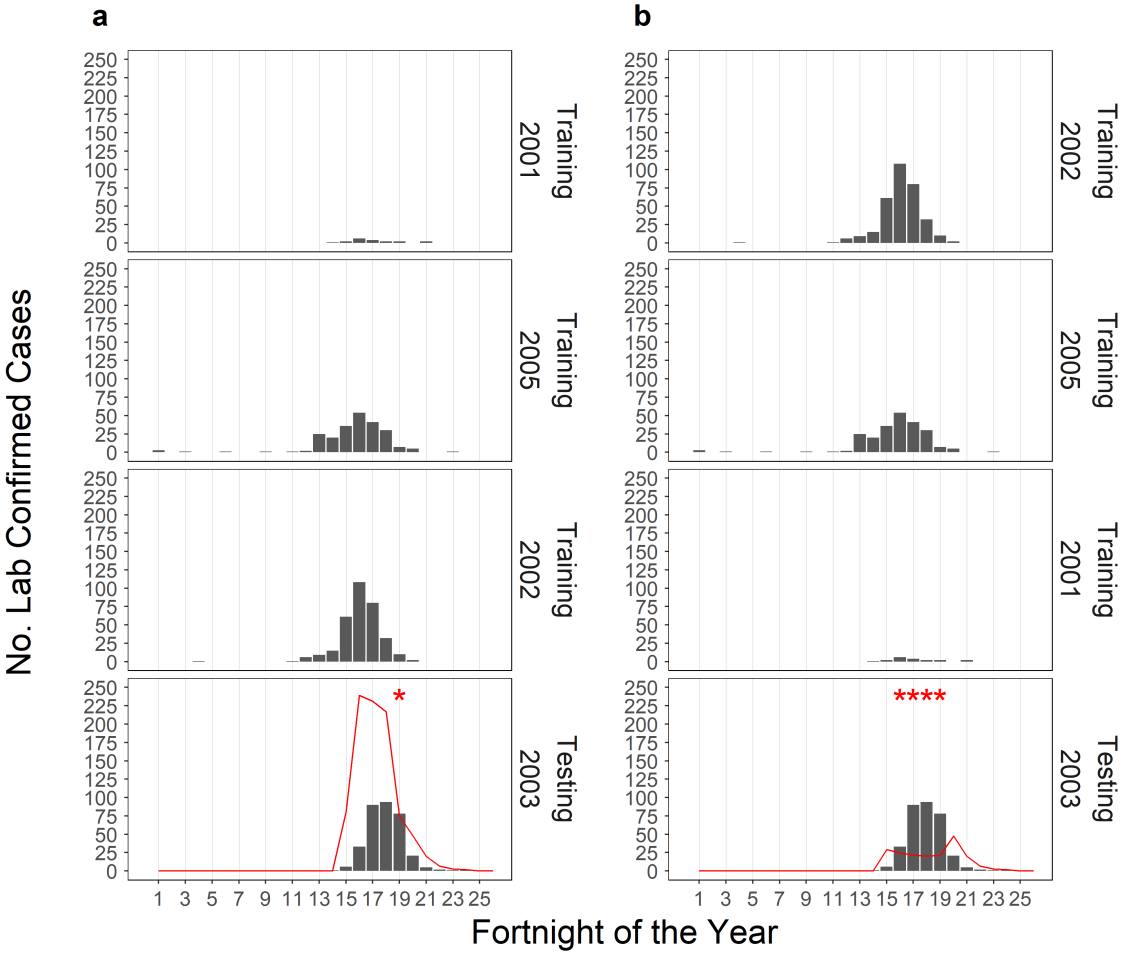


Fig. 3.7 Limitations of the Farrington Method. Here, I select various A/H3 epidemics from seasons in Brisbane and demonstrate how spurious trends can be inferred from empirical data. Both examples use the same seasons within the training and testing data sets; windows of 7 timepoints were taken from each of the three seasons within the training data set ($w = 3$; $b = 2$; Eq. (3.11a)). The training seasons are ordered to create trends of (a) increasingly early and large epidemics; (b) increasingly late and small epidemics. Fortnights with counts exceeding the threshold levels (red lines) are marked with red stars.

3.4 Discussion

In Australia, influenza is a notifiable disease and there is a well-established national surveillance network, which monitors influenza activity throughout the year using multiple state-level systems[88,241]. A wide range of modalities of data are collected, ranging from syndromic ILI cases collected by the online self-reporting FluTracking system to clinical presentations and laboratory-confirmed influenza positive cases at sentinel general practitioners

and hospitals. A subset of these influenza positive cases undergo further typing, subtyping and antigenic characterisation, which is carried out by the WHOCC in Melbourne. With this systematic approach, one can be confident that the collated time series for case counts are representative of underlying antigenic variant-specific activity occurring at the local level; in most countries, influenza positive cases are rarely subtyped, let alone antigenically characterised.

Initial visual inspection of the time series revealed some key features: 1. the number of cases recorded during summer months and the magnitude of epidemics can vary substantially between seasons; 2. the timing of epidemics vary between cities and seasons; 3. within the same season, the timing of local epidemics differs between subtypes and individual antigenic variants. These heterogeneous features result in an overall reduction in the signal to noise ratio, especially since the additional resolution derived from virus characterisation comes at the cost of sample size. A previous study of seasonal influenza in Australia limited to just 8 seasons utilised a data set of 450,000 influenza positive cases with partial virus typing; in contrast, my Australia 15-year data set consisted of 18,250 influenza positive cases with full antigenic characterisation.

These particular features of local epidemics of antigenic variants are often incompatible with the underlying parametric models and their assumptions of many existing detection frameworks. Since the exact timing and magnitude of epidemics for each subtype are highly variable between years, the lack of clear periodicity or trends across years means that frameworks that rely upon long-term historical data, such as the Serfling[220] and Farrington[74] Methods, suffer from poor specificity[39,50]. This heterogeneity is most likely a product of intrinsic variation in virus activity, although systematic changes in surveillance intensity cannot be ruled out definitively. Until the sources of this stochasticity are fully elucidated and can be accounted for, these frameworks are wholly unsuitable for analysing highly resolved virus surveillance data. In fact, model misspecification results in performance that is comparatively worse (Fig. 3.6) than that of the most basic fixed threshold approach (Fig. 3.1a).

Clearly, the analysis of city-level and antigenically characterised data requires a framework that considers each season independently. Fixed thresholds, whilst easy to implement, need to be decided on a case-by-case basis, since they are not readily generalisable across cities[263,264] or indeed subtypes. To overcome this lack of consistency, a more statistically principled approach needs to be implemented: whilst limited in number, case counts during tentative interseasonal periods provide some measure for the stochasticity in virus activity for the focal season. In the context of ILI, this task is relatively trivial when data is plentiful and the magnitude of epidemics is consistently and substantially larger than that of the

background noise. All existing short-term methods are able to consistently detect the onset of large epidemics (Fig. 3.1), with the exception of EARS C1, which struggles when the data is zero-inflated and can get erroneously triggered by isolated and sporadic activity.

Retrospective detection methods, such as segmented linear models and MEM, draw upon the case counts across the entirety of a season; intuitively, identification of epidemic onset, through growth rates and clustering of counts respectively, would benefit from having a larger set of interseasonal reference values. However, this becomes a liability when epidemic activity is obscured by the presence of substantial interseasonal activity (Fig. 3.3b, Fig. 3.3d). Furthermore, these methods assume that every season can always be divided into distinct non-epidemic and epidemic periods, which taken to its logical extreme, results in model misspecification in seasons where there is clearly an absence of heightened or epidemic activity.

These limitations can be partially bypassed by limiting reference values for a focal fortnight to only the most recent time points, which reflects virus activity and case counts are more likely to short-term than long-term correlations. In interseasonal periods, the existing EARS C1 algorithm can perform well, so long as there is protracted virus activity; it fails to derive a meaningful threshold when confronted with consecutive time points with no recorded cases.

To a certain extent, these limitations could be remedied by the addition of rules for what constitutes an epidemic, such as establishing a minimum number of case counts, or by disregarding all erroneous detections outside of a vaguely defined wintertime period. Fundamentally, however, this would merely append additional layers of subjectivity, further complicating attempts to assess the robustness of epidemic timing estimates and downstream analyses towards model parameterisation.

My proposed Poisson count detection framework is optimised for use with highly resolved surveillance data and overcomes the specific challenges posed by the relatively small data set. Like EARS C1, my method need only make minimal assumptions about the exact timing of epidemic onset, only that it is highly unlikely to occur at the beginning of the calendar year and the height of summer in Australia. The impact of isolated activity is reduced through the initial median smoother, whilst preserving the important trends during sustained and increasing epidemic activity. The threshold levels are dynamically informed by season, city and virus-specific activity. It uses a limited set of recent time points to estimate the distribution for underlying count activity; the choice of Poisson model avoids the zero-inflation issue. Model sensitivity is only dependent on two parameters (Eq. (3.12b)) and does not require additional arbitrary criteria for epidemics; the robustness of timing estimates

towards detection sensitivity and specificity can be easily evaluated in a systematic manner (see Chapter 4).

Chapter 4

Epidemiology of seasonal influenza in Australia

In this chapter, I analyse a 15-year city-level dataset of 18,250 laboratory-confirmed and antigenically-characterised influenza virus infections from Australia and investigate the effects of previously hypothesised environmental and virological drivers of influenza epidemics. This chapter is an abridged form of the work that has been previously published in Nature Communications (see Appendix A); the epidemic detection algorithm and its development are described in Chapter 3. The work was undertaken in collaboration with Dylan Morris at Princeton University, Aeron Hurt at the University of Melbourne, Ian Barr at Federation University and my supervisor, Colin Russell at the University of Amsterdam.

Aeron Hurt and Ian Barr generated and kindly provided access to the dataset. Colin Russell provided guidance in the overall conceptualisation of the study and interpretation of results. I performed the data curation and statistical analysis, with assistance on the Bayesian Modelling (see Section 4.2.7) from Dylan Morris. I wrote the first draft of the manuscript; all authors contributed to the critical review and revision of the manuscript.

4.1 Introduction

Seasonal influenza virus epidemics are a substantial source of disease burden and result in 650,000 deaths each year[269]. Although seasonal influenza viruses circulate globally, prevention and treatment occur at the level of regions, cities, and communities. At these scales, the timing, duration and size of local influenza virus epidemics can vary substantially from year-to-year[63,79], but the underlying causes of this variation are poorly understood. Better understanding of the factors that govern epidemic onset and magnitude could allow

for accurate and timely epidemiological forecasts[86] and more efficient allocation of public health resources[203].

In temperate regions of the Northern and Southern Hemispheres, influenza virus activity is most common in winter months but the mechanistic basis of this seasonality remains unclear. Experimental studies demonstrated that reductions in temperature and absolute humidity enhance viral stability and aerosol transmission[161,162,221]. However, epidemics in tropical and subtropical regions often occur during periods of high temperature and humidity[240].

Climatic fluctuations have been implicated as triggers for influenza epidemics in temperate regions. A study of state-level epidemiological data from the United States found that influenza epidemics sometimes follow two-week periods of anomalously low absolute humidity[222]. Subsequent studies of epidemiological activity have found similar results using prefecture-level data from Japan[225], city-level data from the New York Metropolitan Area[62] and region-level data from France[219].

Influenza virus evolutionary dynamics are another theorised driver of influenza virus epidemiology. Within each type and subtype of seasonal influenza virus, new major antigenic variants arise every 3-8 years[146,231]. New variants partially escape the immunity induced by prior infections and vaccinations, rendering a higher fraction of individuals susceptible to infection. Epidemiological theory predicts that epidemics caused by a new antigenic variant should therefore be larger than epidemics of previously-circulated variants[130,131].

Antigenic change could also produce earlier and more spatio-temporally synchronous epidemics. When more individuals are susceptible, fewer transmission chains go stochastically extinct so each new introduction of a virus into a population has a higher chance of causing an epidemic. Consistent with this, studies have suggested that antigenic change is associated with earlier epidemics in Israel[29] and with more synchronous epidemics among cities in the United States[39,102,256], Japan[213] and Australia[92].

Studies of environmental and virological drivers of influenza virus epidemiology, including the studies referenced above, have been limited by three factors: 1. the reliance on influenza-like illness (ILI) data; 2. the aggregation of ILI or virologically confirmed data over large geographical scales (state/province/country); 3. where virologically confirmed data are available, the use of data without subtype and antigenic variant-level resolution.

ILI data frequently includes a wide variety of respiratory infections[217] and limited laboratory characterisation obscure influenza virus type/subtype- and antigenic variant-specific patterns. These patterns become superimposed upon each other due to aggregation of ILI or virologically confirmed data to ecological scales (state/province/country) that sum over multiple local epidemics (county/city/town), which can individually vary substantially in timing, magnitude and influenza virus composition. All together, these sources of obfuscation

make it difficult to disentangle local-level, antigenic variant-specific patterns and critically investigate the impact of virus antigenic change.

Here I use a 15-year dataset of 18,250 typed, subtyped, and antigenically characterised seasonal influenza viruses from the five most populous cities in Australia to investigate the impact of environmental and virological factors on the timing and magnitude of city-level influenza virus epidemics. I find that climatic fluctuations and virus antigenic change have no consistent effects on epidemic onset timing or size while epidemic onset timing itself and heterosubtypic competition have substantial impacts on epidemic size and virus subtype composition. The lack of consistent effect of easily measured climatic and virus antigenic properties and seeming dominance of noisy short-term transmission processes likely diminishes the feasibility of meaningful long-term influenza epidemic forecasting at local scales.

4.2 Methods

4.2.1 Australian surveillance data

The description of the surveillance dataset from Australia is covered within Chapter 3. Pertinent to the subsequent analyses in this chapter, all epidemic activity of all subtypes for the 2009 season was excluded from all analyses because of the 2009 A/H1N1 pandemic. Unsurprisingly, patterns of virus circulation during the pandemic were anomalous compared to typical seasonal influenza virus epidemics and potentially distortive of the patterns I sought to characterise.

4.2.2 Estimation of epidemic timing

The exact timing of interseasonal periods of sporadic activity and epidemic onset for each subtype is highly variable between years, even for individual cities, so I used my Poisson count epidemic detection framework (see Chapter 3 for methodological development) to determine the onset and end of each epidemic independently for each antigenic variant, season and city.

Aggregation of cases by two-week periods was deemed necessary, to smoothen the time series in light of the relatively low number of cases within the data set; this relatively long timescale could however potentially obscure fluctuations in weather that occur at shorter scales. Whilst weekly time series were appreciably noisier, I found a high degree of correspondence in estimated epidemic onset and end timings with values calculated from

data aggregated by two-week periods: indeed, my results were robust to aggregation by week (see Appendix B.2 in Appendix B for sensitivity analyses).

I deemed an antigenic variant to have failed to cause an epidemic if, within a season, the algorithm was unable to define an epidemic period; I confirmed all putative failures by visual inspection of the raw time series. Once the epidemic period was defined, the size of an epidemic per antigenic variant was calculated using the estimated resident population for that particular year and city.

4.2.3 Normalisation of epidemic incidence

For each epidemic, the incidence of laboratory-confirmed cases per million people was calculated from the number of raw counts. Given the positive skew in the distribution of epidemic incidences, individual incidence values were log transformed. To enable comparisons within subtypes, I needed to account for potential differences in surveillance intensity and normalise values between cities: I subtracted off the overall city-specific mean log transformed incidence from each individual value. Although the apparent heterogeneity in the effect of antigenic change and prior immunity between subtypes suggests that data should be stratified by subtype, I repeated my analyses with data aggregated and normalised across subtypes in order to increase statistical power. Individual log transformed values for each epidemic were instead subtracted by the overall city- and subtype-specific mean of log transformed values.

4.2.4 Virus antigenic characterisation by haemagglutination inhibition assay

For my analyses, I defined an antigenic variant as in Smith et al.[231], where an antigenic variant is sufficiently different from preceding variants to warrant an update of the seasonal influenza virus vaccine. To this end, my analyses only accounted for major antigenic changes and did not account for the possibility of small or gradual antigenic changes (neither of which are well studied for seasonal influenza viruses).

The haemagglutination inhibition (HI) assay data used in this study only compared the test virus and the then current reference vaccine strain to assess whether or not viruses had changed antigenically. However, this comparison to a single reference point is potentially problematic given that new Southern Hemisphere's influenza vaccine composition recommendations are made every September. This is usually after the end of the influenza season in Australia and may lead to misidentification during antigenic characterisation of submitted samples during the preceding season where samples containing a novel antigenic

variant may have been tested with sera raised against its predecessor variant. To ameliorate this potential source of bias, I compared the antigenic characterization data against phylogenetic data. This comparison revealed two instances for A/H3 viruses where the reference strain comparison by HI was misleading regarding the antigenic composition of an epidemic. There were a substantial number of laboratory-confirmed cases attributable to A/H3/Fujian/411/2002-like viruses in 2004 but phylogenetic analyses of sequences dated 2004 show that the Fujian/411/2002-like viruses had already been replaced by the novel California/7/2004 variant viruses. Similarly, in 2005, a substantial number of samples initially identified as A/H3 California/7/2004-like viruses were phylogenetically in the new A/Wisconsin/67/2005 variant group.

To account for the likelihood of misidentification due to delays in updating nomenclature, I assumed that all A/H3 cases in 2004 were California/7/2004-like and in 2005 were Wisconsin/67/2005-like antigenic variants. Additional analyses were also carried out with the raw data set without these corrections (see Figs. B.9–B.13 and Table B.6 in Appendix B), and lead to no significant or substantive differences to my findings.

4.2.5 Demographic data

I retrieved estimated resident populations for Adelaide, Brisbane, Melbourne, Perth and Sydney on 30 June of each year from 2000 to 2015 from the Australian Bureau of Statistics (http://stat.abs.gov.au/).

4.2.6 Climate data

For each of the five cities, I compiled the mean temperature (°C) and relative humidity (%) from TuTiempo (https://en.tutiempo.net/) and calculated the mean absolute humidity (gm^{-3}) for each two-week period from 1985 to 2015. For each of the 26 two-week periods of the calendar year, I calculated 31-year mean temperature $\overline{T}$ and absolute humidity $\overline{AH}$ values, see Eqs. (4.1)–(4.2).

$$T' = T - \overline{T} \tag{4.1}$$

$$AH' = AH - \overline{AH} \tag{4.2}$$

Following Shaman et al.[222], I generated a synthetic distribution of wintertime climatic values by bootstrap sampling. In order to maintain the sampling structure and control for anomaly variability among the cities, 15 n-week continuous blocks were randomly sampled

from 01 April - 31 August, 1985-2015 for each of the five cities. These 75 samples were then averaged to produce a mean $\overline{T}$ and $\overline{AH}$ values. This was repeated 100,000 times to produce a bootstrapped distribution of average values. The statistical significance for the mean $\overline{T}$ and $\overline{AH}$ values derived from the 75 empirically observed earliest-in-the-season epidemics was then calculated non-parametrically, by determining the quantile for the observed values within the bootstrap distributions. This bootstrap was repeated at the city level to see if there were geographical differences with individual bootstrap distributions were created for each city.

I also evaluated whether or not epidemic onset is associated with climatic fluctuations that are anomalous for that particular time of the year. By definition, for any given 2-week period of the year, the 31-year mean for T' and AH' is 0. I used a Wilcoxon one-sample test to assess whether there were reductions in climatic values in the observed set of T' and AH' values in each of the 2-week blocks preceding the onset of the earliest epidemic of the season.

4.2.7 Bayesian hierarchical regression

To estimate reasonable bounds on the possible effects of climate and antigenicity on epidemic size, I used a Bayesian hierarchical model that partially pooled effect size estimates across subtypes, increasing the capacity to detect any potential effects without assuming a priori that effects should be the same across different subtypes. I fit the model using Markov Chain Monte Carlo (MCMC) with Stan[233] and its R interface rstan[234]; Stan implements a no-u-turn sampler (NUTS)[233]. All data and code needed to reproduce the analysis and figures is provided in the project Github repository, along with directions in a README file.

In the model notation that follows, the symbol "~" is a "sampling statement"; it denotes that a random variable is distributed according to the given distribution. Normal distributions are parametrised as Normal(mean, standard deviation), generalised Student T distributions are parametrised as Student-T(degrees of freedom, location, scale). Positive constrained normal distributions (Half-Normal) are parametrised as Half-Normal(mode, standard deviation).

I predicted log incidence minus city-and-subtype-specific mean log incidence as a function of the following predictor variables:

X_1: whether the epidemic was the first epidemic for an antigenic variant in the city (binary, yes or no)

X_2: cumulative prior incidence of the antigenic variant (measured as log(total prior cases / city-and-subtype-specific mean cases per epidemic))

X_3: mean absolute humidity during the epidemic, from the start to end date of the epidemic (measured as fortnight of the year)

X_4: start date of the epidemic (measured as fortnight of the year)

X_5: whether the epidemic was the earliest epidemic (of any subtype) in the city that year (binary, yes or no)

X_6: the cumulative amount of influenza activity (of any subtype) in the city that year prior to the epidemic

X_7: mean rainfall during the epidemic, from the start to end date of the epidemic (measured as fortnight of the year)

I omitted mean epidemic temperature as a predictor as it was highly collinear with absolute humidity. Any observed large effect of absolute humidity could therefore theoretically have been attributable to temperature, though in practice I estimated an effect near zero for absolute humidity.

I made a linear prediction of an epidemic's normalised size given its values for $X = (X_1, \ldots, X_7)$. Effect sizes b_i for each predictor X_i were subtype-specific, with b_{ij} denoting the effect of variable i for subtype j. I also estimated subtype-specific intercepts a_j.

I included cumulative antigenic variant activity and prior activity in the year only for old antigenic variants and epidemics that were not first of the year, respectively, that is, as interaction terms with one minus the corresponding binary variables. So the predicted mean centred log size $\langle y_k \rangle$ of an epidemic of subtype j is given by Eq. (4.3), where X_{ik} denotes the value of X_i for epidemic k. Following Gelman[90], I mean-centred and scaled continuous predictors so that effect sizes b would be directly comparable between binary and continuous predictors.

$$\begin{aligned}\langle y_k \rangle = a_j + b_{1_j} X_{1_k} + b_{2_j} X_{2_k}(1 - X_{1_k}) + b_{3_j} X_{3_k} + b_{4_j} X_{4_k} + \\ b_{5_j} X_{5_k} + b_{6_j} X_{6_k}(1 - X_{5_k}) + a_j + b_{7_j} X_{7_k}\end{aligned} \tag{4.3}$$

I assumed that observed epidemic sizes y_k were normally distributed about their predicted sizes $\langle y_k \rangle$ with an unknown, estimated standard deviation σ_y (Eq. (4.4)):

$$y_k \sim Normal(\langle y_k \rangle, \sigma_y) \tag{4.4}$$

I assumed subtype effect sizes b_{ij} for each predictor i and subtype j were normally distributed about a general mean effect size $\langle b_i \rangle$, with an unknown, estimated predictor-specific standard deviation σ_{bi} (Eq. (4.5)):

$$b_{ij} \sim Normal(\langle b_i \rangle, \sigma_{bi}) \tag{4.5}$$

Likewise, I assumed intercepts a_i were normally distributed about a mean intercept $\langle a \rangle$ with an unknown, estimated standard deviation σ_a (Eq. (4.6)).

$$a_i \sim Normal(\langle a \rangle, \sigma_a) \tag{4.6}$$

I assumed predictor-specific effect size standard deviations σ_{bi} were half-normally distributed with mode 0 and an unknown, estimated standard deviation σ_b (Eq. (4.6)).

$$\sigma_{bi} \sim Half-Normal(0, \sigma_b) \tag{4.7}$$

I placed weakly informative[91] positive-constrained half-normal priors on the intercept, effect size, and error term standard deviations σ_a, σ_b and σ_y (Eqs. (4.8)–(4.10)). Weakly informative priors rule out biologically or mathematically implausible parameter values while allowing data rather than assumptions to inform inferences regarding plausible values.

$$\sigma_a \sim Half-Normal(0, 0.5) \tag{4.8}$$

$$\sigma_{bi} \sim Half-Normal(0, 1) \tag{4.9}$$

$$\sigma_y \sim Half-Normal(0, 1) \tag{4.10}$$

I placed a weakly informative Gaussian prior on the mean intercept $\langle a \rangle$ (Eq. (4.11)) and a weakly informative Student-T prior on the mean effect sizes $\langle b_i \rangle$ (Eq. (4.12)):

$$\langle a \rangle \sim Normal(0, 1) \tag{4.11}$$

$$\langle b_i \rangle \sim Student-T(3, 0, 2.5) \tag{4.12}$$

The intercept prior was based on the degree of variation in the normed outcome variable to cover it while ruling out intercepts much larger or smaller than the largest and smallest observations. The effect size prior was based on a recommendation for weakly informative regression effect size priors (for scaled predictors) from the Stan prior recommendation wiki (https://github.com/stan-dev/stan/wiki/Prior-Choice-Recommendations).

I ran four MCMC chains, each with a 1000 step sample warmup period followed by 1000 saved posterior samples, for a total of 4000 posterior draws. I verified convergence by inspecting trace plots and confirming that all parameters had sufficiently low $\hat{R}$ values (all

$\hat{R} < 1.005$) and sufficiently large effective sample sizes (all $n_{eff} > 16\%$ of total sample size). I visualised posteriors as quantile dotplots[77] to aid in visual estimation of distributions.

4.2.8 Data availability

All of the data and code for these statistical analyses and statistical models are available at the following Github repository: https://github.com/edwardkslam/australian_seasonal_flu.

4.3 Results

4.3.1 Australia laboratory-confirmed influenza

I aggregated 18,250 laboratory-confirmed and antigenically characterised cases of seasonal influenza viruses from 2000 to 2015 by two-week (14-day) periods, creating a set of subtype- and antigenic variant-specific time series for the five most populous cities in Australia: Sydney (~5.5 million people), Melbourne (~5.0 million), Brisbane (~2.4 million people), Perth (~2.3 million), and Adelaide (~1.4 million) (Fig. 4.1). I excluded all virus cases from the 2009 season from all analyses because the 2009 A/H1N1 virus pandemic was atypical compared to seasonal epidemics and likely to be driven by different processes, affecting both epidemic dynamics and data collection of A/H1pdm09, as well as the other subtypes. Using a Poisson count detection method (Chapter 3), I identified periods of sustained, above-baseline levels of epidemic activity for each antigenic variant in each city. To facilitate comparisons among cities, I calculated the laboratory-confirmed incidence per 106 individuals using the annual estimated resident population values of each city[10].

Epidemic magnitude and most common virus subtype varied substantially among cities (Fig. 4.1). For example, during the 2002 season, A/H3 and B/Vic viruses were the most common strains in both Brisbane and Sydney. Absolute A/H3 virus incidence in Brisbane was much higher than in Sydney (186 versus 38.0 cases per 106 individuals), as was absolute B/Vic incidence (40.3 versus 22.7 cases per 10^6 individuals). But B/Vic had a substantially higher relative incidence in Sydney than in Brisbane (37% of all cases, versus only 18%). In some seasons, a virus antigenic variant caused a major epidemic in one or more cities but failed to produce any observable above-baseline activity in another city. For example, in 2006, the A/Wisconsin/67/2005 (H3N2) virus variant caused epidemics in Brisbane, Perth and Melbourne, while above-baseline levels of activity were completely absent in Adelaide.

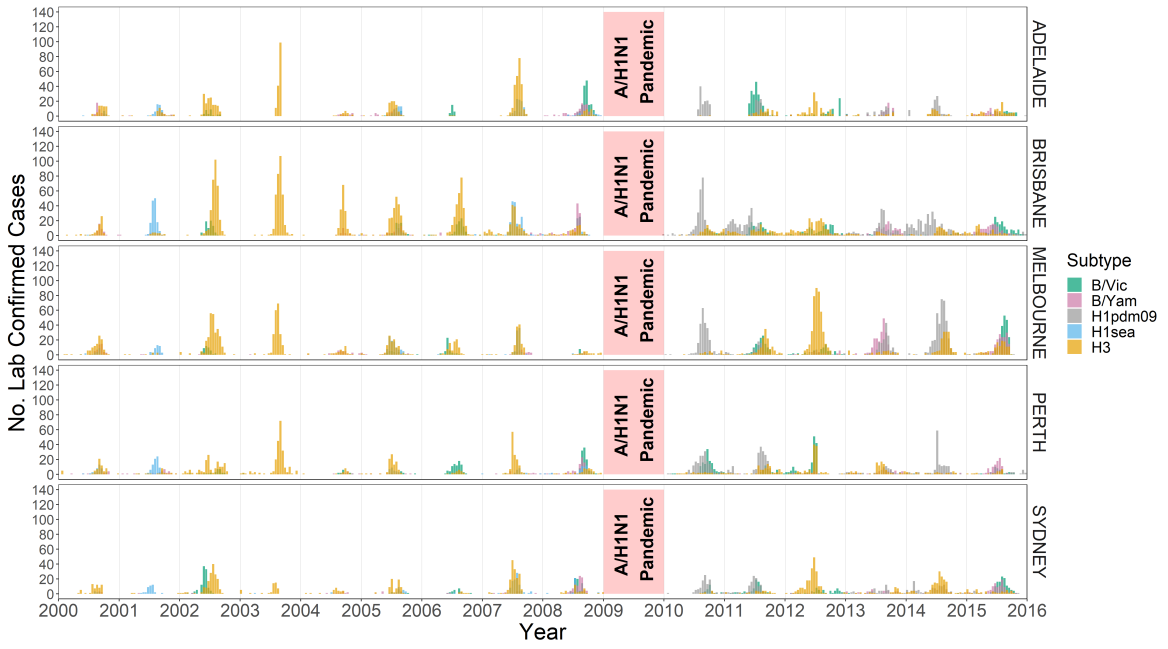


Fig. 4.1 Number of laboratory confirmed seasonal influenza virus infections from 2000 – 2015 for the five largest cities in Australia. Cases are aggregated by two-week periods, stratified by city and coloured by subtype/lineage.

4.3.2 Effect of climatic factors

Epidemic onset timing varied substantially within and among cities and virus subtypes (Fig. B.1). Previously, Shaman et al.[222] showed that the two-week period preceding the onset of state-level ILI epidemics in the United States was often marked by unusually low temperatures (T) or absolute humidities (AH). Fluctuations in these climatic factors from the historic averages expected for that specific day of the year (T' and AH' respectively) were anomalously large and negative when compared against a bootstrapped distribution of random samples from the historical records of observed daily climatic fluctuations recorded over wintertime (defined as 01 October - 28 February).

Following the same bootstrap sampling method (see Section 4.2) and aggregating epidemics across all five Australian cities, there were no statistically significant differences (all $p > 0.05$, see Table B.1 in Appendix B) between the bootstrapped distribution of random samples of typical wintertime fluctuations (01 April - 31 August for Australia) and the observed fluctuations in anomalous temperature and absolute humidity over the two, four and six week periods immediately prior to the onset of the earliest epidemics from 2000-2015 (excluding 2009, 15 years x 5 cities = 75 epidemics in total). Individual city-by-city analyses (Fig. B.2 and Table B.2) showed that there was substantial local variation but no consistent

patterns. Epidemic onset times coincided with both high and low temperature and absolute humidity periods, and there were no statistically significant patterns in four of the five cities.

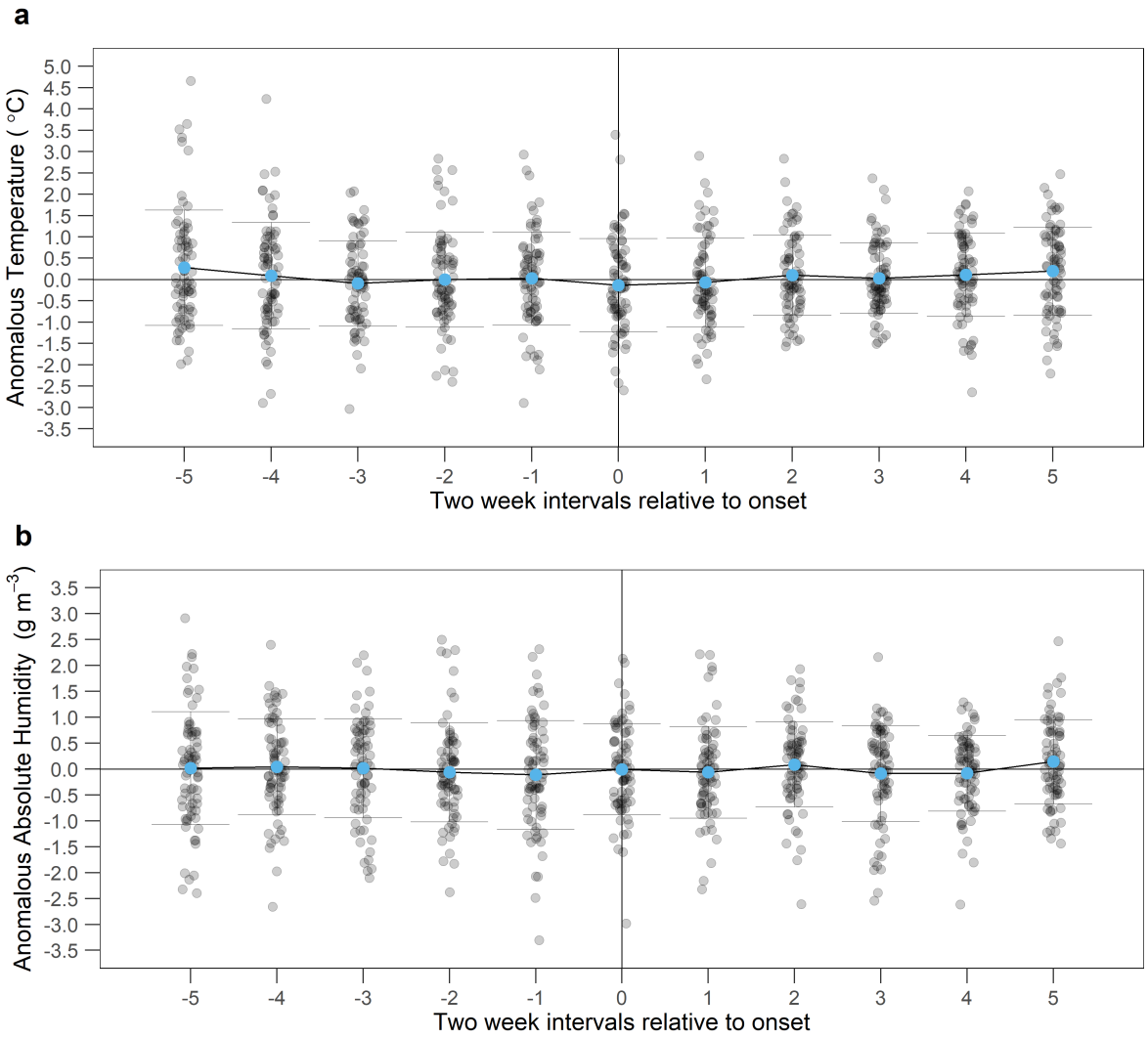


Fig. 4.2 Climatic conditions around epidemic onset. (a) Anomalous temperature T' and (b) absolute humidity AH' prior to and after epidemic onset across all five cities. Epidemic onset is marked by the vertical line at 0. For the earliest onset epidemic in each season and city (15 years x 5 cities = 75 epidemics), T' and AH' for each time point are represented by grey points: a point below the horizontal line denotes that the value is lower than the 31 year city-specific mean. Blue points show the mean T' and AH' for that two week period for all epidemics across all cities within the study period. There were no time periods with statistically significantly ($p < 0.05$; Non-parametric bootstrapping) reductions in mean T' or AH' from the 31-year average.

Even if anomalous fluctuations in temperature and humidity do not necessarily affect epidemic onset, climatic factors could have an impact on virus transmission[221] and overall

epidemic size: for example, influenza mortality in New York Metropolitan Area was shown to be negatively associated temperature and humidity[62]. Overall epidemic incidence should depend strongly on the initial exponential growth phase of the epidemic, where transmission may be facilitated by favourable climatic conditions. I therefore investigated the impact of mean temperature and mean absolute humidity during each epidemic, as well as just the period from epidemic onset to the peak, on that epidemic's size. For both time periods considered, epidemic incidence was not associated with mean absolute humidity (Fig. B.3). I found that epidemic incidence was weakly negatively associated with the mean temperature during the epidemic and the period from start to the peak, but this relationship appears to be primarily driven by two instances, where small epidemics occurred during the early and warmer part of the season; on balance, the highly variable epidemic sizes observed over a range of climatic conditions, suggests that climatic factors have limited and noisy effects (Fig. B.3).

A recent study by Geoghegan et al.[92] estimated epidemic onset timings for influenza A virus epidemics in Australian postcodes for the seasons from 2007 to 2016. Despite the lack of subtype level resolution, their data set is substantially larger (450,000 entries) than the one used here and offers an opportunity to compare findings. I repeated the anomalous temperature and absolute humidity analyses on the Geoghegan et al.[92] dataset. As with the original dataset, there were no consistent statistically significant relationships between climate anomalies and epidemic onset (Appendix B.1, Tables B.3–B.4 and Figs. B.4–B.5).

Other climatic factors have been proposed as drivers of influenza dynamics, notably relative humidity and rainfall[162,240]. I repeated the above analyses for relative humidity and rainfall. There were some city-level associations but no consistent pattern and no pattern when aggregating across cities. Epidemic onset was not associated with statistically significant fluctuations in anomalous relative humidity and rainfall.

4.3.3 Effect of antigenic change

I next examined the effect of antigenic evolution on epidemic dynamics. Between 2000 and 2015, 7 A/H3, 3 A/H1sea, 1 A/H1pdm09, 3 B/Vic and 5 B/Yam virus antigenic variants circulated in Australia. All A/H1pdm09 virus epidemics from 2009 to 2015 were excluded for this set of analyses for two reasons. First, I could not accurately estimate the size of the 2009 pandemic. Second, there was no subsequent, detectable antigenic change observed for A/H1pdm09 viruses during the study period. I normalised epidemic sizes (see Methods) to enable comparisons between cities. Stratifying by subtype/lineage, I compared the size of the first epidemic caused by an antigenic variant against the sizes of epidemics of the same antigenic variant in subsequent years (Wilcoxon two-sample test; Fig. 4.3). Contrary to the

predictions of previous theoretical studies [130,131], newly emerged antigenic variants caused epidemics both larger and smaller than city-specific mean epidemic sizes and there was no evidence of a consistent effect of antigenic change on epidemic size.

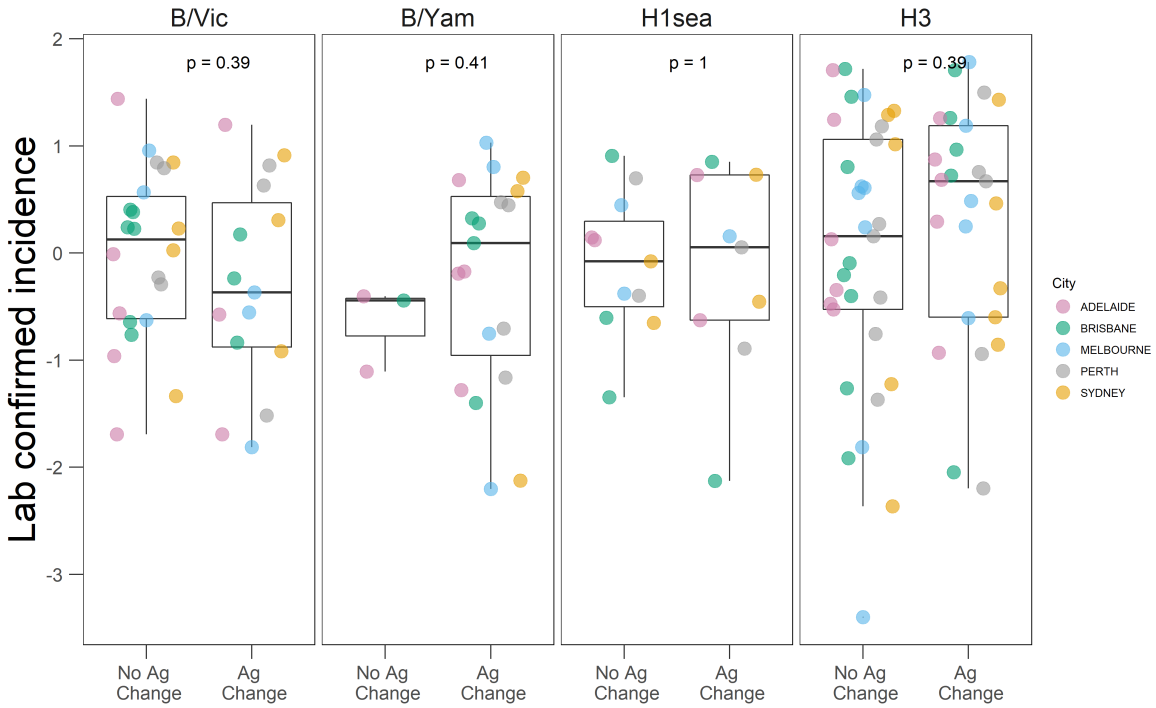


Fig. 4.3 Effect of antigenic change on epidemic incidence. Epidemic incidence was compared between seasons associated with and without the epidemic level circulation of a new major antigenic variant. Within each subtype, incidence for individual epidemics were log transformed and subtracted by the city-specific mean of log incidence, to allow for comparison between cities. p values are from Wilcoxon two sample tests (n = 37, 26, 22 and 63 for B/Vic, B/Yam, A/H1sea and A/H3 respectively). Each point corresponds to one epidemic in a city and the box plots show the median, first and third quartile of the transformed values, and range.

I also compared the timing of the first epidemic caused by an antigenic variant against the timings of subsequent epidemics (Fig. B.6) to test the hypothesis that new variants cause earlier epidemics. The range of onset timings was very broad, with epidemics starting from very early to late into the season, and there were no statistically significant differences in epidemic onset timing between new and extant variant epidemics.

To investigate the impact of antigenic change on the spatio-temporal synchrony of epidemics, I examined the timing of epidemic activity across cities for years when a new major antigenic variant circulated in all five cities. New antigenic variants often failed to initiate epidemics across all five cities in a given year. I compared the synchrony of epidemics

(defined as the reciprocal of the variance in epidemic onset timings) in the season in which an antigenic variant first emerges to the synchrony in subsequent seasons. There were no statistically significant differences in epidemic synchrony associated with antigenic novelty (Fig. B.7).

To check the robustness of this result, I repeated these analyses using estimated onset timings from Geoghegan et al.[92]. There was again no discernible effect of antigenic change on the timing or synchrony of epidemics (Appendix B.1).

4.3.4 Effect of prior immunity

After an antigenic variant causes an epidemic in a city for the first time, the accumulated population immunity to that variant should lead to smaller subsequent epidemics and eventually render further epidemics of that variant less likely. For each epidemic caused by a given antigenic variant, I investigated the relationship between that epidemic's size and the cumulative number of cases caused by that antigenic variant in preceding seasons. To account for differences in population size and surveillance intensity among cities, I normalised epidemic and cumulative case counts by the city-specific mean epidemic size. Antigenic variants that emerged prior to the start of the study period, such as A/Moscow/10/99 (A/H3) and A/New Caledonia/20/99 (A/H1sea) and all A/H1pdm09 epidemics from 2009 to 2015 were excluded from this analysis, since it was not possible to calculate a cumulative case counts for them. Specific B/Yam antigenic variants rarely caused more than one epidemic in a given city but specific antigenic variants of A/H3 and B/Vic viruses caused repeated epidemics in the same city. For A/H3 and B/Vic viruses, epidemic size and cumulative prior incidence were not correlated (Pearson's correlation test; Fig. 4.4).

The accumulation of population immunity should also reduce the probability of successful epidemic initiation, making epidemics, regardless of size, less likely to start after an antigenic variant has already caused an epidemic in that city. For B/Vic and A/H1sea viruses, binary logistic regression showed non-significant associations between the cumulative incidence over prior seasons and the probability of successful epidemic initiation (all $OR < 1$; all $p > 0.05$, Fig. B.8 and Table B.5). This partially resulted from the small number of A/H1sea epidemics during the study period, most of which were caused by newly emerged antigenic variants. However, B/Yam and H3 viruses showed significant negative relationships between cumulative prior incidence and epidemic probability suggesting that prior incidence may have a substantial impact on the probability of successful epidemic initiation.

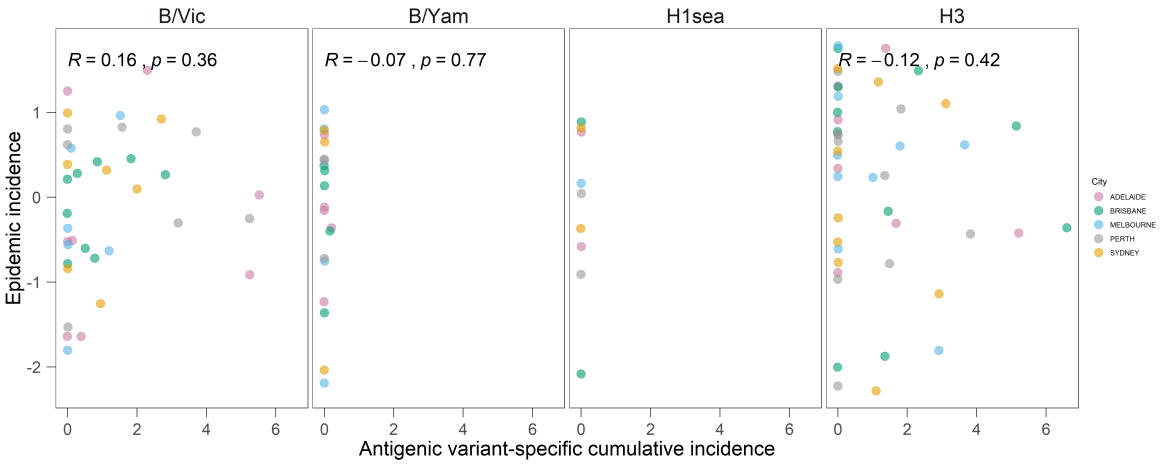


Fig. 4.4 Effect of prior immunity on epidemic incidence. Within each subtype, incidence for individual epidemics were log transformed and subtracted by the city-specific mean of log incidence, to allow for comparison between cities. Antigenic variant-specific cumulative incidence was measured relative to the city-specific mean epidemic size, where 1 is equivalent to the mean epidemic incidence. r and p values are from Pearson's correlation tests (n = 37, 20, 9 and 45 for B/Vic, B/Yam, A/H1sea and A/H3 respectively). Antigenic variants of B/Yam rarely initiated multiple epidemics during the study period and it was not possible to calculate a correlation coefficient for A/H1sea because the one new antigenic variant to emerge during the study period caused only a single epidemic per city.

4.3.5 Aggregating across subtypes

There may be subtype/lineage specific differences in the effect of antigenic change and prior immunity. Notably, B/Yam antigenic variants typically cause only one epidemic per city. I repeated these analyses with epidemics aggregated together, across all subtypes and cities to increase statistical power (see the project Github repository for the analyses and code). As before, there were no statistically significant differences in the magnitude of epidemics between the first and subsequent epidemics of an antigenic variant, nor any association between epidemic size and the cumulative incidence over prior seasons. Binary logistic regression showed that the probability of successful epidemic initiation may be moderately reduced by the cumulative incidence over prior seasons. The findings were robust to the method of normalisation used to allow for comparison between cities and subtypes/lineages (see Section 4.2).

4.3.6 Effect of competition among subtypes

Competition among virus subtypes for hosts should create a first-mover advantage for the first subtype to sustain above-baseline epidemic activity in a city in a given season.

Subsequent epidemics of other subtypes within that same season would be reduced in size. I considered two measures of this kind of inter-subtypic interference: the cumulative amount of epidemic activity prior to the onset of a subtype's epidemic and the lag between the focal epidemic and the season's earliest epidemic. To allow for comparisons across cities and subtypes, I normalised log epidemic case counts by subtracting off the city- and subtype-specific mean log epidemic case count. There was a strongly negative and statistically significant correlation between prior epidemic activity and epidemic size (Pearson's correlation test, $r = -0.420$; $p = 8.7e-5$; Fig. 4.5). An important caveat is that seasonality in the transmission rate could result in epidemics that start later in a season being smaller than those that started earlier regardless of intersubtypic competition.

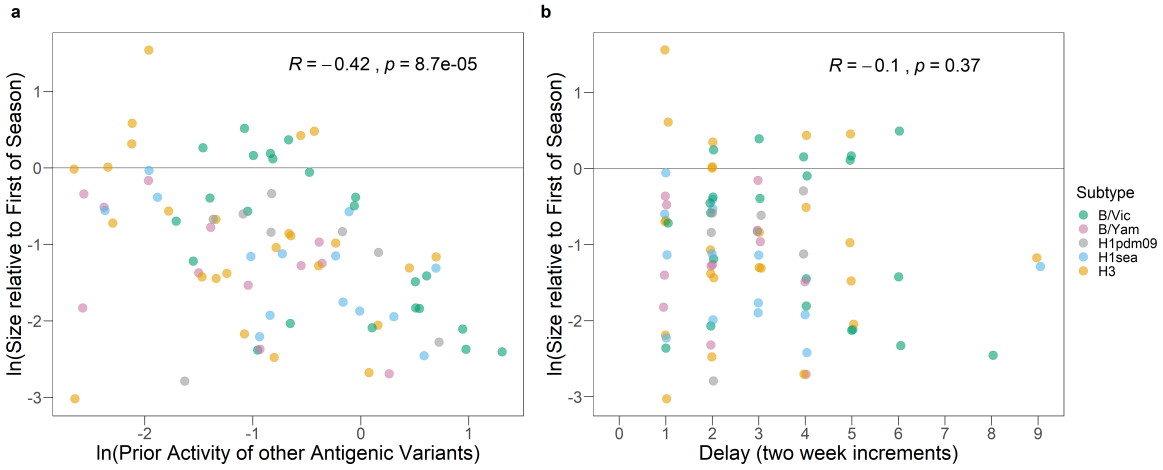


Fig. 4.5 Effect of competition among subtypes on epidemic incidence. The relationship between the size of an epidemic and (a) the amount of prior activity of all other antigenic variants and subtypes and (b) the delay in epidemic onset. The size of each epidemic, relative to the earliest epidemic of that season, was log transformed: the horizontal line at 0 denotes that the size of an epidemic is equal to that of the earliest epidemic of that season and city. In panel (a), prior activity by other subtypes within the same season was measured relative to the city-specific mean epidemic size. In panel (b), delay in epidemic onset was measured relative to the onset timing of the earliest epidemic of that season. r and p values are from Pearson's correlation tests (n = 82).

4.3.7 Joint contributions of climatic and virological factors

Whilst the magnitude of the effects of the climatological and virological factors may be individually subtle, it could be the case that they are only able to effect observable changes on the magnitude and timing of epidemics when acting in concert, or that large effects in opposing directions mask each other. I used a Bayesian multi-level regression model to

identify which putative predictor variables affected epidemic incidence and estimate posterior distributions for their effects on epidemic size. The model included the following variables: antigenic change, cumulative prior cases of the antigenic variant, mean absolute humidity during the epidemic, activity by other subtypes earlier in the season, epidemic start date, and rainfall during the epidemic. Mean temperature during the epidemic was omitted as a predictor, since it was highly collinear with absolute humidity; analyses were subsequently repeated using mean temperature and omitting absolute humidity with no substantial changes in overall results.

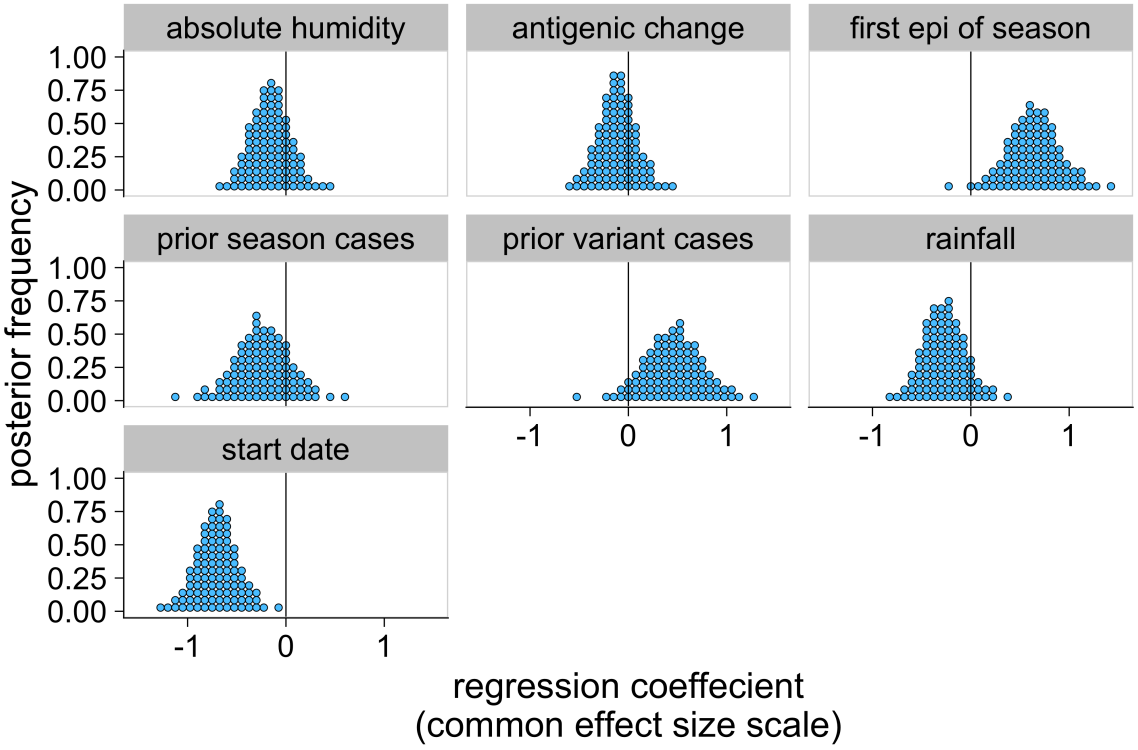


Fig. 4.6 Joint contributions of climatic and virological factors on epidemic incidence. The mean estimated effects across all subtypes were estimated using the Bayesian multilevel model. Predictors were mean-centred and scaled so effects sizes are shown on a common scale.

The model suggested that epidemics that were the first of the season or had early start dates should be modestly larger (Fig. 4.6). Start date had the largest estimated effect and the clearest posterior support for a non-trivial effect size. Posterior modes for the mean effects of antigenic change and absolute humidity across subtypes were near zero (Fig. 4.6), with tight credible intervals ((95% cred. intervals: (-0.56, 0.27) for absolute humidity, (-0.50, 0.30) for antigenic change. Prior cases of an old variant given no antigenic change (95% cred. int.

(-0.26, 0.85)), prior cases of all variants for non-first epidemics (95% cred. int. (-0.84, 0.33)), and rainfall during the epidemic (95% cred. int. (-0.68, 0.20)) also showed no strongly discernible effects, though with less posterior certainty. The model could not explain much of the variation in the data: the median estimated standard deviation of epidemic size about the expected size is 0.77 (95% cred. interval (0.67, 0.90). Since exp(0.77) is approximately 2.15, this implies that it is not unusual to see epidemics half or twice the expected incidence. The model estimated that effects were very similar across subtypes (Fig. B.14, median estimated SDs for the distribution of subtype-specific effect sizes about the overall mean effect size near zero, Fig. B.15). Only the effect of whether an epidemic was first of the season showed meaningful heterogeneity: the model estimates that it is somewhat weaker for B/Vic than for other subtypes (Fig. B.14).

4.4 Discussion

Based on city-level analyses of a subtyped and antigenically characterised influenza virus dataset covering the five largest cities in Australia, I find that climate and antigenic novelty have limited effects on epidemic sizes. The results presented here suggest that, at least in temperate areas, epidemics are governed by factors other than host immunity at local scales, where global fitness advantages for new antigenic variants may not be realised. Conversely, competition for hosts among influenza virus types and subtypes have strong effects on local dynamics. The first virus subtype to establish above baseline epidemic activity in a city and season typically dominates.

A recent study of fine-scale influenza epidemiology in Australia[92] showed there was substantial heterogeneity among Australian cities in the activity of influenza A and B viruses. My subtyped and antigenically characterised dataset allowed us to confirm that further heterogeneity exists at the level of antigenic variants. In particular, specific antigenic variants often cause large epidemics in some cities while not causing detectable activity in others.

While prior studies found that the onset of epidemics in the United States and France were preceded by a two week period of anomalously low absolute humidity[219,222], I found no evidence for climatic effects when aggregating across the five Australian cities. Anomalous fluctuations in temperature and absolute humidity were sometimes positive, sometimes negative but on average approximately zero. Importantly, the overall effect size reported by Shaman et al.[222], after aggregating across all 48 contiguous States of the USA, was very small (with mean AH' being approximately $-0.25 kgkg^{-1}$ or $-0.21 gm^{-3}$, compared against $0 gm^{-3}$, the mean of the null distribution of historic wintertime values). 55-60% of epidemics

were preceded by negative AH' values: a moderate increase upon the null hypothesis being a baseline of 50%.

Shaman et al.[222] also found regional differences in the associations between fluctuations in absolute humidity and epidemic onset. Strong associations were found in the Southeastern US but not in Western states. In Australia, there does not appear to be an aggregate effect at the country level and there were no consistent patterns at the level of individuals cities (Fig. B.2 and Table B.2). The small effect sizes and lack of consistency in climatic patterns across regions and cities in the USA and Australia may reflect the fact that climatic factors alone are unlikely to account for the differences in the patterns of influenza seasonality between temperate and tropical regions[239].

Seasonal epidemic waves in the US appear to begin in the Southern states, which have warmer and more humid climates[39,40], casting some doubt on the role of low humidity as a trigger for influenza epidemics. Rather than acting as specific triggers, it is plausible that climatic factors are acting on longer time scales than the anomalous fluctuations reported by Shaman et al.[222] to more generally enhance transmission and increase incidence[64]. However, in Australia, epidemic size does not appear to be strongly associated with the mean temperature or absolute humidity over the epidemic period.

Given the interest in influenza virus as a model system for phylodynamics of a pathogen that consist of multiple co-existing antigenic variants[141], there is interest in understanding how competition between these related variants, typified by cross-immunity, shapes epidemiological dynamics. Studies have hypothesised that antigenic change should result in larger[24,130,270,274] and earlier[29] local epidemics, which exhibit greater spatio-temporal synchrony at the national level[39,92,102,213,256]. The sequential replacement of old antigenic variants by new ones is indicative that antigenicity and population immunity are important for the global level phylodynamics of influenza viruses. In contrast, at the local level, I find for A/H3 and B/Vic viruses that neither antigenic change nor the accumulation of antigenic variant specific immunity are strong drivers of epidemic size, though accumulating variant-specific immunity may moderately reduce the probability of successful epidemic initiation.

It is striking that individual antigenic variants of A/H3 and B/Vic viruses are capable of re-invading the same city multiple times over consecutive years, despite a lack of substantial antigenic change. A/H1pdm09 viruses had previously been shown to cause repeat epidemics without antigenic change[20,118], but this work establishes that this occurs for multiple types and subtypes of human influenza. One possible explanation for the lack of evidence for the year-on-year depletion of susceptible hosts is that influenza virus infection often fails to confer strain-specific and effective immunity. In some individuals, antigenic seniority

and existing immunity against previously encountered antigenic variants may suppress novel strain-specific antibody responses, leading to only modest specific protection against reinfection[140,151]. Similarly, there is evidence from vaccine trials[191] and infection challenge studies[100] to suggest that children, particularly those below ten years of age[172], require multiple exposures to develop protective immunity.

There may also be substantial and previously unaccounted heterogeneity in individual susceptibility towards the same virus strain. The notion that population level strain-specific immunity to influenza viruses is monolithic may be an artefact of the single-infection ferret models typically used to estimate antibody-mediated protection. In humans, there is substantial individual-to-individual variation in the antigenicity of amino acid escape mutations for influenza haemagglutinin[148]. Such heterogeneity between individuals stems from their varied exposure histories to different influenza viruses. Unfortunately, the corresponding age records for this data set were too incomplete to allow us to study age-specific heterogeneities in demographics and attack rates between cities and whether such patterns change over seasons.

Spatial and social connectivity structures among hosts in a city may also limit the spread of epidemics. Heterogeneous contact patterns between hosts can have a substantial impact on resulting epidemiological dynamics[53,55]. Epidemics may be inherently frail processes: relatively minor human behavioural or environmental perturbations could prematurely terminate epidemics before they exhaust the pool of susceptible hosts, preserving a substantial number of susceptibles and permitting subsequent epidemics of the same antigenic variant.

While this dataset is substantially smaller (>450,000 vs 18,250 cases) than the one analysed by Geoghegan et al.[92] and is thus more likely to be affected by noise in epidemic and surveillance processes, the differences between my findings and theirs highlight the importance of subtyping and antigenic characterisation, particularly for drawing conclusions about the effects of antigenic change. Geoghegan et al.[92] had cautiously suggested, given only virus type data, that the 2009, 2012 and 2014 influenza A virus epidemics in Australia exhibited greater spatio-temporal synchrony potentially due to the emergence of the novel A/H1pdm09 subtype in 2009 and novel A/H3 antigenic variants in 2012 and 2014. However, with further subtype resolution and antigenic characterisation, I find that the majority of influenza A activity in Adelaide and Melbourne in 2014 was attributable to A/H1pdm09, rather than the (antigenically novel) A/H3; in fact, there was no above baseline A/H3 activity in Perth. The fact that different virus subtypes caused these apparently synchronous epidemics implies that the epidemic synchrony described by Geoghegan et al.[92] was not due to the antigenic evolution or regional spread of a single virus strain.

Apart from competition between antigenic variants, previous epidemiological studies have hypothesised the existence of heterosubtypic competition where prior infection by a virus of one subtype is negatively associated with subsequent infection by a virus of another subtype[51,232]. In agreement with a previous US study of national level ILI activity augmented with limited virus subtyping[99], I also find evidence for a first-mover advantage and competition to infect hosts within a city, where the subtype or type that initiates above baseline levels of activity first is most likely to have the largest epidemic of that season.

There are multiple caveats to this study that merit explicit consideration. The most important ones derive from the use of passive surveillance data that might not accurately reflect true underlying influenza virus activity. For example, surveillance intensity could plausibly vary between cities and years. While variation in surveillance efforts is evident among cities, there was no evidence of systematic increases or decreases in the number of laboratory-confirmed cases or changes to surveillance practices within each city during the study period. Despite this, the longer duration of epidemics recorded after 2009 could be indicative of enhanced surveillance in the post-pandemic era: to mitigate this possibility, I repeated the analyses on the effect of antigenic change on epidemic size, splitting between pre- and post-pandemic eras and epidemic sizes normalising by their respective era-specific means. In either era, there were no consistent effect of antigenic change on epidemic size, with the caveat that splitting across eras reduced the number of observations in each era and thus the statistical power (Fig. B.18).

Intensity of surveillance could also vary over the course of an epidemic. For example, sentinel physicians could become more likely to submit samples for further testing as an epidemic unfolds or conversely, testing could prematurely cease as facilities become overwhelmed with samples. Despite being unable to definitively rule out the former scenario, the latter is unlikely to affect the data. If reporting ceased after a certain number of samples had been tested, the distribution of epidemic sizes would be truncated and each epidemic would be unlikely to have an exponentially declining tail. No such patterns exist in the data.

The intensity of surveillance could also potentially vary across subtypes and lineages. The mean age of infection for A/H3 is greater than influenza B[25] viruses and healthcare seeking behaviour may differ between adults, parents with children and children. Furthermore, it is commonly thought that A/H3 virus infections result in more severe clinical presentations and greater risk of mortality[243] than influenza B viruses, potentially resulting in differences in the likelihood of detection by a sentinel health practitioner, though this may not be the case[34].

Another important caveat is that while I were able to include antigenic data in this study, these data were all derived from HI assays. HI assays do not measure virus antigenic changes that occur away from the receptor binding site and thus likely represent an incomplete picture

of antigenic change. Reference viruses and sera used in the haemagglutination inhibition assays can also impact the interpretation of the assay readout and the HI data used in this study were therefore treated with caution (see Section 4.2).

In this study, I attempted to identify associations between population susceptibility and epidemic incidence. Accurately quantifying the former is a complex challenge so cumulative antigenic variant-specific epidemic incidence was used as a proxy but that itself is subject to the limitations listed above. Besides natural infection, immunity can also be derived from vaccination, the contribution and effectiveness of which could not be determined due to a lack of temporally and geographically complete vaccination records over the study period. Regardless, I hypothesise that the impact of seasonal vaccination would be limited, particularly in the context of Australia, given the low uptake of vaccination historically[21]. Crucially, the uptake by children, who are important in driving local community transmission, has often below 10% during the study period[238].

While the Bayesian multilevel model estimated negligible effects on epidemic size stemming from climatic factors and prior cases attributed to the same antigenic variant, the estimated credible intervals were not tight enough to rule out these effects conclusively (Fig. 4.6). However, my results suggest that climatic and antigenic factors are unlikely to be strong drivers of local influenza epidemiological dynamics. Indeed, the effects of these specific factors are dwarfed in magnitude by more generic epidemiological drivers: seasonality not directly captured by climate (measured as the fortnight in the year when the onset of the epidemic occurs) and competition for hosts among subtypes (measured by whether an epidemic is the first of the season) (Fig. 4.6). I also find that even with all generic and specific factors considered, precise predictions of epidemic size remain difficult because of substantial noise in the local epidemic process.

The Bayesian multilevel model for epidemic size avoids explicitly modelling underlying transmission processes and may fail to fully capture the nature of the relationship (linear vs non-linear) between transmission rates/R_0 and the total cases in an epidemic. However, based on previous virus transmissibility studies[221], if climatic factors are strong drivers of epidemiological dynamics, one would expect the climatic variabilities observed in Australia to have a substantial impact on transmission rates and produce detectable differences in epidemic size, but this is not the case.

Climatic drivers of seasonality and homosubtypic competition between virus antigenic variants are thought to be strong drivers of seasonal influenza epidemiology, but seasonal influenza virus epidemiological dynamics in major Australian cities appear to be more substantially shaped by other factors; particularly the establishment of sustained virus transmission activity and subsequent competition among virus types and subtypes. This

implies that the time horizon for meaningful forecasting of epidemic subtype composition is very short (days-to-weeks) and forecasting efforts aimed at longer term predictions will require further insights into the dynamics of virus introduction and epidemic establishment and into the accumulation of population immunity to seasonal influenza viruses.

These results form the impetus behind Chapters 5 and 6, where I examine how the network structure of populations and immunity may limit the extent of epidemic spread, providing a potential explanation for recurrent epidemics by the same antigenic variant over consecutive seasons.

Chapter 5

The effects of host contact structure and cross-immunity on the recurrence of epidemics

5.1 Introduction

Previous attempts at quantifying the effects of antigenic evolution for seasonal influenza viruses have often relied upon mean-field, *SIR* type mathematical models [130,131] or relatively simple metapopulation representations of population structure [29,278]. Once an epidemic has concluded, the disease-free equilibrium state prevents pathogen reinvasion, due to the critical depletion of susceptibles [112,113,249]. In a closed system without demography, recurrent epidemics can only occur if this susceptible fraction is sufficiently restored, through a combination of phenotypic changes in virus antigenicity and loss of host immunity through waning. Of course, it is likely that epidemics fail to run to their theoretical limits and reach critical depletion of the susceptible pool in the first place, leaving pockets of susceptible individuals and reducing the amount of restoration required.

My analyses of the Australian data set in Chapter 4 suggests that reinvansion is occurring in the absence of major antigenic change, as quantified using HI assays [231]. In lieu of phenotypic changes in virus antigenicity, the restoration of population susceptibility could be achieved by the loss of immunity within hosts: rather than long-lasting homotypic protection [94], the duration of protection conferred by natural infection could be shorter than previously anticipated [176].

Another potential explanation is that population structure limits epidemic spread and consequently, the build-up of immunity at a local scale. A similar hypothesis [124] has been

proposed as the basis for the recurrent epidemics in 2010 and 2011 by A/H1pdm09, which did not exhibit marked changes in antigenicity[20,118]. For seasonal influenza viruses, if human population structure limits the build-up of immunity reinvasion could become possible without necessitating a rapid turnover in population immunity.

More broadly speaking, the impact of host contact structure has been of great theoretical interest: predictions made by network models differ substantially from those of classical random mixing models[67,157,177,197], since interactions are limited to a fixed set of contacts, the number of which differ between individuals. The degree distribution of the network shapes the patterns of disease spread[53,192] and the geometry of immunity imparted upon the residual network[78]. This in turn determines the vulnerability of the network towards subsequent reinvasion[19,78] and competition between related strains[125,153,194].

Indeed, the patterns of influenza virus spread may be intimately associated with underlying host contact structures and their age-specific heterogeneities. Children have been implicated by many epidemiological studies to be important for community transmission, due to the frequency and intensity of their contacts in schools. The apparent shift in age-specific attack rates towards adults over subsequent A/H1pdm09 epidemics may have resulted from the preferential depletion and accumulation of immunity within the highly connected children[19]. These patterns highlight how population structure can be potentially leveraged for more targeted and efficacious interventions[184].

It is evident that model specification is critical since different assumptions on how individuals mix within a population can lead to drastically divergent model dynamics[199,248]. For many predominantly childhood diseases, such as measles[103], pertussis[210] and varicella[174], social contact data can be used to successfully parameterise mathematical models and parsimoniously recapitulate age-related patterns of transmission and disease[152]. Such minimal explanations could account for the importance of children in the community transmission of seasonal influenza but fails to account for the substantial amount of cross-immunity that is widely distributed across the population.

In this chapter, I investigate how the effects of population structure and its interactions with host immunity can facilitate the recurrent epidemics of seasonal influenza. Whilst previous studies into network models and the dynamics of pandemic influenza have provided insight into the important implications of population structure[19,124], their model frameworks need to be adapted for the specific context of seasonal influenza. Transmission across the network is likely to be affected by age-related correlations between host contact patterns and individual immunity. Rather than using a configuration model approach and various prescribed degree distributions[119,124], I draw upon data from empirical contact surveys and construct networks that better characterise the structure of real populations. Furthermore,

I add multiple immune classes to reflect the complex exposure histories over the course of an individual's lifespan and account for the geometry of cross-immunity. Using this more sophisticated model framework, I analyse the structural evolution of the network over two sequential seasons.

5.2 Method

5.2.1 Contact network model

Within a network model, individuals and their epidemiologically-relevant contacts are represented by nodes and edges respectively. Rather than using prescribed degree distributions to describe the number of contacts that an individual has, I used an exponential random graph model (ERGM) to produce a more realistic caricature of population structure. ERGMs define a distribution of networks by considering the likelihood of an edge between a given pair of nodes, based on their individual attributes. These models can be fitted using empirical data to reflect salient features of population structure, such as the number of edges between individuals of the same household, by summarising egocentric social survey data, where "egos" report the number and characteristics of their contacts with "alters".

To construct my synthetic networks, I used the Australian census data for Sydney and randomly sampled 1000 households, each of which could be a family unit or a group of cohabitating adults. Individuals were randomly assigned ages and age-appropriate, extra-household activities (day-care centre for 0 - 5 years, school for 6 - 18 years and workplaces for >18 years of age), taking into account the attendance rates, class sizes[11], employment rates and workplace sizes. Unfortunately, I found no contact survey studies for Australia; I derived the expected number of edges within the various social contexts used the POLYMOD study[186] data set, which contains detailed diaries of daily social contacts and their locations for over 7000 individuals across 8 European countries.

For a disease that affects both children and adults, it is clear that households form a key component of epidemic spread. From various survey studies[27,96,186,202], it is known that within a household, the probabilities differ between homophilous (child-child and adult-adult) and heterophilous (child-adult) contacts. These probabilities are further dependent on the size of the household, with network density decreasing as size increases[96]. Additionally, the odds of physical contact within a household decreases with the age of its members, particularly the contacts between father and child[96,202]. However, the additional attributes of household size, gender and ages were not identifiable: there was insufficient coverage of individuals from each possible attribute combination. Furthermore, higher-order network statistics, such

as stars and clustering coefficients, could not be recapitulated since egocentric sampling in the POLYMOD study[186] did not exhaustively collect information for all other individuals within the household of the respondent. Similarly, with regards to extra-household contacts, it was not possible to take into account the effects that the sizes of day-care, schools and workplaces can have on contact probabilities.

The model incorporating "*social structure*" was fitted using the ERGM package[32,250] in R, with model terms (Table C.1) accounting for the number of child-child, adult-adult and child-adult edges observed within households, day-care, schools and workplaces. Uniform homophily for all these attributes is assumed, whereby the propensity for within-group ties between individuals is the same across all groups. In other words, within each and every school, the probability that two children nodes are connected by an edge is the same. Since it was not possible to further specify network structure beyond basic descriptive statistics, an additional degree term was used to forcibly introduce a positive skew into the degree distribution, so that it better recapitulates the empirical observation that a small fraction of individuals have many contacts, which is particularly prominent for children.

To evaluate the effects of population structure and contact heterogeneity in network models, I also generated two additional ERGMs to use as baselines for comparison (Fig. C.1). 1) A "*no structure*" null model that ignores all nodal attributes and has randomly distributed edges; 2) a "*household*" model that has assortative mixing between individuals of the same household and randomly distributed edges between individuals of different households. To enable fair comparison, all three ERGMs have the same overall number of edges and network densities.

For each realisation, a simulated network is drawn from the ERGM distribution and assumed to remain static throughout the epidemic. This is primarily done in the interests of computational tractability but is also justified by an inability to parameterise a dynamic model, given the lack of suitable cross-sectional survey data across time. It is reasonable to assume that membership of households, schools, etc. are unlikely to fluctuate drastically over shorter timescales, although the exact contact patterns within maybe more labile. Since I considered short timescales, demographic processes were also omitted; otherwise, the need for the properties of temporal networks to stay invariant in response to the removal and introduction of new nodes could have introduced complications.

5.2.2 Modelling host immunity and virus transmission

The epidemiological model is based upon the *SIR* framework but transmission is dictated by the underlying network structure and can only occur between connected nodes. Hosts are randomly assigned an immunity class. Fully naïve S_{naive} individuals have a susceptibility

$\sigma_{naive} = 1$. Individuals with prior exposure and partial immunity are classified (S_i), based on the most recent antigenic cluster of viruses to which they possess immunity towards. For instance, hosts with immunity to the wild-type cluster and the cluster before would be in class S_0 and S_1 respectively. The susceptibility for individuals, who have most recently encountered cluster i, towards virus phenotype from cluster j is governed by the parameter σ_{ij}, which is the strength of cross-immunity interactions between antigenic variants. σ_{ij} decreases with antigenic distance and is defined by saturating linear (Eq. (5.1a)) and multiplicative (Eq. (5.1b)) functions of the escape factor θ.

$$\sigma_{ij} = \begin{cases} \theta \cdot |i-j| & \text{if } \theta \cdot |i-j| \leq 1 \\ 1 & \text{if } \theta \cdot |i-j| > 1 \end{cases} \tag{5.1a}$$

$$\sigma_{ij} = 1 - (1-\theta)^{|i-j|} \tag{5.1b}$$

From first principles, the basic reproduction number R_0 is the expected number of secondary cases generated by a single infected individual during the infectious period, in a fully susceptible population. For a fully susceptible population, consider a newly infected node (the ego) within the network: it cannot re-infect the neighbour (the alter) who infected it. Its excess degree, or the number of alters that are susceptible to onwards transmission, is thus its degree k less one. The excess degree distribution q_k is a function of the degree distribution p_k and its first moment $\langle p_k \rangle$ [192,193] (Eq. (5.2a)). R_0 is thus the product of the probability of disease transmission along a random edge (T) and the average excess degree $\langle q_k \rangle$ [192,258,276] (Eq. (5.2b)).

$$q_k = \frac{(k+1)p_{k+1}}{\langle k \rangle} \tag{5.2a}$$

$$R_0 = T \cdot \langle q_k \rangle \tag{5.2b}$$

Analytical solutions for the average excess degree $\langle q_k \rangle$ have been derived for random bipartite networks [1,17], where the two types of nodes differ in their degree distributions or susceptibility. Multiple immune classes introduce additional asymmetry into the calculation of the effective excess degree of a node: the choice of infecting alter affects the number of remaining alters that can be subsequently infected (Fig. 5.1a,b). The probability that a particular alter is the originator of infection is not uniform: a naïve alter is more likely than a partially immune one to be the originator. Onward transmission is conditional on the ego node becoming infected, the probability of which is dependent on its immune class

(Fig. 5.1c,d). Furthermore, the degree of a node is determined by a joint distribution over the distributions for within- and extra-household contacts. The overall average effective excess degree for the network is not simply the first moment of q_e and needs to account for both the degree and susceptibility of the individual nodes.

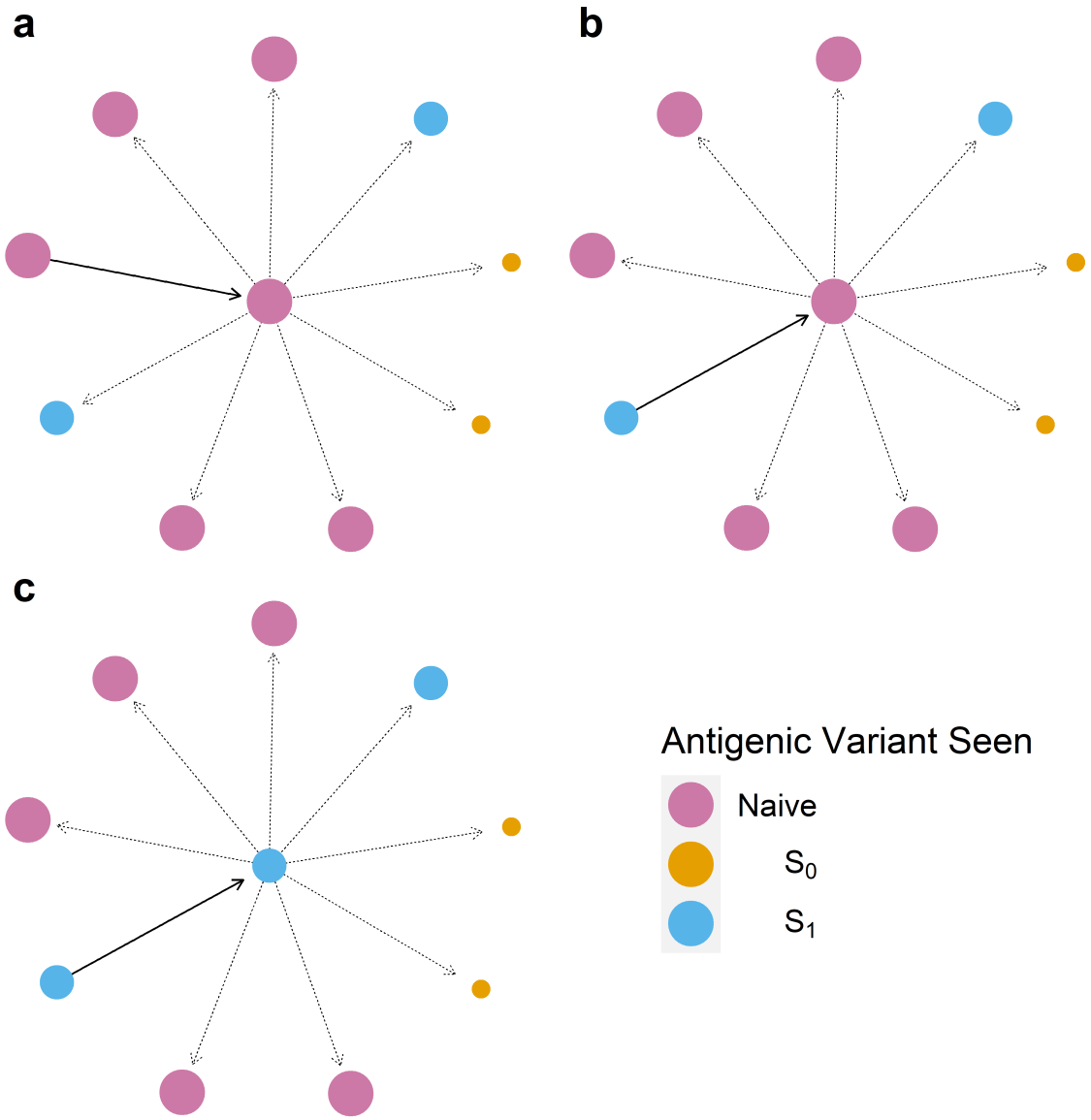


Fig. 5.1 Diagrammatic representation of the Effective Excess Degree for a node. The colours of a node denotes its immune class and hence its susceptibility towards infection by the current wild-type; the larger the node, the greater the susceptibility. In each panel, the central node (ego) is infected by one of its alters along the bolded edge. Whilst the composition of the alters do not change between panels, the remaining alters that the ego can then infect differ: (a) 4 naïve, 2 S_0 and 2 S_1; (b) 4 naïve, 2 S_0 and 1 S_1. In (c), the ego node is now S_1 and possesses partial cross-immunity, reducing the likelihood of initially becoming infected and thus the probability of its alters becoming infected.

Calculating the average effective excess degree of the network is necessary for parameterising the transmission probability and ensuring that the resultant effective reproduction number R_e and epidemiological dynamics are plausible. The dimensionality of this problem

quickly escalates, due to the number of immune classes and types of interactions: rather than adopting an analytical approach, I calculated the average effective excess degree through simulation. For each combination of epidemiological parameters, I generated 1000 distinct networks from the ERGM distributions and randomly reassigned immune classes. In order to calculate the average effective excess degree for one such realisation, consider the ego node a of immune class S_a with c alters $\{c_1, \ldots, c_b\}$ of immune classes $\{S_{b_1}, \ldots, S_{b_c}\}$. With respect to the current wild-type virus of cluster 0, the susceptibility of ego and alters are thus $\sigma_{a,0}$ and $\{\sigma_{b_1,0}, \ldots, \sigma_{b_c,0}\}$.

Accounting for partial immunity of the alters, the effective degree (ED) and effective excess degree (EED) for node a is as follows (Eqs. (5.3a)–(5.3b)).

$$ED_a = \sum_{i=1}^{c} \sigma_{b_i,0} \tag{5.3a}$$

$$EED_a = \frac{\sum_{i=1}^{c} \sigma_{b_i,0} \cdot (ED_a - \sigma_{b_i,0})}{\sum_{i=1}^{c} \sigma_{b_i,0}} \tag{5.3b}$$

The overall average effective excess degree across a network of n nodes is given by Eq. (5.4).

$$\text{Average } EED = \frac{\sum_{i=1}^{n} \sigma_{i,0} \cdot EED_i}{\sum_{i=1}^{n} \sigma_{i,0}} \tag{5.4}$$

Using the average EED values for the 1000 replicates and a given effective reproduction number R_e, the probability T that transmission occurs over the duration of the infectious period along an edge connecting an infected ego and a susceptible alter node can be calculated (Eq. (5.5a)). The infectious period γ^{-1} is assumed to be exponentially distributed. From the overall transmission probability T and the duration of the infectious period γ^{-1}, the daily probability that transmission occurs along an edge β can then be calculated (Eq. (5.5b)). As a means of validating my simulation-based approach to model parameterisation, I selected a target R_e and ran a set of simulations using my calculated β, ensuring that the R_e estimated from the early part of the epidemic trajectories matched the original target value.

$$R_e = T \cdot \overline{\text{Average } EED} \tag{5.5a}$$

$$T = 1 - (1 - \beta)^{\gamma^{-1}} \tag{5.5b}$$

For reference, Table 5.1 summarises the modelling notation used above.

Term	Meaning
q_k	Excess degree distribution (fully susceptible population)
R_0	Basic reproduction number
σ_{ij}	Cross-immunity term: susceptibility towards cluster j for individuals with immunity towards i
ED	Excess degree for a given node
EED	Effective excess degree for a given node
R_e	Effective reproduction number
T	Probability that, over the duration of the infectious period, an infectious ego will transmit to its susceptible alter
β	Daily probability that an infectious ego will transmit to its susceptible alter
γ	Daily probability of that an infectious individual recovers

Table 5.1 Notation used in network model framework

I compared two immune class distributions, which had the same overall number of individuals in each immune class. For the "*homogenous*" immune class distribution, which acts as the null model, the immune classes are distributed equally amongst the two age groups (top row of Fig. 5.2a). This is contrasted against a distribution of immune classes that differs between age groups (bottom row of Fig. 5.2a): children ($\leq$18 years of age) have no exposure towards more historic antigenic variants and are more likely than adults ($>$18 years of age) to be in the S_{naive} class. The proportions within the immune classes reflect the upper end of estimates for seasonal incidence, which is thought to vary between 5-20% [181,245,274].

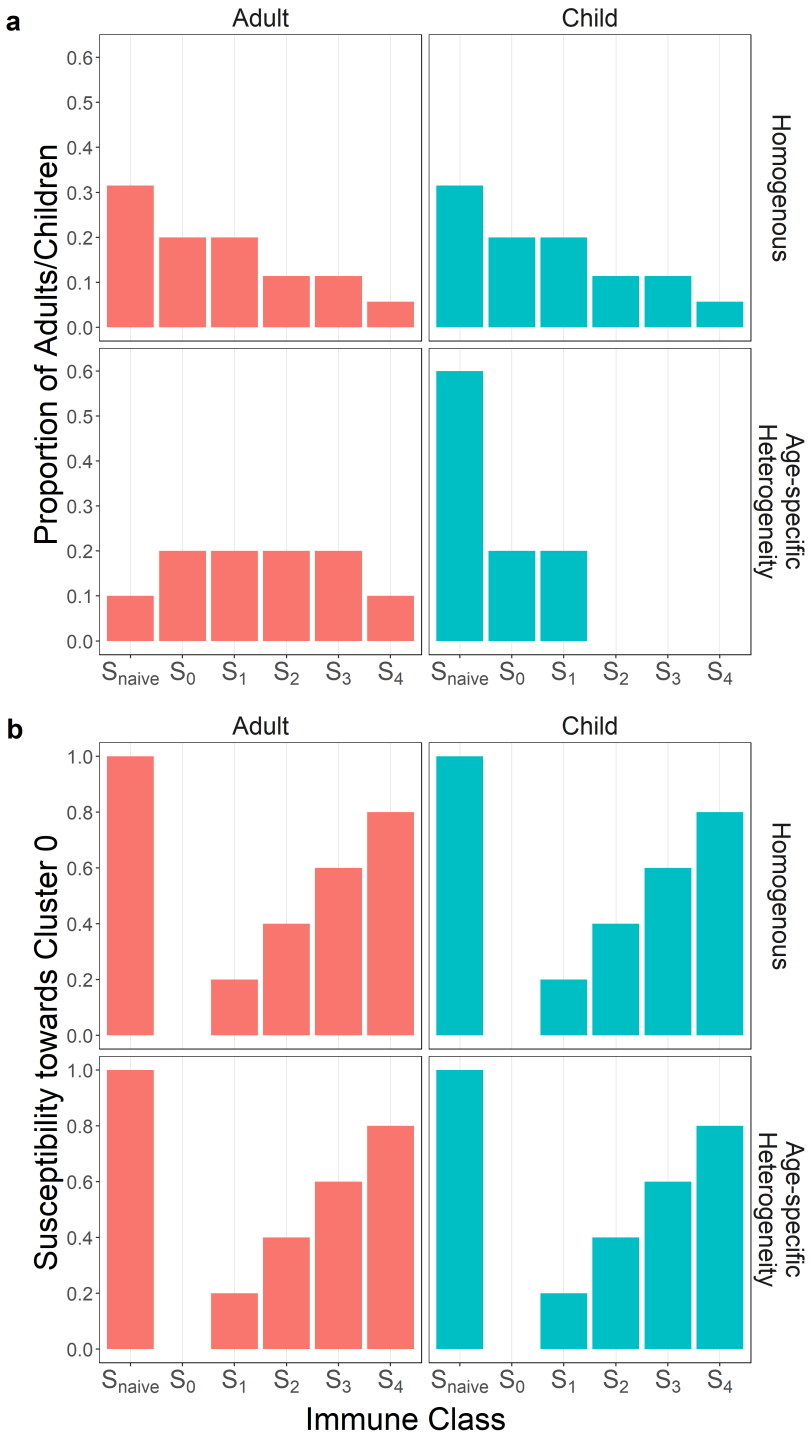


Fig. 5.2 The distribution of (a) immune classes within the population and (b) their corresponding susceptibility towards the current antigenic variant of cluster 0. In order to investigate the interactions between network structure and the geometry of cross-immunity, I compare two scenarios: the total number of individuals in each immune class are identical, but differ in their distribution between children and adults. For the susceptibility profiles shown, I used a linear model of cross-immunity with $\theta = 0.2$ (Eq. (5.1a)).

5.2.3 Simulating epidemics

Once β is calculated for the specific set of epidemiological parameters and target R_e, 1000 independent stochastic simulations were run, each with a unique network drawn from the ERGM distribution and unique distribution of immunity amongst the individuals. The epidemic was seeded by random infection of five hosts and simulated with daily timesteps.

At each timestep, infections can occur along an edge connecting infected and non-infected individuals. The outcome of such events are evaluated based on a draw from a Bernoulli distribution with probability equal to β multiplied by the susceptibility $\sigma_{i,0}$ of the at-risk individual (Eq. (5.6)).

$$p(\text{Infection along an edge}) = \beta \cdot \sigma_{i,0} \tag{5.6}$$

When an individual becomes infected at time t, their infectious period t_I is drawn from the exponential distribution $T_I \sim Exponential(\gamma^{-1})$, where γ is the recovery rate, which is assumed to be the same for both adults and children; the infected individual will recover at the end of time $t + t_I$. Recovered individuals have their immunity class updated to S_0 but cannot be re-infected within the same season. To explore the conditions necessary for reinvasion, I adopt a two-step framework, which separates the shorter timescale for epidemics from the longer timescale for changes in host susceptibility resulting from immune waning. Two sequential epidemics (*S1* and *S2*) are initiated on the networks sequentially and I investigate their structural evolution and changes. This approach is similar to that of Andreasen[6], Jaramillo et al.[124], whereby demographic processes and antigenic drift act in-between seasons to alter the immunity structure of the host population. In my framework, virological and demographic conditions are assumed to stay invariant over such short timescales; instead, I focus on the effects of host immune waning. A proportion δ of individuals possessing immunity towards the current wild-type (class S_0) are randomly selected and become re-susceptibilised (class S_{naive}), which partially restores the connectivity of the residual network.

5.2.4 Model scenarios considered

To characterise the impacts of and interactions between network structure and cross-immunity, I compared and contrasted a series of model scenarios (Table 5.2):

Network Structure	Immune Class Distribution
No structure	Homogenous
	Age-specific Heterogeneity
Household	Homogenous
	Age-specific Heterogeneity
Social structure	Homogenous
	Age-specific Heterogeneity

Table 5.2 Model scenarios considered

The ERGMs representing the different levels of population structure are made comparable by having the same overall number of edges within the network. Similarly, the number of individuals in each immune class are identical but differ in their age distributions.

5.3 Results

5.3.1 Pre-Season 1

The underlying topology and properties of a network fundamentally dictates epidemic spread. In particular, there is a propensity for outbreaks to affect highly connected nodes disproportionately[78,192]. For a population with pre-existing partial cross-immunity, I examined the effective excess degree for the various networks, which accounts for both nodal degree and immunity (Fig. 5.3). Compared to the "*no structure*" null case, the addition of "*household*" structure has limited effects on the effective excess degree (the average number of alters that are susceptible to onward transmission by any given node) of the overall network, since the majority of edges are between individuals of different households and remain randomly distributed. However, there is a modest difference between age groups, due to contact between siblings occurring more frequently than those between parent and child.

The difference in effective excess degree distribution between age groups becomes more marked upon the incorporation of "*social structure*": in addition to the differences in contact patterns within households, children make a greater number of extra-household contacts within day-care and schools than adults at their workplaces. Intuitively, this difference is further exacerbated upon the incorporation of age-specific heterogeneities in the distribution of immune classes (bottom row of Fig. 5.3), since children not only have a higher degree but now also tend to be more susceptible than adults.

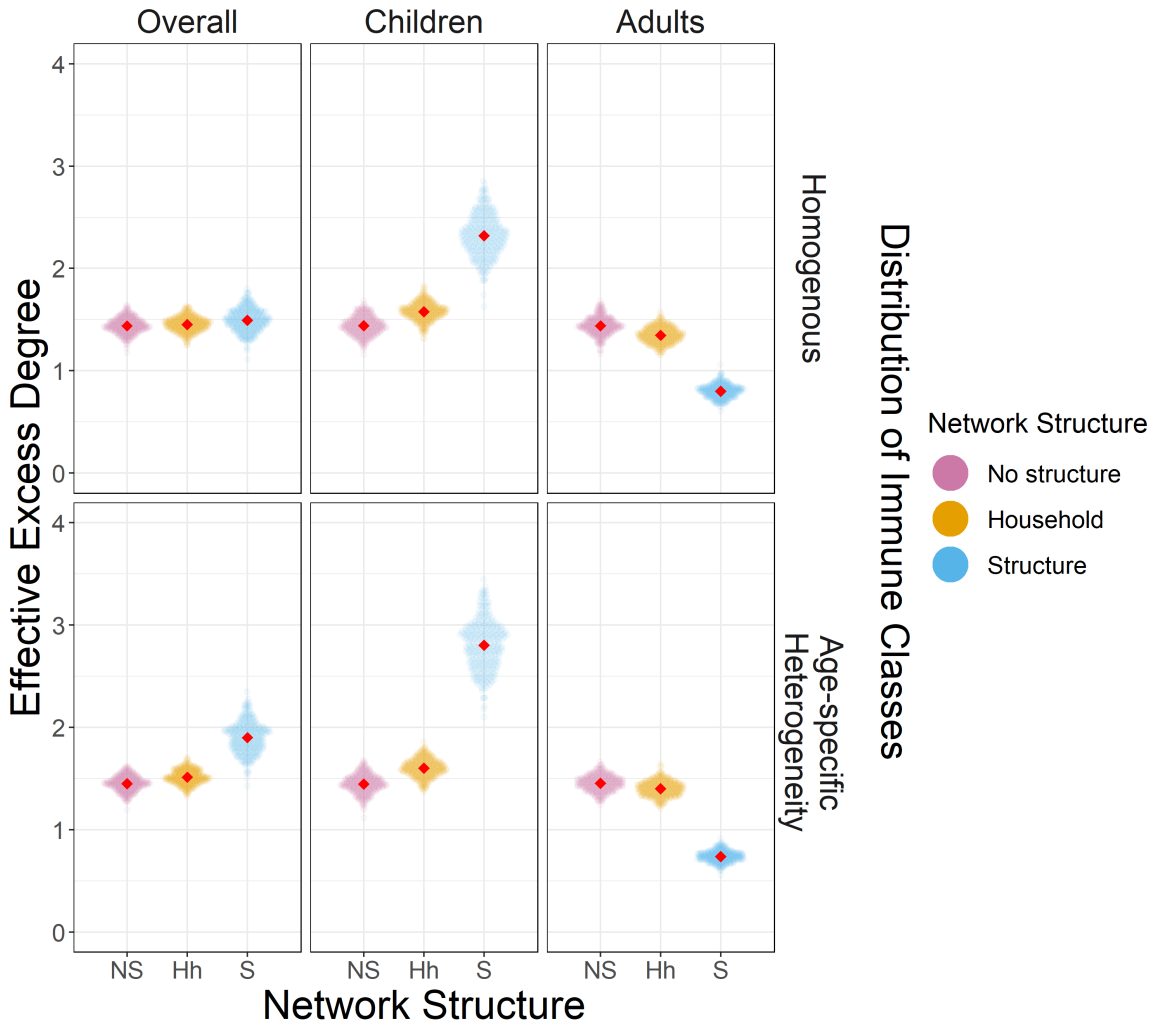


Fig. 5.3 The effect of network structure and age-specific heterogeneities in the distribution of immune classes on the average effective excess degree. A thousand simulated networks were independently drawn from their respective ERGM distributions, with immune classes randomly allocated accordingly: individual coloured points on the plot represent the average effective excess degree for each realisation. Red points denote the mean of the distributions.

5.3.2 Season 1 final size

Using my stochastic network models, I first investigated how network structure and immune topology affects the spread of the *S1* epidemic. To ensure comparability across all scenarios, R_e was kept constant at 1.2, which is typical for seasonal influenza viruses[274], by reducing the transmission probability accordingly when the underlying average effective excess degree was greater. The addition of network structure results increases the likelihood of premature stochastic termination and failure to initiate major epidemics. When epidemics

do occur, the final size is reduced: the reduction in attack rate for adults is disproportionately large when compared to that for children. This differential reduction in attack rates become more pronounced, when adults possess greater levels of immunity than children (bottom row of Fig. 5.4). Preferential infection of children changes the composition of epidemics (Fig. C.2) and is in agreement with the classical results, as discussed above (Fig. 5.3), where high connectivity nodes are particularly vulnerable.

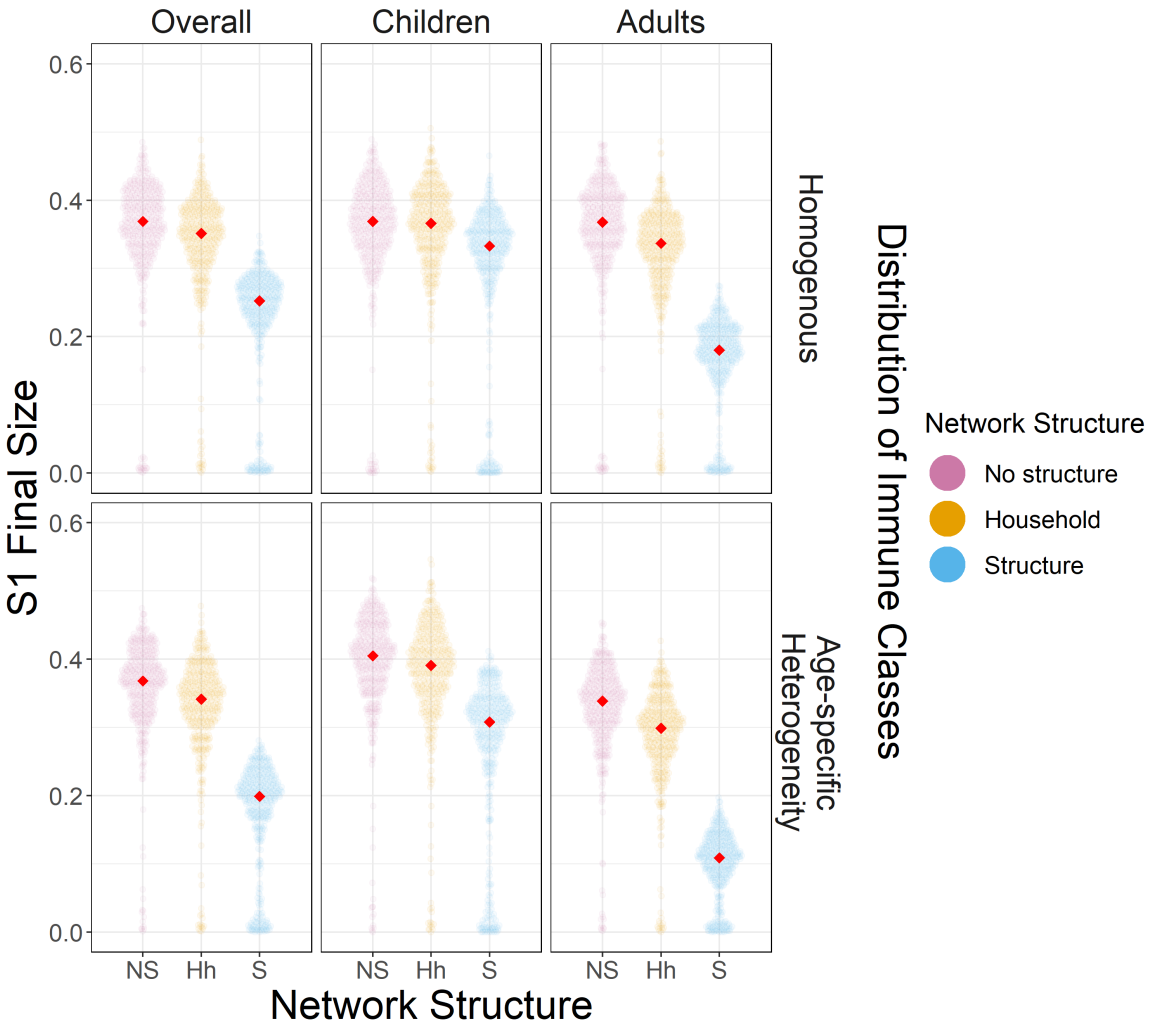


Fig. 5.4 The effect of network structure and age-specific heterogeneities in the distribution of immune classes on the final size of the Season 1 epidemic. Individual points represent the overall, child- and adult-specific attack rates from one stochastic simulation; the red point represents the median value.

5.3.3 Season 1 changes in effective excess degree

Epidemics confer immunity, particularly to nodes predisposed to infection, and thus the state of the residual network is critical to reinvasion. I therefore investigated how the effective excess degree distributions changed after *S1*. When epidemics are successfully initiated, the addition of social structure and heterogeneous immunity results in smaller epidemics. Despite infecting fewer people, the removed individuals are of greater connectivity so the residual *post-S1* network has a lower effective excess degree (Fig. 5.5a), which makes reinvasion more difficult (Fig. 5.5b). Children form a crucial conduit for transmission and consequently suffer from the greatest reductions in connectivity after the epidemic has cascaded throughout the network (Fig. C.3). To allow for comparisons between networks that differ in their underlying *pre-S1* connectedness, individual *pre-S1* effective excess degree values were log transformed and subtracted by the group mean. Similarly, the *post-S1* values were log transformed and subtracted by the *pre-S1* mean, in order to quantify the relative change.

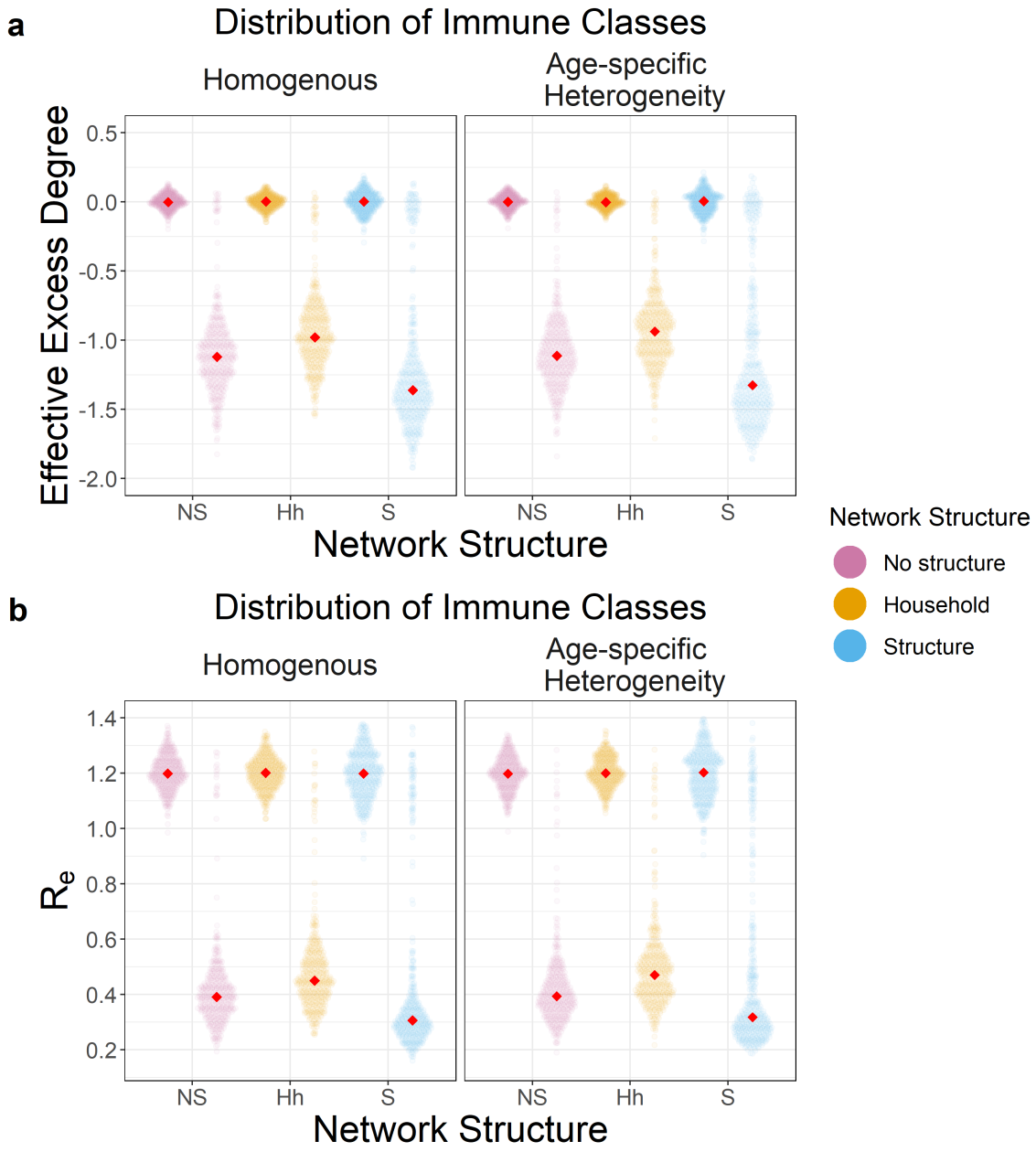


Fig. 5.5 The impact of the Season 1 epidemic on network connectivity. For each network structure and immunity combination, the left and right distributions show the original *pre-S1* and residual *post-S1* distributions for (a) average effective excess degree and (b) corresponding R_e distributions; the red point denotes the median value. To make the effective excess degree distributions comparable, *pre-S1* and *post-S1* values were log-transformed and normalised relative to the pre-S1 mean for that particular combination.

5.3.4 Interseasonal immune waning

In the case of standard homogenously mixing, deterministic, non-spatial *SIR* models, epidemics end due to the critical depletion of susceptibles; reinvasion is not possible unless the proportion of susceptibles is restored to a level, where $R_e > 1$. In my model framework, restoration of network connectivity is achieved through immune waning, where a proportion δ of individuals of the S_0 immune class, irrespective of their age group, are randomly selected and transition into S_{naive}. I investigated how network and immune structure would affect the rate at which immune waning would restore network connectivity.

The amount of effective excess degree restored increases concurrently with δ (Fig. 5.6; see Fig. C.4 for raw distributions). When there is "*no structure*", age-specific heterogeneities in the immune class distributions has negligible effects (purple lines in Fig. 5.6). The addition of "*household*" structure (yellow lines in Fig. 5.6) and/or heterogeneities in immune class distribution have little further impact on the ability of immune waning to restore network connectivity. Again, the random distribution of extra-household contacts facilitates overall transmission through the population by mitigating the effects of limited assortative mixing within households.

Addition of "*social structure*" alone (dashed blue line in Fig. 5.6) moderately hinders the restoration of overall network connectivity. The differences in contact patterns and degree distribution shifts the burden of disease towards children; since δ remains constant between age groups, for any given waning rate, network connectivity is restored to a greater extent for adults than children. By $\delta = 0.4$, the effective excess degree in adults has been restored to their *pre-S1* levels, whilst for children, only 0.74 (95% quantile: $(0.52, 1.05)$) of the effective excess degree lost had been restored.

In the presence of both "*social structure*" and heterogeneities in immune class distribution between age groups (solid blue line in Fig. 5.6), the ability to restore overall network connectivity becomes further compromised; through skewing the degree distribution, these two factors act in conjunction to heighten the differential effectiveness of resusceptibilisation between children and adults. In this scenario, the adult effective excess degree can be fully restored by approximately $\delta = 0.275$, whilst over 0.6 of the effective excess degree initially lost by children remains unaccounted for.

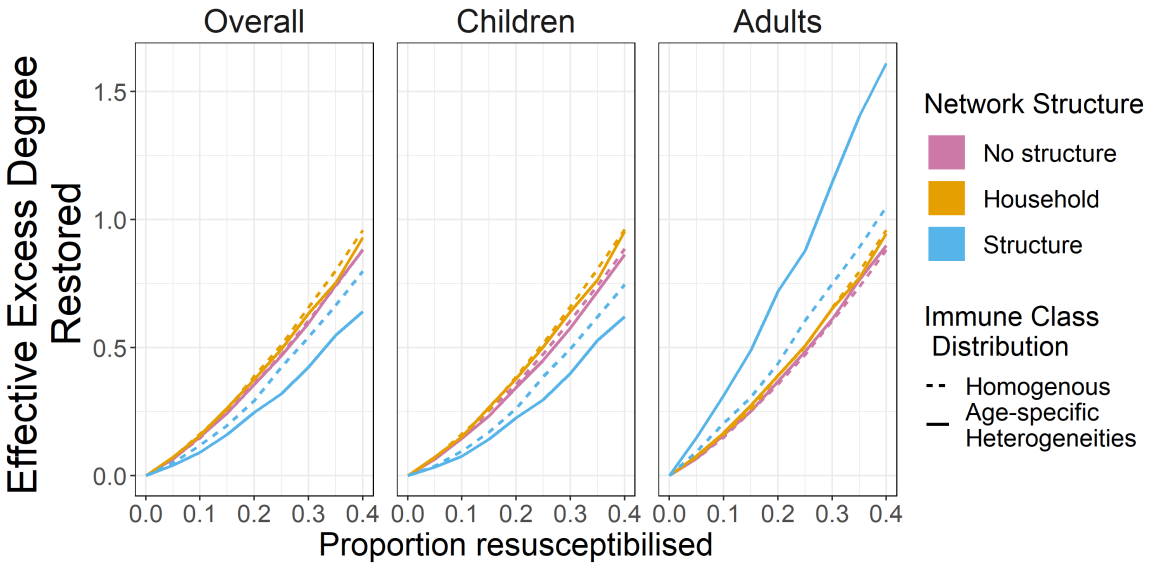


Fig. 5.6 The effect of resusceptibilisation on restoring the effective excess degree of the *post-S1* network. The lines represent the mean amount of effective excess degree restored (Eq. (5.7)) for a given proportion of S_0 individuals resusceptibilised during the interseasonal period between *S1* and *S2*. The amount of effective excess degree restored is measured relative to the amount lost: a value of 1, implies that EED_{PreS2}=EED_{PreS1}.

$$\text{Restored Effective Excess Degree} = \frac{EED_{PreS2} - EED_{PostS1}}{EED_{PreS1} - EED_{PostS1}} \tag{5.7}$$

5.3.5 Season 2 final size

Having establishing how network and immune structures shape the spread of the *S1* epidemic and modulate the effectiveness of interseasonal immune waning, I investigated how these changes in network connectivity would then affect the dynamics observed in *S2*. In the absence of immune waning, *S1* epidemics can, if successfully initiated, render the network sufficiently sparse and prevent the occurrence of major *S2* epidemics (leftmost column in Fig. 5.7a). However, inherent stochasticity can lead occasionally to the premature termination of the *S1* epidemic, which becomes more likely in the presence of network structure or segregation of immune classes based on age groups (Fig. 5.7b): this provides an opportunity for the occurrence of a minor or major *S2* epidemic, depending on the extent of spread by the *S1* epidemic.

With increasing values of δ and restoration of network connectivity through immune waning, the probability of successfully initiating *S2* epidemics increases ($OR > 1$; Fig. 5.8 and Table C.2); by $\delta = 0.4$, major epidemics that are comparable in size to those of *S1* can

be readily observed (Fig. 5.7). However, the binary logistic regression models (Fig. 5.8 and Table C.2) show that the ability of resusceptibilisation to enable *S2* epidemics is reduced by "*social structure*" and to a lesser extent, by age-specific heterogeneities in immune class distribution. After all, heterogeneous immune class distribution by itself is unable to exert any impacts, if the edges that connect individuals are randomly distributed.

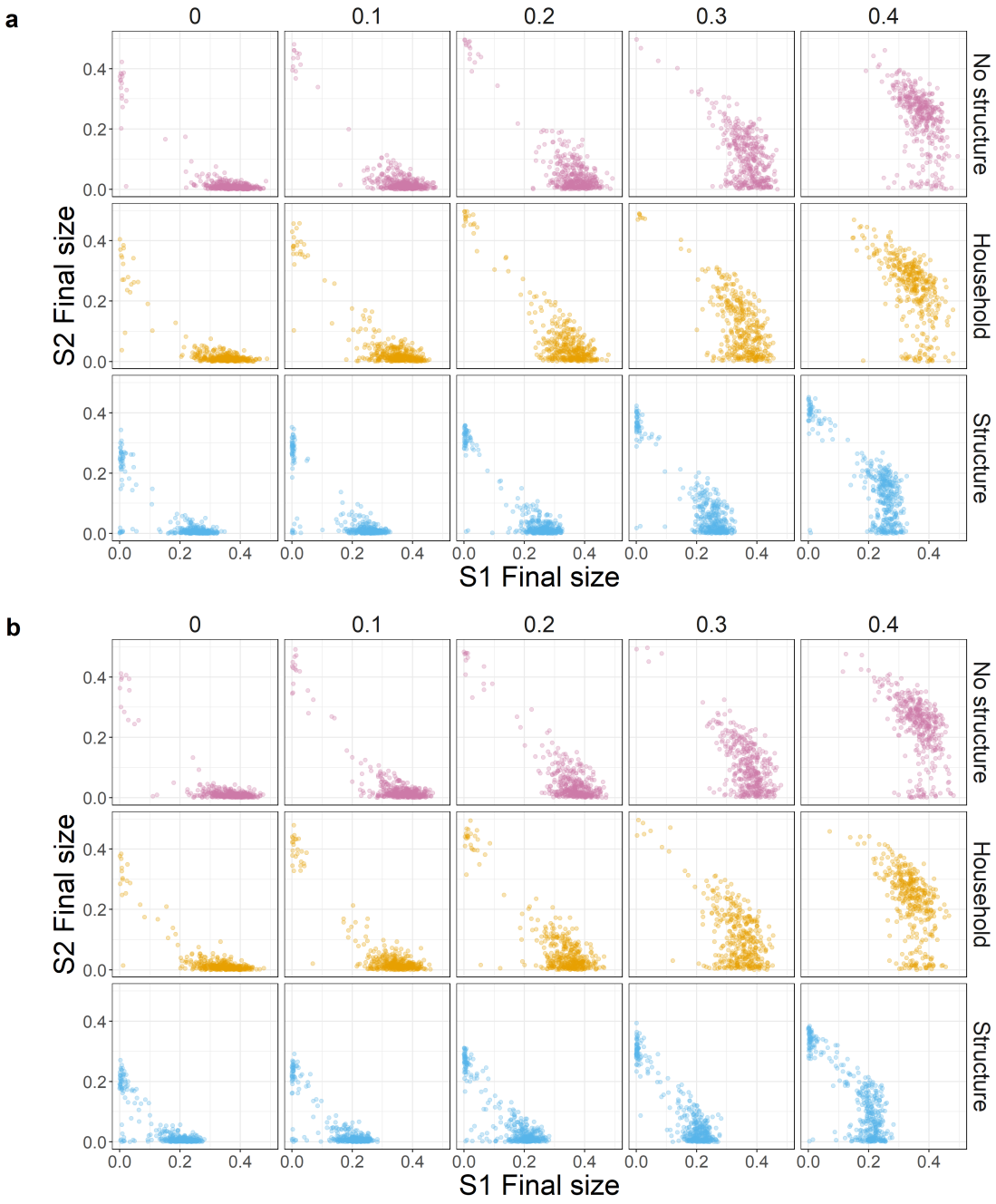


Fig. 5.7 The final size of consecutive epidemics. Immune classes are distributed (a) homogenously or (b) segregated based upon age groups. Individual points denote the final size for the *S1* and *S2* epidemics for a single model realisation. The proportion of S_0 individuals resusceptibilised (δ) between *S1* and *S2* is shown by the top facet label, steadily increasing from left to right. Note that at high waning rates, it is possible for *S2* epidemics to be substantially larger in magnitude, due to the lack of substantial *S1* activity and resusceptibilisation of existing S_0 individuals.

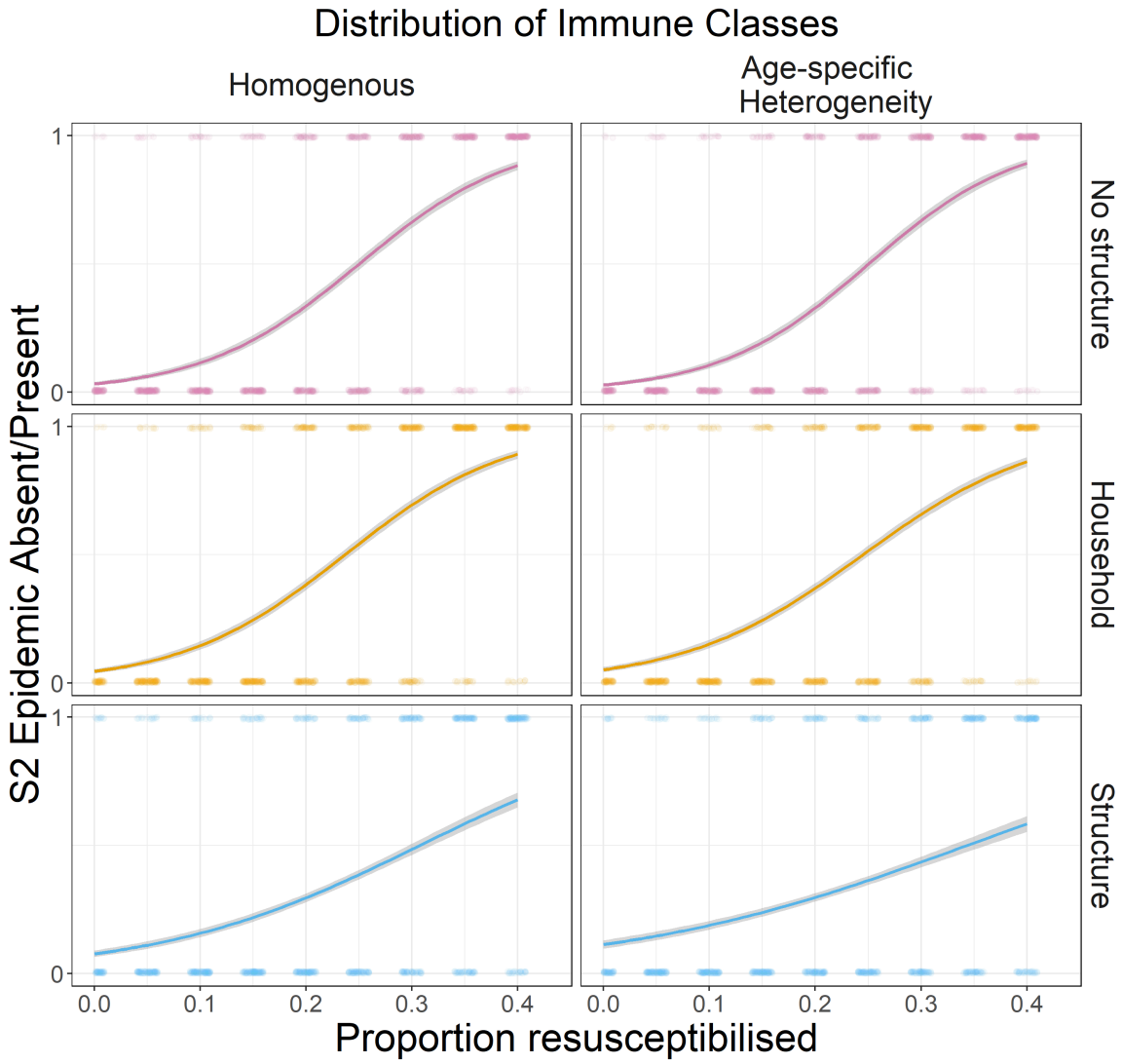


Fig. 5.8 Effect of immune waning on the probability of a major *S2* epidemic. For each value of immune waning, the probability of a successful *S2* epidemic is given by the proportion of realisations, with total incidence above 0.05. Binary logistic regression models were fitted for each combination of network structure and immune class distribution. The 95% confidence interval is denoted by the grey shaded area. See Table C.2 for *OR* from the binary logistic regressions.

5.4 Discussion

It has been widely acknowledged that the structure of host populations have profound effects on disease transmission, which cannot be readily recapitulated by models that adopt the classical random mixing assumption or variants thereof[18]. In an attempt to reduce model

complexity, existing influenza virus models have focused on characterising either the effects of cross-immunity[130,131] or population structure[19,124]. Some approaches have attempted to incorporate elements of both[29,278], but have failed to capture their finer aspects: crucially, age structure, which imposes additional layers of heterogeneity by inducing correlations between the two. I build upon the frameworks of previous studies to explicitly investigate how interactions between the structure of host contacts and topology of immunity can shape network connectivity and define the region of parameter space permissive to recurrent epidemics by the same antigenic variant.

By constraining the interactions between individuals, network topology dictates pathogen dynamics: in unipartite networks with arbitrary degree distributions, individuals differ only in the number of contacts, so nodal degree determines their risk of infection[19,78,177,192]. Aside from the inclusion of a non-trivial network structure with degree correlations, the addition of partial cross-immunity within the network introduces further asymmetry: the frailty of individual nodes towards infection needs to account for the susceptibility of not only themselves but also that of their immediate neighbours. As an extension to the classical results from unipartite networks[78,192], I find that the burden of infection disproportionately falls upon nodes that have a greater effective excess degree (Figs. 5.3 and C.2), a composite measure of network connectivity. Within the context of my network models that are parameterised using empirical contact survey data and plausible immune class distributions, I show that children are particularly vulnerable to infection due to their assortative mixing within education settings, resulting in the formation of subgroups and clusters of susceptibility.

Differential susceptibility and assortative mixing[192,229] can skew the effective excess degree distribution and can appear to enhance disease spread. However, these previous findings stem from the fact that the transmission probability along an edge is invariant across networks: in other words, R_e is greater in networks with assortative mixing than those without. In my framework, when R_e is held constant instead, I find that segregation of extra-household contacts through "*social structure*" hinders the overall spread throughout the network. Despite the fact that more structured networks result in smaller epidemics (Fig. 5.3), the negative impact on the connectivity of the residual network is greater (Fig. 5.5), due to the concentration of edge density within a limited number of children, who are also preferentially infected. These effects can be similarly achieved by partially segregating immune classes between age groups, which yet again acts to concentrate susceptibility and the edges that are most permissive to transmission within a limited number of children.

Residual networks are more resilient towards reinvasion since epidemics preferentially remove highly connected individuals and their connections: this pruning leaves behind sparser chains of susceptible individuals so subsequent *S2* epidemics are more prone to

premature stochastic truncation. Similar to the concept of herd immunity and threshold values in classical *SIR* models, recurrence of major epidemics can only become viable once susceptibility within the epidemiologically active portion of the network is sufficiently restored through changes in virus antigenicity or host immunity. By limiting the size of *S1* epidemics, network structure has been theorised to reduce the amount of immunity lost that is necessary to enable *S2* epidemics[124].

This previous study[124] runs contrary to my results where I find that a greater amount of immune waning is required for networks with "*socially structure*" or age-specific immune class distributions than for random networks, to restore network connectivity between epidemics (Fig. 5.6). In fact, this apparent quandary highlights the importance of specifying network structure, beyond that of just the degree distribution. Configuration models[124] are random graphs and lack assortative mixing that arises from clustering and degree correlations. Compared to the assortative mixing captured by "*social structure*" in my more complex network models, disassortative mixing results in high degree nodes being more broadly distributed throughout random graph networks[192]. This means that these individuals, who are crucial in driving disease dissemination, are more readily accessible: for any node within such a network, the path to a high degree individual is shorter, involving fewer intermediary nodes, so it is less difficult to re-establish network connectivity.

Counterintuitively, whilst network structure and segregation of immune classes between age groups act to limit the size of the *S1* epidemic, network connectivity is disproportionately affected and is less effectively restored by interseasonal immune waning. Consequently, for any given immune waning rate, assuming that a major epidemic was present in *S1*, *S2* epidemics are more prone to failure in these "*socially structured*" networks with assortative mixing than in random networks (Fig. 5.8). Even with large immune waning rates ($\delta = 0.4$), *S2* epidemics still fail approximately 40% of the time. Furthermore, this level of immunity lost between seasons appears biologically implausible, given estimates of cross-immunity[8] and cross-reactivity[140] of approximately 88% and 75% respectively, between strains of the same antigenic variant over consecutive seasons.

My models provide novel insights into how the interactions between network structure and topology of immunity can have far-reaching effects on multi-season dynamics. These factors define the epidemiologically active contacts within a network, forming cyclical feedback loops as the patterns of disease then induces correlations between effective excess degree and immune status. However, these more complex caricatures of empirical populations remain unable to recapitulate recurrent epidemics without necessitating large amounts of immune waning. This is in itself an important result, by demonstrating the dependence of transmission upon high degree and susceptible children nodes, which through their intrinsic

clustering contributes to the frailty of the network. This also highlights the need to incorporate additional factors into the model to limit the acquisition or maintenance of immunity between seasons.

The spread of epidemics could be prematurely curtailed by a mismatch between epidemic activity and underlying seasonal drivers. In my model, I omitted seasonal forcing, since my previous analyses in Chapter 4 suggest that the impact of climatic factors on epidemiological dynamics may be limited. Furthermore, each discrete epidemic is brief in duration relative to the length of the year.

Epidemiological dynamics could also be affected by changes in the underlying contact network over time. In the short-term, individuals could exchange contacts amongst themselves, within the same class or workplace could be labile: on one hand, this can enhance transmission and cause larger epidemics[246] but could also enhance the effects of immune waning by opening up alternative pathways to reach the critical high degree nodes. In the long-term, demographic processes could replenish population susceptibility, through natural births and deaths. Whereas endemic sexually transmitted diseases are affected by the turnover of sexual partnerships[67,89], the acute nature of infection means that transient influenza epidemics should be less affected by demographic processes or network rewiring. The use of static networks can thus be justified since individuals are unlikely to change their membership of households, schools or workplaces over the course of an epidemic and such short-term timescales.

It is plausible that mixing between individuals of the same class or workplace is less rigid than supposed by a static network but justifying and parameterising a more complex dynamic model can only be possible with improved social contact data of sufficient temporal granularity and exhaustive egocentric sampling of individual social units. Since respondents enrolled in the POLYMOD study[186] only reported the number of daily contacts made within each social context, it is not possible to infer whether or not the identity of their alters had changed between days. Nor was it possible to infer the wider structure of contacts within each class or workplace, beyond the edges connecting ego to its immediate alters. More broad changes in host behaviour, such as school closures at the end of term[36], could also have an impact on seasonal influenza activity, but this remains poorly established[239]

Successful reinvasion could be more likely if residual networks are less frail than expected: there is growing evidence that infection may not necessarily confer long-lasting strain-specific immunity. Notably, children have proven notoriously difficult to seroconvert, with seronegative individuals requiring multiple doses of seasonal vaccine to elicit a protective antibody response[191]. Given the edge density within this age group, individual children could potentially be capable of facilitating transmission over multiple seasons, before becoming

immunised and removed from the network. Adults could also form a persisting, albeit less efficient, conduit of partially immune nodes and provide some level of network connectivity. Here, older individuals could be stuck in a perpetual state of susceptibility[100], due to antigenic seniority and the reliance of broadly neutralising antibodies[172], which blocks the development of antibody responses towards newly encountered viruses.

Population immunity may also be poorly maintained between epidemics. The duration of vaccine-induced protection has long been acknowledged to be limited[28] but rather surprisingly, homologous challenge studies have suggested that protection conferred by natural infection may be similarly transient[176]. Indeed, serological studies suggest that the half-life of protection for children is shorter than for adults[206]. Compared to the homogenous immune waning implemented in my framework, a greater amount of network connectivity can be restored with differential age-specific waning rates: for the same total number of S_0 individuals resusceptibilised, preferential resusceptibilisation of children will have a greater effect overall effect on network connectivity.

In Chapter 6, I explicitly incorporate broadly neutralising immunity and immune waning rates that differ between age groups. I assess whether these mechanisms can overcome the effects of network structure and increase the likelihood of recurrent epidemics.

Chapter 6

Changes in host immunity and its effects on the recurrence of epidemics

6.1 Introduction

In Chapter 5, I showed that interactions between host contact structure and partial cross-immunity towards seasonal influenza viruses can induce substantial correlations between nodal degree and susceptibility. These correlations increase network frailty and fragmentation, necessitating an inordinate amount of interseasonal resusceptibilisation and immune loss to overcome. In this chapter, I consider more explicitly how an individual's immunity changes with age and assess whether these additional immune mechanisms can modulate the effectiveness of immune waning and increase the likelihood of recurrent epidemics.

Since virus reinvasion is intrinsically linked to the immunity profile of the population, serological studies have proven invaluable in elucidating potential mechanisms for maintaining or restoring network connectivity. Specifically, these mechanisms must ameliorate one or both of the disruptive effects of epidemics: 1. frailty: the targeted and preferential immunisation of highly connected individuals; 2. interference: a more generalised depletion of susceptibles, rendering potential chains of transmission more sparse[78]. Resulting from the aforementioned correlations between nodal degree and susceptibility, both effects of epidemics disproportionately affect children so potential mechanisms are likely to draw upon differences in the immune responses between age groups.

Children tend to produce antibodies targeting the antigenically variable head of HA; as such, seropositivity towards currently circulating strains peaks in school-aged children[138]. Seropositivity then decreases with age, reaching a minimum for middle-aged adults[138]. This change could be the result of interference from antigenic seniority or reduced adaptability

in one's immune response with age. Over the course of one's lifetime, sequential exposure to divergent virus subtypes[136] could result in increased recognition of conserved but subdominant epitopes[134,237]. The effects of antigenic seniority are thus hypothesised to prevent older adults from being able to mount an effective and specific antibody response[172]. Alternatively, similar outcomes and the shift towards anti-stalk antibodies could be attributable to immunosenescence, which restrict somatic hypermutation and the clonal diversity of B cells, thus resulting in reduced capacity of antibodies to target antigenically variable HA epitopes[110].

Adults could be trapped in a perpetual state of partial susceptibility[100], since their immune response becomes increasingly dominated by broadly neutralising antibodies, potentially forming chains of partially susceptible individuals within the network that are not readily depleted by epidemic activity. However, intuition suggests that this may only have limited effects on overall network connectivity, given how adults are sparsely connected; compared to children, the contribution of adults towards community transmission is limited[72,182,255]. If such a resilient component comprising of adults exists, its weak impact on epidemiological dynamics has proven difficult to detect, with circumstantial evidence only coming to light once the highly connected children are sufficiently immunised after multiple epidemics. Typically, heightened activity is first detected in children but after two consecutive A/H3 virus dominated seasons in the USA, peak levels of A/H3 activity were instead first detected in middle-aged and older adults during the 2017-18 season[87,100].

Reliance on different immune responses could contribute to differences in the duration of protection conferred by infection, which is estimated to be shorter for children than adults[206]: greater interseasonal loss of immunity in children could help to mitigate their particular frailty and restore connectivity in parts of the network that are most affected. The shorter duration of immunity within children can be ascribed to the high degree of individual variability in post-infection HI titres, which is a correlate of the more specific anti-HA head antibody response. In turn, this variability can be attributed to underlying heterogeneities in both the pre-existing titres and the magnitude of the acute boost due to infection.

In this chapter, I build upon the modelling framework laid out in Chapter 5 and investigate the effects of broadly-neutralising antibodies and variable rates of immune waning, quantifying their overall effectiveness in enabling virus reinvasion over consecutive seasons. Whilst the biological basis suggests that these putative mechanisms are capable of lessening the immunological imprints imparted by individual epidemics, there is a need for explicit consideration of underlying network structure.

6.2 Method

I utilise the same general framework that I developed in Chapter 5, with regards to generating the underlying contact structure, distribution of immune classes and simulating the trajectories of epidemics over consecutive seasons. I modified the model by adding a broadly immune class of hosts and specifying age group-specific immune waning rates; model outputs were then contrasted against those generated by the base model from Chapter 5. For clarity, key results presented in Figs. 6.4–6.7 are for networks with "*social structure*" to highlight the impact of the various immune waning scenarios on the disproportionate frailty of children nodes; results for "*no structure*" and "*household*" networks can be found in Appendix D (Figs. D.3–D.6).

6.2.1 Incorporating broadly immune class

An additional immune class S_{broad} was incorporated to represent older adults, who possess broadly neutralising immunity towards all antigenic variants ($\delta_{(broad)j} = 0.5$, for all j). After recovering from infection, these hosts cannot become reinfected within the same season and remain in the same S_{broad} class, reflecting their inability to develop a specific antibody response towards the current wild-type virus. I replaced the immune class S_4, which consists of individuals who encountered and were able to develop strong immunity towards viruses from four antigenic clusters ago, with S_{broad} (Fig. 6.1). This corresponds to approximately ten to twenty years since last successfully immunised. For adults, the ability to mount a specific response after virus exposure declines dramatically with age beyond the late 20s[138], reaching a trough at around the age of 40.

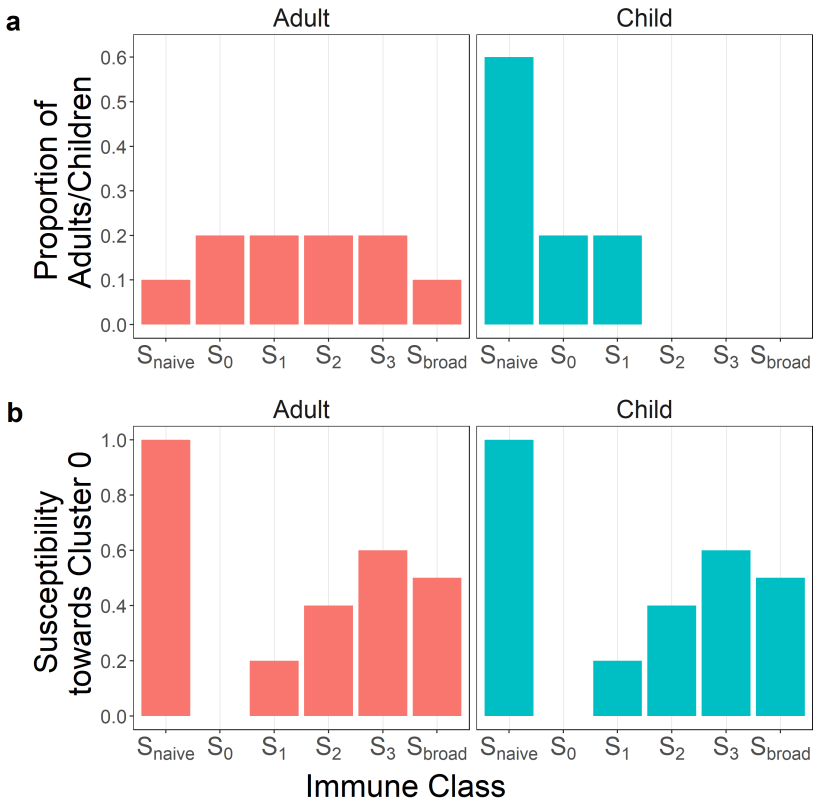


Fig. 6.1 The distribution of (a) immune classes within the population and (b) their corresponding susceptibility towards the current antigenic variant of cluster 0. For the susceptibility profiles shown, I used a linear model of cross-immunity with $\theta = 0.2$ (Eq. (5.1a)).

6.2.2 Varying the interseasonal loss of immunity between age groups

In the original model framework, during the interseasonal period between *S1* and *S2*, a proportion δ of individuals possessing immunity towards the current wild-type (class S_0) are randomly selected and become re-susceptibilised (class S_{naive}). Here, I further delineate between age groups (class $S_{children_0}$ and S_{adults_0}) and apply age group-specific immune waning rates ($\delta_{children}$ and δ_{adults}). Parameter values were guided by experimental studies: Ranjeva et al.[206] estimated the half-lives for immune protection towards A/H3 viruses to be $t_{\frac{1}{2}children} = 3.5$ years, 95% CI:$(1.4, 5.2)$ and $t_{\frac{1}{2}adults} = 4.1$ years, 95% CI:$(3.2, 5.5)$. From these half-lives, I calculated the year-on-year rates of immune waning, which I used as parameter values for my model: $\delta_{children} = 0.18$, 95% CI:$(0.12, 0.39)$ and $\delta_{adults} = 0.16$, 95% CI:$(0.12, 0.19)$.

To characterise the extent to which network connectivity can be more effectively restored through variable immune waning rates, I considered a series of immune waning scenarios

(Table 6.1), which are contrasted against the null model with homogenous immune waning as discussed in Chapter 5.

Duration of Immunity	$\delta_{children}$	δ_{adults}
Long	0.12	0.12
Mid	0.18	0.16
Short	0.39	0.19
Mixed	0.39	0.12

Table 6.1 Immune Waning Scenarios

In order to make the varying $\delta_{children}$ and δ_{adults} across these scenarios comparable against the single parameter δ of the homogenous waning null model, I calculate the number of S_0 individuals resusceptibilised during the interseasonal step for each stochastic realisation. This common metric provides a per capita measure for the effectiveness of resusceptibilisation in restoring network connectivity and enabling *S2* epidemics.

6.3 Results

6.3.1 Effects of broadly immune individuals

Interseasonal immune waning

I compared networks with and without the S_{broad} immune class to evaluate the hypothesis that adults could form a perpetually susceptible and resilient component of the network. In the context of *S1* epidemic dynamics, the two cases are identical since I ensured that the overall immunity profile for both cases remain unchanged; furthermore, S_{broad} individuals cannot be reinfected over the course of a single season. In networks that have little to no assortative mixing within age groups ("*household*" and "*no structure*" respectively), the presence of S_{broad} reduces the amount of effective excess degree lost *post-S1*: network connectivity can be readily restored by immune waning to original *pre-S1* levels (Fig. 6.2). In what are essentially random graphs, these highly resilient S_{broad} individuals are interspersed widely and easily accessible by all nodes, boosting nodal effective excess degree for both age groups (Fig. D.1). This effectively mitigates the effects of interference from epidemics.

In networks with "*social structure*", the underlying contact patterns of adults and the limited number of S_{broad} individuals means this immune class exerts limited effects on network connectivity: the effective excess degree distribution for adults is slightly affected

(Fig. D.2). At low values of δ, S_{broad} marginally increases the connectedness of adults (Figs. 6.1 and D.2). Paradoxically, at high values of δ the effective excess degree of adults is lower since the expansion of the S_0 class is blocked, so there are fewer individuals to resusceptibilise; this effect is somewhat analogous to a weaker form of interference from epidemics. However, these changes in the network properties of adults are not manifested within children or the overall network: individuals possessing broadly neutralising immunity are akin to dead-ends in chains of transmission.

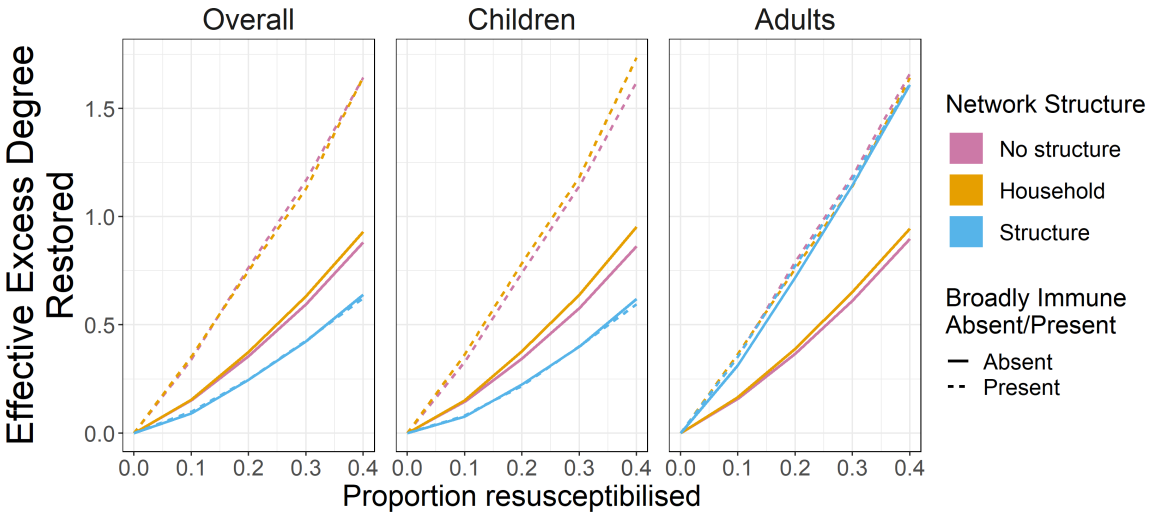


Fig. 6.2 The impact of the S_{broad} immune class on the ability of interseasonal immune waning in restoring network connectivity. At each value of δ, the lines give the mean amount of effective excess degree restored, relative to the amount lost (Eq. (5.7)).

Likelihood of S2 Epidemic

The impact of broadly neutralising immunity varies between the network structures. These differences in *pre-S2* connectivity affect the likelihood of *S2* epidemics (Fig. 6.3): with the random mixing in "*no structure*" and limited clustering in "*household*" networks, the presence of S_{broad} facilitates successful *S2* epidemic initiation. This effect is particularly evident at low values of δ, where the *pre-S2* network remains highly fragmented (Fig. 6.3 and Table D.1). At large values of δ, the large turnover of population immunity negates the interference effects of epidemics so the impact of S_{broad} becomes negligible.

Since the addition of S_{broad} does not substantially alter the *pre-S2* effective degree distributions for networks with "*social structure*", it is unsurprising that the likelihood of *S2* epidemics remains mostly unaffected; when there is substantial immune waning, the likelihood is reduced slightly (Fig. 6.3), as a consequence of limiting the size of S_0 and extent to which resusceptibilisation can occur, as discussed above.

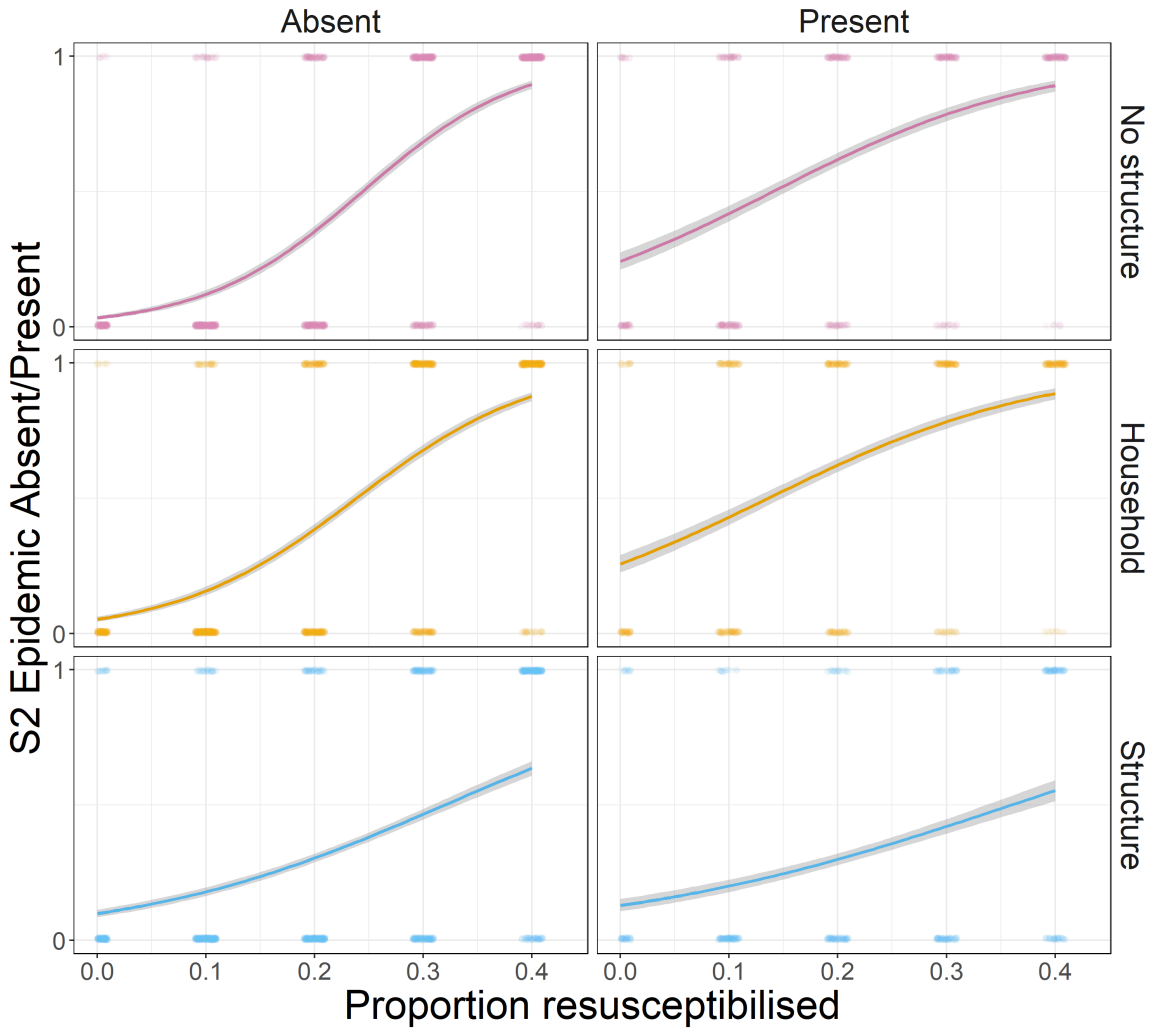


Fig. 6.3 The impact of the S_{broad} immune class on the probability of a major *S2* epidemic. For each value of immune waning, the probability of a successful *S2* epidemic is given by the proportion of realisations, with total incidence above 0.05. Binary logistic regression models were fitted for each combination of network structure and immune class distribution. The 95% confidence interval is denoted by the grey shaded area. See Table D.1 for *OR* from the binary logistic regressions.

6.3.2 Effects of variable immune waning rates between age groups

Interseasonal immune waning

Consistent with the results in Chapter 5, I find that as the duration of immunity is reduced (left to right in Fig. 6.4), the average *pre-S2* effective excess degree is restored to levels closer to *pre-S1* values. When comparing the "*mixed*" and "*short*" immune waning scenarios, it is evident that increasing δ_{adults} from its minima to maxima has little further effect on

increasing network connectivity: after all, the confidence interval for δ_{adults} is narrow, relative to the range for $\delta_{children}$. Additionally, due to correlations in nodal degree and susceptibility, children nodes suffer especially from frailty so the restoration of network connectivity is most sensitive to $\delta_{children}$ (Fig. 6.5). Indeed, Fig. 6.5 shows that immune waning that varies between age groups is more effective than homogenous waning at reducing network frailty: for the same number of S_0 individuals resusceptibilised, 0.58 of the initial reduction in average effective excess degree is restored in the "*mixed*" waning scenario, as compared to 0.31 with homogenous waning. In other words, if the objective is to maximise the amount of network connectivity restored whilst minimising the overall level of population immunity turnover, resusceptibilising a child, rather than an adult, would have a greater impact.

Whilst the confidence interval for the duration of protection conferred by infection is substantially wider for children[206], even so, at the upper bound of experimentally informed rates, the *pre-S2* effective excess degree for children is still lower than original *pre-S1* levels (Fig. 6.4).

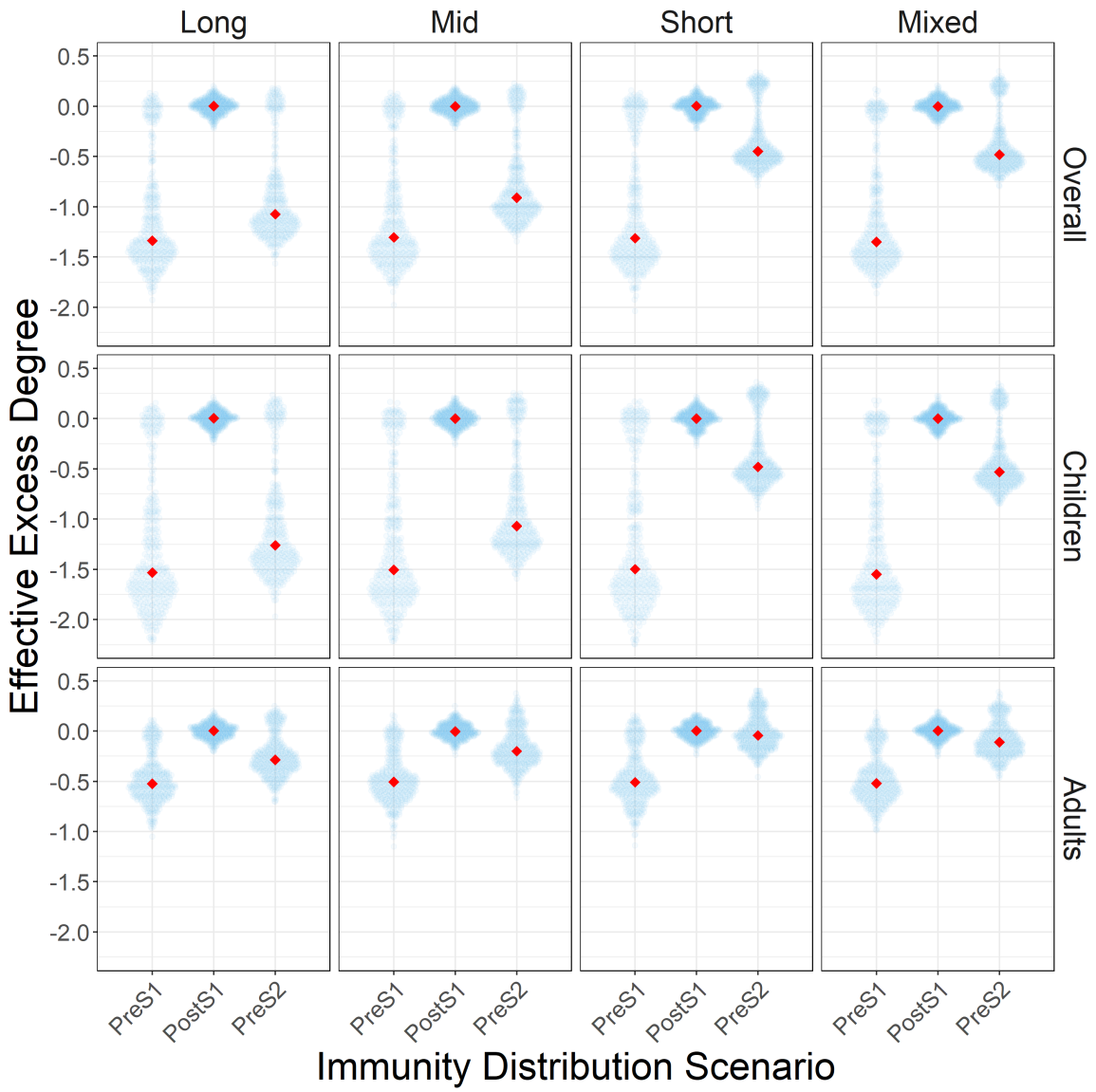


Fig. 6.4 Changes in network connectivity across different variable immune waning scenarios. Immune waning scenarios and the corresponding age group-specific rates are listed in Table 6.1. The left, central and right distributions show the original *pre-S1*, residual *post-S1* and resusceptibilised *pre-S2* distributions for average effective excess degree; the red point denotes the median value. To make the effective excess degree distributions comparable, *pre-S1*, *post-S1* and *pre-S2* values were log-transformed and normalised relative to the *pre-S1* mean for that particular combination. Results for "*no structure*" and "*household*" networks are shown in Appendix D (Fig. D.3).

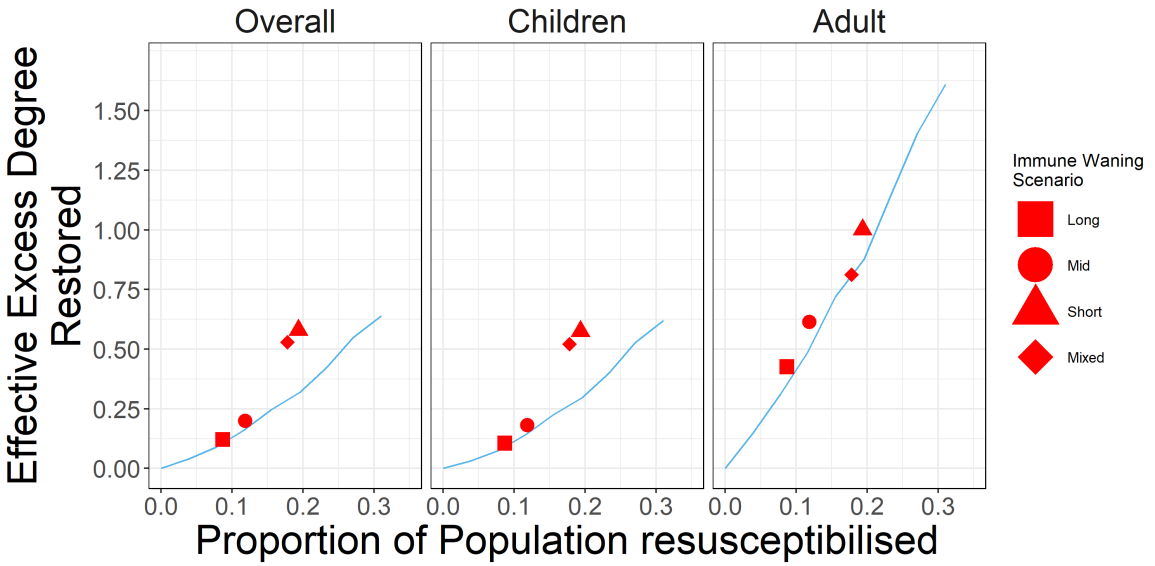


Fig. 6.5 The effect of variable interseasonal immune waning on restoring the effective excess degree of the *post-S1* network. The various immune waning scenarios (see Table 6.1 for corresponding waning rates) are compared against the null homogenous waning model, which is denoted by the blue line showing the mean amount of effective excess degree restored (Eq. (5.7)) for a given number of S_0 individuals resusceptibilised. Results for "*no structure*" and "*household*" networks are shown in Appendix D (Fig. D.4).

Season 2 Final Size

Reducing the duration of immunity (left to right in Fig. 6.6) increases the likelihood of a major *S2* epidemic, especially if a major *S1* epidemic has occurred: when children are preferentially resusceptibilised, this increase in likelihood is relatively larger than what would otherwise have been expected from homogenous waning (Fig. 6.7). This reflects the underlying changes in network connectivity and the importance of children in facilitating disease transmission. However, *S2* epidemics are still prone to failure, underlining the fact that the *pre-S2* effective excess degree levels are still lower than *pre-S1* levels.

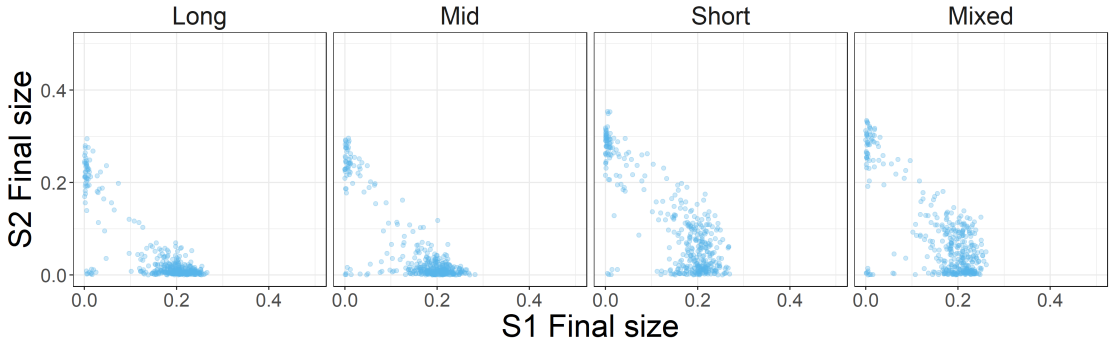


Fig. 6.6 The effect of variable interseasonal immune waning on the final size of consecutive epidemics. The corresponding immune waning rates for each scenario are listed in Table 6.1. Individual points denote the final size for the *S1* and *S2* epidemics for a single model realisation. Results for "*no structure*" and "*household*" networks are shown in Appendix D (Fig. D.5).

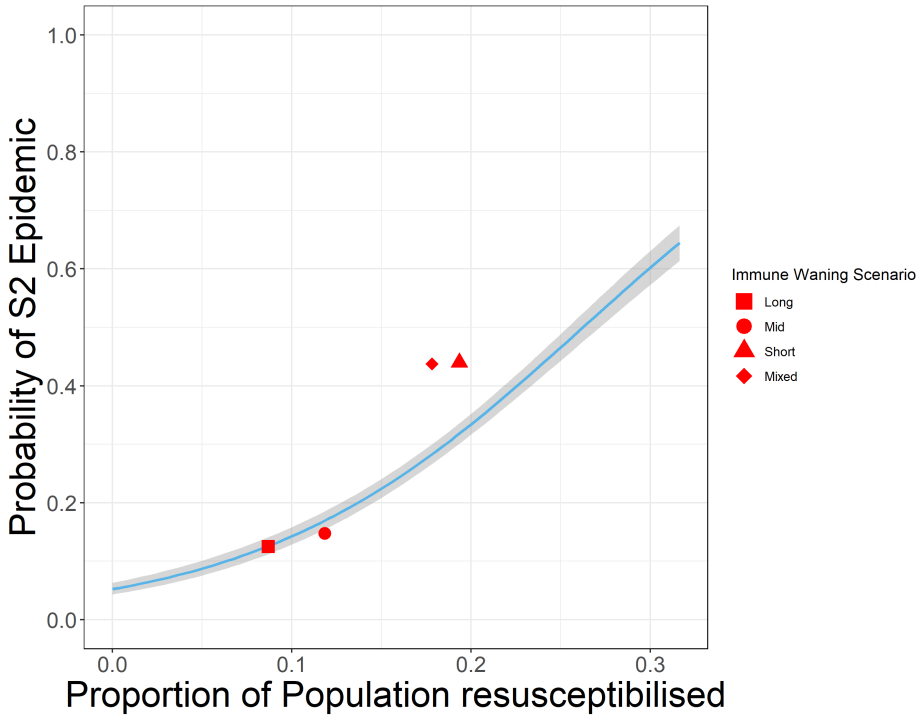


Fig. 6.7 The effect of variable interseasonal immune waning on the probability of a major *S2* epidemic. The various immune waning scenarios (red points) were contrasted against the null homogenous waning model: a binary logistic regression model (blue line) was fitted to the latter and shows the probability of successful *S2* epidemics for a given number of S_0 individuals resusceptibilised. The probability of a successful *S2* epidemic is given by the proportion of realisations, with total incidence above 0.05. Results for "*no structure*" and "*household*" networks are shown in Appendix D (Fig. D.6).

6.4 Discussion

The epidemiology of an antigenically variable pathogen is inherently linked to the nature and distribution of host immunity. Notably, an individual's serological response is moulded by sequential exposures to different sets of related strains of seasonal influenza viruses, which culminate in immune responses that differ between age groups. These resultant shifts in individual immunity could lessen the disruption of network connectivity by epidemics, thus making recurrent epidemics more likely. In this chapter, I assess whether recurrent epidemics within structured populations can be facilitated by immunological mechanisms involving broadly neutralising antibodies in adults or differences in the duration of immunity between age groups.

It has been hypothesised that the development of broadly neutralising antibodies may confine middle-aged adults in a perpetual state of susceptibility[100] by preventing individuals from mounting a novel and specific antibody response. Cross-sectional serological studies have readily detected the presence of these antibodies[237], with increasing prevalence with age[178,189]; however, baseline titres in individuals are typically lower than what could be considered protective levels[237] and the associated plasmablasts have proven difficult to isolate[47]. The significance of such immunity at the population level is thus questionable, especially since these serological studies often have limited sample sizes.

Using my network models, I demonstrate that adults with broadly neutralising immunity are unlikely to strengthen connectivity within networks with social structure (Figs. 6.2 and 6.3). Adults have lower edge densities than children: creating isolated clusters of susceptibility at the periphery of the graph does little to facilitate transmission within the core of the network, which is predominantly formed by children. Despite the fact that the exact titre levels that correlate with protection have not been determined for broadly neutralising antibodies[237], it is likely that $\delta_{(broad)j}$ of 0.5 is an over-estimate of the level of individual immunity, so the *in vivo* significance in community transmission is expected to be further reduced.

Serological studies have highlighted that the duration of immunity conferred by infection may be shorter in children than adults[206]. Children rely upon a specific response towards immunodominant epitopes of HA, resulting in a high degree of variability in the baseline HI titres, post-infection titre boosts and overall levels of protection, which is further reduced by virus antigenic drift. On the other hand, the shift in recognition towards conserved stalk epitopes results in a less labile response by adults.

This difference in immune waning rates between age groups specifically target the disruptive effects of epidemics, which disproportionately fall upon children. Indeed, I show that these differences can lead to more effective restoration of network connectivity,

achieving greater per capita effects than applying homogenous waning equally across all age groups (Figs. 6.5–6.7). Within the range of experimentally informed parameter values (Table 6.1), antigenic variants are readily able to reinvade the same population, even after a major epidemic. However, the *S2* epidemic is by no means guaranteed and still prone to failure. Indeed, my analysis of seasonal influenza activity in Australia in Chapter 4 shows that failure to initiate subsequent epidemics occurs more frequently for A/H1sea and B/Yam than A/H3 and B/Vic viruses.

As I demonstrated in Chapter 5, neither host contact structure nor immunity can be considered in isolation: their particular topologies can induce unexpected correlations between nodal attributes and produce joint effects that counterintuitively exacerbate network fragmentation. Results from this chapter further reinforce this sentiment: the effectiveness of broadly neutralising immunity is curtailed by the poor connectedness of adults. In contrast, the effects of immune waning can be maximised by varying its rates between age groups, capitalising on their intrinsic differences in connectedness. These results further reinforce the importance of children in the spread of seasonal influenza viruses. Likewise, our intervention strategies should be targeted towards children: by taking advantage of their nodal properties, interventions would be disproportionately effective at reducing the overall population level burden of disease and exceed the impact that could be achieved through untargeted, blanket approaches. In order to elucidate the effects of immunity at the population level, there is a need to clarify how individual immunity and host contact patterns evolve over time. Cross-sectional or cohort studies are required to further characterise the patterns in the distribution of humoral immunity and disease-causing contacts between hosts, with a particular focus on elucidating how immunity accumulates in children.

Chapter 7

Discussion

7.1 Summary of findings

Our ability to better control seasonal influenza viruses is partially limited by our poor grasp on how interactions between virus and host translate across scales, from infection and immunity within an individual to virus transmission through a population and the emergence of new virus strains at the global level. Although inconclusive, great strides have been made over the past decade in characterising the spectrum of individuals' serological responses[140,151], from which attempts have been made to infer their history of virus exposure[81] and the mechanisms that underlie the development of immunity[140]. Similarly, our understanding of global evolutionary dynamics has been enriched by improvements in virus surveillance[22,25,205,212], albeit still confounded by historical systemic biases in regional coverage.

Local epidemics form an important crucible in which virus fitness interacts with host immunity in a reciprocal manner. Identifying and quantifying the drivers behind observed epidemiological dynamics at this intermediary scale are thus crucial. For local public health authorities, such insights could facilitate more accurate epidemiological forecasting, enabling more timely resource allocation and targeted interventions. Globally, vaccine strain selection can be improved through a greater understanding of how the emergence of new antigenic variants is shaped by local immune selection pressures. However, previous studies have only incompletely explored the interactions between virus, host and the environment, due to a lack of highly resolved data that is geographically disaggregated and clearly delineates between virus subtype and antigenic variants. The aim of this thesis was to address this specific knowledge gap: I characterised the phylodynamical interactions between seasonal influenza viruses, climate, and human host populations, integrating analyses from observational study and theoretical modelling approaches.

I firstly develop a statistically-principled approach to epidemic detection, which overcomes the specific challenges posed by highly resolved surveillance data, where antigenic characterisation comes at the expense of sample size. With this tool, I inferred the timing of individual epidemics, with which I was then able to critically investigate the effects of environmental and virological factors on the timing and magnitude of local epidemics for the five most populous cities in Australia (Chapter 4). In contrast to previous studies, I found that climatic fluctuations[219,222] and virus antigenic change[24,29,130,270,274] do not have consistent effects on the epidemic timing or magnitude, highlighting the need to avoid gross geographical aggregation, which obscures important local differences in virus activity. Most surprisingly, I observed that viruses from the same antigenic variants are capable of reinvading the same city over consecutive seasons, which conflicts with the predictions from existing epidemiological models[130,131], where the local depletion of susceptibles is hypothesised to prevent reinvasion.

In the absence of marked phenotypic change, it follows that population susceptibility must somehow remain at higher than expected levels in order for recurrent epidemics to occur. One possibility is that host contact structure limits epidemic spread[124] and the build-up of population immunity. In parallel, susceptibility levels could be restored between seasons through immune waning[176]. Previous studies have considered the effects of network structure and immunity in isolation; in the context of pandemic influenza, the contributions of the latter were ignored since there was an expectation of negligible pre-existing population cross-immunity[124]. Of course, such simplifying assumptions cannot be made for seasonal influenza viruses. In Chapter 5, I show that the differing contact patterns and levels of immunity between adults and children can result in correlations in nodal degree and susceptibility. Whilst this interaction shifts the burden of disease disproportionately onto children and reaffirms the importance of children in facilitating community transmission, it also exacerbates network frailty and fragmentation, thereby reducing the likelihood of virus reinvasion.

Since the structure of host contacts and immunity concentrates epidemiologically active edges within children nodes, I explicitly considered whether changes to an individual's immune response with age could act synergistically with host immune waning and increase the likelihood of recurrent epidemics (Chapter 6). I investigated the impact of broadly neutralising immunity of adults[100,134,172] and variable immune waning rates that differed between age groups[191,206]. I found that incorporating a broadly immune class[100] did little to preserve network connectivity between seasons. Adults are poorly connected so creating pockets of individuals with permanent partial susceptibility within the periphery of the network has minimal impact. In contrast, when I account for the shorter duration of immunity in children relative to adults[206], network connectivity is restored more effectively: immune

waning can be applied more judiciously to counter the specific frailty of children nodes. These results further consolidate the importance of accurately characterising the patterns in the distribution of humoral immunity and disease-causing contacts between hosts.

7.2 Strengths and limitations

The highly resolved virus surveillance data set enabled me to accurately analyse the effects of climatic factors and virus antigenic change on city-level epidemics (Chapter 4). This is in itself a key innovation but the limited size of the data set is an aspect that warrants further improvements. Firstly, antigenically characterising a greater proportion of submitted influenza virus positive samples would greatly reduce the uncertainty surrounding the timing of epidemic onset and allow for aggregation of cases by shorter time periods: aggregation by week could better capture short-term fluctuations in climatic conditions. Additionally, there is a need for a more holistic assessment of virus antigenicity, beyond the major phenotypic changes derived from substitutions in the immunodominant epitopes present on the globular head of HA. Microneutralisation assays can capture subtle changes in antigenicity[145,156] and enable better assessment of the impact of year-to-year variations in antigenic drift. Furthermore, there is a need to move away from relying upon reference sera derived from naïve ferrets, since the lack of varied immune exposure and response result in sera that fails to reflect the heterogeneous nature of *in vivo* responses[149]. More broadly speaking, these changes in the approach to virus characterisation represent much needed improvements to vaccine strain selection that are achievable using current technologies.

However, I acknowledge that such a fastidious approach to virus characterisation may prove to be highly costly and labour intensive for routine surveillance, especially since it involves moving away from established HI assay based protocols. Increasing the geographical resolution for virus surveillance would be another avenue for improvement that is differently ambitious and could yield equally valuable insights into local reinvasion dynamics at the level of individual suburbs or communities. Indeed, it would be fascinating to investigate how influenza viruses spread amongst the individual communities that form the overall population. This would also produce actionable information that can immediately inform public health interventions, identifying risk factors that predispose certain communities to unusually high burdens of disease[279] or patterns[55] in epidemiological dynamics.

There is a need to replicate my analyses across other cities situated in temperate regions to validate my findings on the limited effects of climatic factors and virus antigenicity. Despite being a notifiable disease in many countries, including the UK, many surveillance networks fail to record data that is representative of virus activity within the wider population.

Routine testing and virological confirmation are often limited to patients admitted to hospital, resulting in the under-reporting of less severe cases, which presents more frequently in younger age groups[9,218]. Surveillance intensity and the availability of testing often varies across the year: in the UK, routine laboratory testing is only active during the influenza season, resulting in undersampling of interseasonal periods of the year. More generally, the quality of surveillance data needs to be improved by further virus characterisation beyond virus positivity and typing. Only by adopting a systematic, year-around approach akin to that of Australia, where there are networks of sentinel GP practices that routinely submit samples from 25% of their ILI cases[238] for laboratory confirmation and antigenic characterisation, can suitable data be produced to inform future studies.

Based upon my empirical observations, I developed a network model to serve as a framework to investigate the interactions between host contact structure and immunity. Here, the structures considered are informed by contact studies and reflects social segregation, nuances that have been overlooked by the use of arbitrary degree distributions[124,229]. Indeed, over the course of the latter half of my PhD studies, there has been a growing emphasis on constructing epidemiological models that explicitly incorporate realistic network structures[55,108,164,279]. These innovations in quantifying spatial processes cannot occur in isolation, since I demonstrate in Chapters 5 and 6 that there is a need to account for correlations between nodal degree and susceptibility; characterising the topology of cross-immunity requires a greater understanding into how individual immunity develops and changes over the course of one's lifetime.

Given the role of children in facilitating influenza virus transmission, in-depth social contact studies are required to characterise their interactions within school/social and familial contexts, which are important in virus dissemination across the community and introductions into households respectively. This would involve fully sampling all members within individual school and household units, capturing information about higher-order structural properties, such as the presence of cliques, that were deficient from the POLYMOD study[186]. Pilot attempts have been made in recent years at fine-scale studies of contacts within households, showing how the probability of within-household ties[27,96] and risk of infection[72] vary with family types, composition and sizes. In order to generate robust data from these preliminary insights, the size of such studies will need to be expanded upon: a representative sample of individuals of all ages will be crucial in investigating how the patterns, nature and frequency of disease-causing contacts change with age. Focusing again on children, it is plausible that the number, frequency and risk of transmission is greater for direct contacts between toddlers in day-care; with increasing social development and age, interactions could

involve less direct contact, become less homogenous and more clustered within friendship groups.

Alongside investigating age-specific patterns, accounting for short-term changes to network structure would be another aspect of model development worth consideration. In dynamic networks, the less rigid contact structures could provide alternative pathways to facilitate virus transmission[246] and reduce the frailty of the network. The data required to parameterise such models will again necessitate longitudinal contact studies; there is also a need to clarify the extent and frequency of fluctuations in contacts within each social context, as well as their overall relevance, after accounting for the relatively short duration of infection and transience of seasonal epidemics. Identifying an acceptable level of model complexity suited to the research question will be vital, in order to maintain computational tractability and interpretability, as well as focussing sampling efforts. Network rewiring to account for changes in nodal degree with age may only become important in multi-season models investigating long term evolutionary patterns and changes in host immunity.

Comprehensive household/school contact surveys would also present an excellent opportunity for concurrent longitudinal serological studies: in parallel, it would be possible to associate structural properties with an individual's risk of infection and development of immunity. If children can indeed be routinely infected multiple times over consecutive seasons, this would provide additional evidence to validate the hypothesis that rates of immune waning are greater in children[206]. Furthermore, the role of children in virus dissemination through communities can be scrutinised by analysing the frequency of infection for the different members of a household and screening for correlations. It is plausible that for adults, frequent contact with children predisposes the development of broadly neutralising immunity, which could result in differences in the seroprevalence of broadly neutralising antibodies between adults living in family and non-familial group households.

With regards to the acquisition of broad neutralising immunity, exposure rate to viruses for parents could vary over the years, alongside corresponding changes in household contact patterns: daily contact rates with their parents decrease as the children become older[27,96,186], which may underlie the observed reduction in secondary attack rates for other household members[72]. By the time that middle-aged or elderly parents enter a state of permanent partial susceptibility, their centrality within the network is likely to be substantially diminished; timing-wise, this coincides with their children leaving home. This further reduces the prospect that permanently partially susceptible adults are critical for the maintenance of network connectivity and recurrent epidemics.

It has been suggested that epidemiological dynamics can be modulated by population densities: cities with greater population densities experience more interseasonal activity and

consequently smaller epidemics during the influenza season[55]. In addition to epidemiological dynamics, it is highly plausible that serological profiles would also vary between countries that differ substantially in demography and household compositions. Comparisons between countries could also generate further insights into the importance of child-adult and household contacts in the development of population cross-immunity.

Whilst additional studies will invariably produce increasingly rich and voluminous data, a key aim in mathematical modelling is to identify the minimal set of drivers necessary for a system to produce outputs that are sufficiently accurate to recapitulate some empirical observation of interest. With respect to network models, there is not only a need to further characterise the patterns in host contacts and immunity but also identify which aspects are most important to the epidemiology of seasonal influenza and reinvasion: especially in the context of producing timely and easily interpretable outputs to inform policy development, parsimony is paramount.

7.3 Conclusion

In this thesis, I set out to characterise the phylodynamical interactions between seasonal influenza viruses, climate, and human host populations at local scales. Through careful analysis of empirical data, I demonstrate that climatic factors and antigenic change have little impact on the dynamics of city-level epidemics. Additionally, I find that individual antigenic variants are capable of reinvasion over consecutive seasons. This observation brings into question the extent to which epidemics are capable of depleting susceptibles between seasons. Using a stochastic network model framework, I then demonstrated how interactions between structure in host contact patterns and cross-immunity can hinder reinvasion and evaluated the effects of different immune mechanisms, once again demonstrating that the effects of immunity depend on the connectedness of targeted individuals. In particular, I show that children are drivers of influenza virus transmission within communities due to their high degree of connectedness and levels of susceptibility. Additionally, the relatively higher rates of immune waning for children means that any accrued immunity is lost relatively rapidly, which further accentuates their importance in enabling virus reinvasion season after season. These results suggest that the population level burden of disease can be alleviated most effectively by intervention strategies that target children. Given limited public health resources, my findings help to better define the scope of future studies by highlighting critical aspects of network structure and immunity that need to be further characterised.

References

1. Antoine Allard, Pierre André Noël, Louis J. Dubé, and Babak Pourbohloul. Heterogeneous bond percolation on multitype networks with an application to epidemic dynamics. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, 79(3), 2009. ISSN 15393755. doi: 10.1103/PhysRevE.79.036113.

2. Wladimir J. Alonso, Cécile Viboud, Lone Simonsen, Eduardo W. Hirano, Luciane Z. Daufenbach, and Mark A. Miller. Seasonality of influenza in Brazil: A traveling wave from the amazon to the subtropics. *American Journal of Epidemiology*, 165(12): 1434–1442, 2007. ISSN 00029262. doi: 10.1093/aje/kwm012.

3. Christopher S. Anderson, Sandra Ortega, Francisco A. Chaves, Amelia M. Clark, Hongmei Yang, David J. Topham, and Marta L. Dediego. Natural and directed antigenic drift of the H1 influenza virus hemagglutinin stalk domain. *Scientific Reports*, 7 (1):1–19, 2017. ISSN 20452322. doi: 10.1038/s41598-017-14931-7. URL http://dx.doi.org/10.1038/s41598-017-14931-7.

4. Roy M Anderson, B Anderson, and Robert M May. *Infectious diseases of humans: dynamics and control*. Oxford university press, 1992.

5. V. Andreasen, J. Lin, and S. A. Levin. The dynamics of cocirculating influenza strains conferring partial cross-immunity. *Journal of mathematical biology*, 35(7):825–842, 1997. ISSN 03036812. doi: 10.1007/s002850050079.

6. Viggo Andreasen. Dynamics of annual influenza A epidemics with immuno-selection. *Journal of Mathematical Biology*, 46(6):504–536, 2003. ISSN 03036812. doi: 10.1007/s00285-002-0186-2.

7. Sarah F. Andrews, Yunping Huang, Kaval Kaur, Lyubov I. Popova, Irvin Y. Ho, Noel T. Pauli, Carole J.Henry Dunand, William M. Taylor, Samuel Lim, Min Huang, Xinyan Qu, Jane Hwei Lee, Marlene Salgado-Ferrer, Florian Krammer, Peter Palese, Jens Wrammert, Rafi Ahmed, and Patrick C. Wilson. Immune history profoundly affects broadly protective B cell responses to influenza. *Science Translational Medicine*, 7 (316), 2015. ISSN 19466242. doi: 10.1126/scitranslmed.aad0522.

8. Sarder Mohammed Asaduzzaman, Junling Ma, and P. van den Driessche. Estimation of Cross-Immunity Between Drifted Strains of Influenza A/H3N2. *Bulletin of Mathematical Biology*, 80(3):657–669, 2018. ISSN 15229602. doi: 10.1007/s11538-018-0395-5. URL https://doi.org/10.1007/s11538-018-0395-5.

9. Kate V. Atkinson, Lisa A. Bishop, Glenn Rhodes, Nicolas Salez, Neil R. McEwan, Matthew J. Hegarty, Julie Robey, Nicola Harding, Simon Wetherell, Robert M. Lauder, Roger W. Pickup, Mark Wilkinson, and Derek Gatherer. Influenza C in Lancaster, UK, in the winter of 2014-2015. *Scientific Reports*, 7:1–10, 2017. ISSN 20452322. doi: 10.1038/srep46578.

10. Australian Bureau of Statistics. ERP by SA2 and above (ASGS 2001), 1991 to 2016, 2017. URL http://www.abs.gov.au.

11. Australian Bureau of Statistics. 4402.0 - Childhood Education and Care, Australia, June 2017, 2018. URL https://www.abs.gov.au/ausstats/abs@.nsf/cat/4402.0.

12. A. M. Baetjer. Effect of ambient temperature and vapor pressure on cilia-mucus clearance rate. *Journal of applied physiology*, 23(4):498–504, 1967. ISSN 00218987. doi: 10.1152/jappl.1967.23.4.498.

13. Frank Ball and Peter Neal. A general model for stochastic SIR epidemics with two levels of mixing. *Mathematical Biosciences*, 180(1):73–102, 2002. ISSN 0025-5564. doi: https://doi.org/10.1016/S0025-5564(02)00125-6. URL http://www.sciencedirect.com/science/article/pii/S0025556402001256.

14. Frank Ball, Denis Mollison, and Gianpaolo Scalia-Tomba. EPIDEMICS WITH TWO LEVELS OF MIXING By Frank Ball, Denis Mollison and Gianpaolo Scalia-Tomba University of Nottingham, Heriot-Watt University and Universita La Sapienza. *Annals of Applied Probability*, 7(1):46–89, 1997.

15. Frank Ball, Tom Britton, Thomas House, Valerie Isham, Denis Mollison, Lorenzo Pellis, and Gianpaolo Scalia Tomba. Seven challenges for metapopulation models of epidemics, including households models. *Epidemics*, 10:63–67, 2015. ISSN 18780067. doi: 10.1016/j.epidem.2014.08.001. URL http://dx.doi.org/10.1016/j.epidem.2014.08.001.

16. Sébastien Ballesteros, Elisabeta Vergu, and Bernard Cazelles. Influenza A gradual and epochal evolution: Insights from simple models. *PLoS ONE*, 4(10), 2009. ISSN 19326203. doi: 10.1371/journal.pone.0007426.

17. Shweta Bansal and Lauren Ancel. The impact of past epidemics on future disease dynamics. *Journal of Theoretical Biology*, 309:176–184, 2012. ISSN 0022-5193. doi: 10.1016/j.jtbi.2012.06.012. URL http://dx.doi.org/10.1016/j.jtbi.2012.06.012.

18. Shweta Bansal, Bryan T. Grenfell, and Lauren Ancel Meyers. When individual behaviour matters: Homogeneous and network models in epidemiology. *Journal of the Royal Society Interface*, 4(16):879–891, 2007. ISSN 17425689. doi: 10.1098/rsif.2007.1100.

19. Shweta Bansal, Jonathan Read, Babak Pourbohloul, and Lauren Ancel Meyers. The dynamic nature of contact networks in infectious disease epidemiology. *Journal of Biological Dynamics*, 4(5):478–489, sep 2010. ISSN 1751-3758. doi: 10.1080/17513758.2010.503376. URL http://www.tandfonline.com/doi/abs/10.1080/17513758.2010.503376.

20. I. G. Barr, L Cui, N Komadina, R T Lee, R T Lin, Y Deng, N Caldwell, R Shaw, S. Maurer-Stroh, Barr I.G., Cui L., Komadina N., Lin Lee R.T., Lin Lee R.T., Deng Y., Caldwell N., and Shaw R. A new pandemic influenza A(H1N1) genetic variant predominated in the winter 2010 influenza season in Australia, New Zealand and Singapore. *Eurosurveillance*, 15(42):1–6, 2010. ISSN 1025-496X. doi: 10.2807/ese.15.42.19692-en. URL http://www.eurosurveillance.org/images/dynamic/EE/V15N42/art19692.pdf{%}0Ahttp://ovidsp.ovid.com/ovidweb.cgi?T=JS{&}PAGE=reference{&}D=emed12{&}NEWS=N{&}AN=359896415.

21. Ian G. Barr, D. Vijaykrishna, and S. G. Sullivan. Differential age susceptibility to influenza B/victoria lineage viruses in the 2015 Australian influenza season. *Eurosurveillance*, 21(4):1–9, 2016. ISSN 15607917. doi: 10.2807/1560-7917.ES.2016.21.4.30118.

22. Trevor Bedford, Sarah Cobey, Peter Beerli, and Mercedes Pascual. Global migration dynamics underlie evolution and persistence of human influenza a (H3N2). *PLoS Pathogens*, 6(5), 2010. ISSN 15537366. doi: 10.1371/journal.ppat.1000918.

23. Trevor Bedford, Andrew Rambaut, and Mercedes Pascual. Canalization of the evolutionary trajectory of the human influenza virus. *BMC biology*, 10(1):38, 2012. ISSN 1741-7007. doi: 10.1186/1741-7007-10-38. URL http://bmcbiol.biomedcentral.com/articles/10.1186/1741-7007-10-38{%}5Cnhttp://www.biomedcentral.com/1741-7007/10/38.

24. Trevor Bedford, Marc A. Suchard, Philippe Lemey, Gytis Dudas, Victoria Gregory, Alan J. Hay, John W. McCauley, Colin A. Russell, Derek J. Smith, and Andrew Rambaut. Integrating influenza antigenic dynamics with molecular evolution. *eLife*, 3(3):1–26, feb 2014. ISSN 2050-084X. doi: 10.7554/eLife.01914. URL https://elifesciences.org/articles/01914.

25. Trevor Bedford, Steven Riley, Ian G. Barr, Shobha Broor, Mandeep Chadha, Nancy J. Cox, Rodney S. Daniels, C. Palani Gunasekaran, Aeron C. Hurt, Anne Kelso, Alexander Klimov, Nicola S. Lewis, Xiyan Li, John W. McCauley, Takato Odagiri, Varsha Potdar, Andrew Rambaut, Yuelong Shu, Eugene Skepner, Derek J. Smith, Marc a. Suchard, Masato Tashiro, Dayan Wang, Xiyan Xu, Philippe Lemey, and Colin a. Russell. Global circulation patterns of seasonal influenza viruses vary with antigenic drift. *Nature*, 523(7559):217–20, 2015. ISSN 0028-0836. doi: 10.1038/nature14460. URL http://www.ncbi.nlm.nih.gov/pubmed/26053121.

26. Robert B. Belshe, William C. Gruber, Paul M. Mendelman, Harshvardhan B. Mehta, Kutubuddin Mahmood, Keith Reisinger, John Treanor, Ken Zangwill, Frederick G. Hayden, David I. Bernstein, Karen Kotloff, James King, Pedro A. Piedra, Stan L. Block, Lihan Yan, and Mark Wolff. Correlates of Immune Protection Induced by Live, Attenuated, Cold-Adapted, Trivalent, Intranasal Influenza Virus Vaccine. *The Journal of Infectious Diseases*, 181(3):1133–1137, 2000. ISSN 0022-1899. doi: 10.1086/315323.

27. Guillaume Béraud, Sabine Kazmercziak, Philippe Beutels, Daniel Levy-Bruhl, Xavier Lenne, Nathalie Mielcarek, Yazdan Yazdanpanah, Pierre Yves Boëlle, Niel Hens, and Benoit Dervaux. The French connection: The first large population-based contact survey in France relevant for the spread of infectious diseases. *PLoS ONE*, 10(7):1–22, 2015. ISSN 19326203. doi: 10.1371/journal.pone.0133203.

28. Gordon R Bloomberg. The influence of environment, as represented by diet and air pollution, upon incidence and prevalence of wheezing illnesses in young children. *Current opinion in allergy and clinical immunology*, 11(2):144–149, apr 2011. ISSN 1473-6322 (Electronic). doi: 10.1097/ACI.0b013e3283445950.

29. Jacob Bock Axelsen, Rami Yaari, Bryan T. Grenfell, and Lewi Stone. Multiannual forecasting of seasonal influenza dynamics reveals climatic and evolutionary drivers. *Proceedings of the National Academy of Sciences*, 111(26):9538–9542, 2014. ISSN 0027-8424. doi: 10.1073/pnas.1321656111. URL http://www.pnas.org/lookup/doi/10.1073/pnas.1321656111.

30. E. Brooks-Pollock, M. C.M. de Jong, M. J. Keeling, D. Klinkenberg, and J. L.N. Wood. Eight challenges in modelling infectious livestock diseases. *Epidemics*, 10:1–5, 2015. ISSN 18780067. doi: 10.1016/j.epidem.2014.08.005. URL http://dx.doi.org/10.1016/j.epidem.2014.08.005.

31. David L. Buckeridge, Howard Burkom, Murray Campbell, William R. Hogan, and Andrew W. Moore. Algorithms for rapid outbreak detection: A research synthesis. *Journal of Biomedical Informatics*, 38(2):99–113, 2005. ISSN 15320464. doi: 10.1016/j.jbi.2004.11.007.

32. Carter T Butts, David R Hunter, Martina Morris, Pavel N Krivitsky, and Zack Almquist. Introduction to Random Graph (ERG or p*) modeling with ergm, 2017. URL https://cran.r-project.org/web/packages/ergm/vignettes/ergm.pdf.

33. Saverio Caini, Winston Andrade, Selim Badur, Angel Balmaseda, Amal Barakat, Antonino Bella, Abderrahman Bimohuen, Lynnette Brammer, Joseph Bresee, Alfredo Bruno, Leticia Castillo, Meral A. Ciblak, Alexey W. Clara, Cheryl Cohen, Jeffery Cutter, Coulibaly Daouda, Celina De Lozano, Domenica De Mora, Kunzang Dorji, Gideon O. Emukule, Rodrigo A. Fasce, Luzhao Feng, Walquiria Aparecida Ferreira De Almeida, Raquel Guiomar, Jean Michel Heraud, Olha Holubka, Q. Sue Huang, Herve A. Kadjo, Lyazzat Kiyanbekova, Herman Kosasih, Gabriela Kusznierz, Jenny Lara, Ming Li, Liza Lopez, Phuong Vu Mai Hoang, Cláudio Maierovitch Pessanha Henriques, Maria Luisa Matute, Alla Mironenko, Brechla Moreno, Joshua A. Mott, Richard Njouom, Nurhayati, Akerke Ospanova, Rhonda Owen, Richard Pebody, Kate Pennington, Simona Puzelli, Mai Thi Quynh Le, Norosoa Harline Razanajatovo, Ana Rodrigues, Juan Manuel Rudi, Raymond Tzer Pin Lin, Marietjie Venter, Marie Astrid Vernet, Sonam Wangchuk, Juan Yang, Hongjie Yu, Maria Zambon, François Schellevis, John Paget, Sonam Gyeltshen, Pema Euden, Guy Vernet, Patricia Bustos, Joanna Ellis, Isabella Donatelli, Caterina Rizzo, Julia Guillebaud, Laurence Randrianasolo, Batlazar Nunes, and Pedro Pechirra. Temporal patterns of influenza A and B in tropical and temperate countries: What are the lessons for influenza vaccination? *PLoS ONE*, 11(3):1–15, 2016. ISSN 19326203. doi: 10.1371/journal.pone.0152310.

34. Saverio Caini, Madelon Kroneman, Therese Wiegers, Clotilde El Guerche-Séblain, and John Paget. Clinical characteristics and severity of influenza infections by virus type, subtype, and lineage: A systematic literature review. *Influenza and other Respiratory Viruses*, 12(6):780–792, 2018. ISSN 17502659. doi: 10.1111/irv.12575.

35. C. Castillo-Chavez, H. W. Hethcote, V. Andreasen, S. A. Levin, and W. M. Liu. Epidemiological models with age structure, proportionate mixing, and cross-immunity. *Journal of Mathematical Biology*, 27(3):233–258, 1989. ISSN 03036812. doi: 10.1007/BF00275810.

36. S Cauchemez, A J Valleron, P Y Boelle, A Flahault, and N M Ferguson. Estimating the impact of school closure on influenza transmission from Sentinel data. *Nature*, 452 (1476-4687 (Electronic)):750–754, 2008. ISSN 1476-4687. doi: 10.1038/nature06732. URL nature.com/articles/nature06732.

37. Ning Chai, Lee R. Swem, Mike Reichelt, Haiyin Chen-Harris, Elizabeth Luis, Summer Park, Ashley Fouts, Patrick Lupardus, Thomas D. Wu, Olga Li, Jacqueline McBride, Michael Lawrence, Min Xu, and Man Wah Tan. Two Escape Mechanisms of Influenza A Virus to a Broadly Neutralizing Stalk-Binding Antibody. *PLoS Pathogens*, 12(6): 1–29, 2016. ISSN 15537374. doi: 10.1371/journal.ppat.1005702.

38. DL Dennis L. Chao, HM Elizabeth, IM Longini, M. Elizabeth Halloran, and Ira M. Longini, Jr. School opening dates predict pandemic influenza A (H1N1) epidemics in the USA. *J Infect Dis*, 202(6):877–880, 2010. ISSN 0022-1899. doi: 10.1086/655810. School. URL https://academic.oup.com/jid/article-lookup/doi/10.1086/655810.

39. Vivek Charu, Scott Zeger, Julia Gog, Ottar N. Bjørnstad, Stephen Kissler, Lone Simonsen, Bryan T. Grenfell, and Cécile Viboud. Human mobility and the spatial transmission of influenza in the United States. *PLoS Computational Biology*, 13(2):1–23, 2017. ISSN 15537358. doi: 10.1371/journal.pcbi.1005382.

40. Ishanu Chattopadhyay, Emre Kiciman, Joshua W Elliott, and Jeffrey L Shaman. Conjunction of factors triggering waves of seasonal influenza. pages 1–44, 2018.

41. Ming Wong Chit, Lin Yang, Pan Chan King, Gabriel M. Leung, Kwok H. Chan, Yi Guan, Hing Lam Tai, Anthony Johnson Hedley, and Joseph S.M. Peiris. Influenza-associated hospitalization in a subtropical city. *PLoS Medicine*, 3(4):485–492, 2006. ISSN 15491277. doi: 10.1371/journal.pmed.0030121.

42. Susan S. Chiu, Kwok-Hung Chan, Hong Chen, Betty W. Young, Wilina Lim, Wilfred Hing Sang Wong, Yu Lung Lau, and J. S. Malik Peiris. Virologically Confirmed Population-Based Burden of Hospitalization Caused by Influenza A and B among Children in Hong Kong. *Clinical Infectious Diseases*, 49(7):1016–1021, 2009. ISSN 1058-4838. doi: 10.1086/605570.

43. G. CHOWELL, M. A. MILLER, and C. VIBOUD. Seasonal influenza in the United States, France, and Australia: transmission and prospects for control. *Epidemiology and Infection*, 136(06):852–864, 2008. ISSN 0950-2688. doi: 10.1017/S0950268807009144. URL http://www.journals.cambridge.org/abstract{_}S0950268807009144.

44. Sarah Cobey and Katia Koelle. Capturing escape in infectious disease dynamics. *Trends in Ecology and Evolution*, 23(10):572–577, 2008. ISSN 01695347. doi: 10.1016/j.tree.2008.06.008.

45. Miriam Cohen, Robert Schooley, Miriam Cohen, Xing-quan Zhang, Hooman P Senaati, Hui-wen Chen, Nissi M Varki, and Robert T Schooley. Influenza A penetrates host mucus by cleaving sialic acids with neuraminidase Influenza A penetrates host mucus by cleaving sialic acids with neuraminidase. *Virology Journal*, 10(1):1, 2013. ISSN Virology Journal. doi: 10.1186/1743-422X-10-321. URL VirologyJournal.

46. Ben S Cooper, Richard J Pitman, W John Edmunds, and Nigel J Gay. Delaying the international spread of pandemic influenza. *PLoS medicine*, 3(6):e212, jun 2006. ISSN 1549-1676 (Electronic). doi: 10.1371/journal.pmed.0030212.

47. Davide Corti, Amorsolo L Suguitan Jr, Debora Pinna, Chiara Silacci, Blanca M Fernandez-rodriguez, Fabrizia Vanzetta, Celia Santos, Catherine J Luke, Fernando J Torres-velez, Nigel J Temperton, Robin a Weiss, Federica Sallusto, Kanta Subbarao, and Antonio Lanzavecchia. Heterosubtypic neutralizing antibodies are produced by individuals immunized with a seasonal influenza vaccine. *Journal of Clinical Investigation*, 120(5), 2010. doi: 10.1172/JCI41902DS1.

48. D. Costagliola, A. Flahault, D. Galinec, P. Garnerin, J. Menares, and A. J. Valleron. A routine tool for detection and assessment of epidemics of influenza-like syndromes in France. *American Journal of Public Health*, 81(1):97–99, 1991. ISSN 00900036. doi: 10.2105/AJPH.81.1.97.

49. Laurent Coudeville, Fabrice Bailleux, Benjamin Riche, Françoise Megas, Philippe Andre, and René Ecochard. Relationship between haemagglutination-inhibiting antibody titres and clinical protection against influenza: Development and application of a bayesian random-effects model. *BMC Medical Research Methodology*, 10, 2010. ISSN 14712288. doi: 10.1186/1471-2288-10-18.

50. Benjamin J. Cowling, Irene O.L. Wong, Lai Ming Ho, Steven Riley, and Gabriel M. Leung. Methods for monitoring influenza surveillance data. *International Journal of Epidemiology*, 35(5):1314–1321, 2006. ISSN 03005771. doi: 10.1093/ije/dyl162.

51. Benjamin J. J. Cowling, Sophia Ng, Edward S. K. S. K. Ma, Calvin K. Y. K. Y. Cheng, Winnie Wai, Vicky J. J. Fang, Kwok-hung Kwok-Hung Chan, Dennis K. M. K. M. Ip, Susan S. S. Chiu, J. S. Malik S. Malik Peiris, and Gabriel M. M. Leung. Protective Efficacy of Seasonal Influenza Vaccination against Seasonal and Pandemic Influenza Virus Infection during 2009 in Hong Kong. *Clinical Infectious Diseases*, 51(12):1370–1379, dec 2010. ISSN 1058-4838. doi: 10.1086/657311. URL https://academic.oup.com/cid/article-lookup/doi/10.1086/657311.

52. Benjamin J. Cowling, Sophia Ng, Edward S. K. Ma, Calvin K. Y. Cheng, Winnie Wai, Vicky J. Fang, Kwok-Hung Chan, Dennis K. M. Ip, Susan S. Chiu, J. S. Malik Peiris, and Gabriel M. Leung. Protective Efficacy of Seasonal Influenza Vaccination against Seasonal and Pandemic Influenza Virus Infection during 2009 in Hong Kong. *Clinical Infectious Diseases*, 51(12):1370–1379, dec 2010. ISSN 1058-4838. doi: 10.1086/657311.

53. Benjamin D Dalziel, Kai Huang, Jemma L Geoghegan, Nimalan Arinaminpathy, Edward J Dubovi, Bryan T Grenfell, Stephen P Ellner, Edward C Holmes, and Colin R Parrish. Contact Heterogeneity, Rather Than Transmission Efficiency, Limits the Emergence and Spread of Canine Influenza Virus. *PLoS Pathogens*, 10(10):

e1004455, oct 2014. ISSN 15537374. doi: 10.1371/journal.ppat.1004455. URL https://dx.plos.org/10.1371/journal.ppat.1004455.

54. Benjamin D. Dalziel, Kai Huang, Jemma L. Geoghegan, Nimalan Arinaminpathy, Edward J. Dubovi, Bryan T. Grenfell, Stephen P. Ellner, Edward C. Holmes, and Colin R. Parrish. Contact Heterogeneity, Rather Than Transmission Efficiency, Limits the Emergence and Spread of Canine Influenza Virus. *PLoS Pathogens*, 10(10), oct 2014. ISSN 15537374. doi: 10.1371/journal.ppat.1004455.

55. Benjamin D. Dalziel, Stephen Kissler, Julia R. Gog, Cecile Viboud, Ottar N. Bjørnstad, C. Jessica E. Metcalf, and Bryan T. Grenfell. Urbanization and humidity shape the intensity of influenza epidemics in U.S. cities. *Science*, 362(6410):75–79, 2018. ISSN 0036-8075. doi: 10.1126/science.aat6030. URL http://www.sciencemag.org/lookup/doi/10.1126/science.aat6030.

56. T. DaPalma, B. P. Doonan, N. M. Trager, and L. M. Kasman. A systematic approach to virus-virus interactions. *Virus Research*, 149(1):1–9, 2010. ISSN 01681702. doi: 10.1016/j.virusres.2010.01.002.

57. F M DAVENPORT and A V HENNESSY. A serologic recapitulation of past experiences with influenza A; antibody response to monovalent vaccine. *The Journal of experimental medicine*, 104(1):85–97, 1956. ISSN 0022-1007. doi: 10.1084/jem.104.1.85. URL http://www.ncbi.nlm.nih.gov/pubmed/13332182{%}5Cnhttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC2136633.

58. Fred M Davenport, Albert V Hennessy, Thomas Francis Jr, and With the Technical Assistance of Phyllis Fabisch. Epidemiologic and immunologic significance of age distribution of antibody to antigenic variants of influenza virus. *The Journal of experimental medicine*, 98(6):641–656, 1953.

59. Fred M Davenport, Albert V Hennessy, and With the Technical Assistance of Phyllis H. Fabisch. Predetermination by infection and by vaccination of antibody response to influenza virus vaccines. *The Journal of experimental medicine*, 106(6):835–850, 1957.

60. FredM Davenport, AlbertV Hennessy, C H Stuart-Harris, T Francis Jr, and Others. Epidemiology of Influenza. Comparative Serological Observations in England and the United States. *Lancet*, pages 469–474, 1955.

61. M. L. Davey and D. Reid. Relationship of air temperature to outbreaks of influenza. *British journal of preventive & social medicine*, 26(1):28–32, 1972. ISSN 00071242. doi: 10.1136/jech.26.1.28.

62. Robert E. Davis, Colleen E. Rossier, and Kyle B. Enfield. The Impact of Weather on Influenza and Pneumonia Mortality in New York City, 1975–2002: A Retrospective Study. *PLoS ONE*, 7(3):e34091, mar 2012. ISSN 1932-6203. doi: 10.1371/journal.pone.0034091. URL https://dx.plos.org/10.1371/journal.pone.0034091.

63. Lise Denoeud, Clément Turbelin, Séverine Ansart, Alain Jacques Valleron, Antoine Flahault, and Fabrice Carrat. Predicting pneumonia and influenza mortality from morbidity data. *PLoS ONE*, 2(5):3–6, 2007. ISSN 19326203. doi: 10.1371/journal.pone.0000464.

64. Ethan R. Deyle, M. Cyrus Maher, Ryan D. Hernandez, Sanjay Basu, and George Sugihara. Global environmental drivers of influenza. *Proceedings of the National Academy of Sciences*, 113(46):13081–13086, 2016. ISSN 0027-8424. doi: 10.1073/pnas.1607747113. URL http://www.pnas.org/lookup/doi/10.1073/pnas.1607747113.

65. Richard C Dicker, Fatima Coronado, Denise Koo, and R Gibson Parrish. Principles of epidemiology in public health practice; an introduction to applied epidemiology and biostatistics. 2006.

66. Michael B. Doud, Scott E. Hensley, and Jesse D. Bloom. Complete mapping of viral escape from neutralizing antibodies. *PLoS Pathogens*, 13(3):1–20, 2017. ISSN 15537374. doi: 10.1371/journal.ppat.1006271.

67. Ken T.D. Eames and Matt J. Keeling. Modeling dynamic and network heterogeneities in the spread of sexually transmitted diseases. *Proceedings of the National Academy of Sciences of the United States of America*, 99(20):13330–13335, 2002. ISSN 00278424. doi: 10.1073/pnas.202244299.

68. R. Eccles. An explanation for the seasonality of acute upper respiratory tract viral infections. *Acta Oto-Laryngologica*, 122(2):183–191, 2002. ISSN 00016489. doi: 10.1080/00016480252814207.

69. W P Jr Edmondson, R Rothenberg, P W White, and J M Jr Gwaltney. A comparison of subcutaneous, nasal, and combined influenza vaccination. II. Protection against natural challenge. *American journal of epidemiology*, 93(6):480–486, jun 1971. ISSN 0002-9262 (Print). doi: 10.1093/oxfordjournals.aje.a121282.

70. Maryna Eichelberger, Hana Golding, Maureen Hess, Jerry Weir, Kanta Subbarao, Catherine J Luke, Martin Friede, and David Wood. FDA/NIH/WHO public workshop on immune correlates of protection against influenza A viruses in support of pandemic vaccine development, Bethesda, Maryland, US, December 10-11, 2007., aug 2008. ISSN 0264-410X (Print).

71. Ali H. Ellebedy, Florian Krammer, Gui Mei Li, Matthew S. Miller, Christopher Chiu, Jens Wrammert, Cathy Y. Chang, Carl W. Davis, Megan McCausland, Rivka Elbein, Srilatha Edupuganti, Paul Spearman, Sarah F. Andrews, Patrick C. Wilson, Adolfo Garciá-Sastre, Mark J. Mulligan, Aneesh K. Mehta, Peter Palese, and Rafi Ahmed. Induction of broadly cross-reactive antibody responses to the influenza HA stem region following H5N1 vaccination in humans. *Proceedings of the National Academy of Sciences of the United States of America*, 111(36):13133–13138, 2014. ISSN 10916490. doi: 10.1073/pnas.1414070111.

72. Akira Endo, Mitsuo Uchida, Adam J. Kucharski, and Sebastian Funk. Fine-scale family structure shapes influenza transmission risk in households: Insights from primary schools in Matsumoto city, 2014/15. *PLoS Computational Biology*, 15(12):1–18, 2019. ISSN 15537358. doi: 10.1371/journal.pcbi.1007589.

73. Suzanne L. Epstein. Prior H1N1 Influenza Infection and Susceptibility of Cleveland Family Study Participants during the H2N2 Pandemic of 1957: An Experiment of Nature. *The Journal of Infectious Diseases*, 193(1):49–53, 2006. ISSN 0022-1899. doi: 10.1086/498980.

74. C P Farrington, N J Andrews, A D Beale, M A Catchpole, Royal Statistical Society, and Royal Statistical Society. A Statistical Algorithm for the Early Detection of Outbreaks of Infectious Disease. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 159(3):547–563, jul 1996. ISSN 09641998, 1467985X. doi: 10.2307/2983331. URL http://www.jstor.org/stable/2983331.

75. Paddy Farrington and Nick Andrews. Outbreak detection: application to infectious disease surveillance. In Ron Brookmeyer and Donna F Stroup, editors, *Monitoring the Health of Populations: Statistical Principles and Methods for Public Health Surveillance*, pages 203–231. OUP USA, New York, NY, USA, dec 2003. ISBN 9780199682676. doi: 10.1093/acprof. URL http://oro.open.ac.uk/22646/.

76. Neil M. Ferguson, Alison P. Galvani, and Robin M. Bush. Ecological and immunological determinants of influenza evolution. *Nature*, 422(6930):428–433, 2003. ISSN 0028-0836. doi: 10.1038/nature01509. URL http://www.nature.com/doifinder/10.1038/nature01509.

77. Michael Fernandes, Logan Walls, Sean Munson, Jessica Hullman, and Matthew Kay. Uncertainty Displays Using Quantile Dotplots or CDFs Improve Transit Decision-Making. In *2018 CHI Conference on Human Factors in Computing Systems*, pages 1–12, 2018. doi: 10.1145/3173574.3173718. URL https://idl.cs.washington.edu/files/2018-UncertaintyBus-CHI.pdf.

78. Matthew J. Ferrari, Shweta Bansal, Lauren A. Meyers, and Ottar N Bjørnstad. Network frailty and the geometry of herd immunity. *Proceedings of the Royal Society B: Biological Sciences*, 273(1602):2743–2748, nov 2006. ISSN 0962-8452. doi: 10.1098/rspb.2006.3636. URL https://royalsocietypublishing.org/doi/10.1098/rspb.2006.3636.

79. Brian S. Finkelman, Cécile Viboud, Katia Koelle, Matthew J. Ferrari, Nita Bharti, and Bryan T. Grenfell. Global patterns in seasonal activity of influenza A/H3N2, A/H1N1, and B from 1997 to 2005: Viral coexistence and latitudinal gradients. *PLoS ONE*, 2 (12), 2007. ISSN 19326203. doi: 10.1371/journal.pone.0001296.

80. K A Fitzner, S M McGhee, A J Hedley, and K F Shortridge. Influenza surveillance in Hong Kong: results of a trial Physician Sentinel Programme. *Hong Kong medical journal = Xianggang yi xue za zhi*, 5(1):87–94, 1999. ISSN 1024-2708.

81. J M Fonville, S H Wilks, S L James, A Fox, M Ventresca, M Aban, L Xue, T C Jones, N M H Le, Q T Pham, N D Tran, Y Wong, A Mosterin, L C Katzelnick, D Labonte, T T Le, G van der Net, E Skepner, C A Russell, T D Kaplan, G F Rimmelzwaan, N Masurel, J C de Jong, A Palache, W E P Beyer, Q M Le, T H Nguyen, H F L Wertheim, A C Hurt, A D M E Osterhaus, I G Barr, R A M Fouchier, P W Horby, and D J Smith. Antibody landscapes after influenza virus infection or vaccination. *Science (New York, N.Y.)*, 346(6212):996–1000, 2014. ISSN 1095-9203. doi: 10.1126/science.1256427. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4246172{&}tool=pmcentrez{&}rendertype=abstract.

82. Hall C.E Fox JP, Cooney MK and Foy HM. Influenza Virus Imfections in Seattle Families, 1975-1979, 1982.

83. Thomas Francis. On the Doctrine of Original Antigenic Sin. *Proceedings of the American Philosophical Society*, 104(6):572–578, sep 1960. ISSN 0003049X. URL http://www.jstor.org/stable/985534.

84. Christophe Fraser, Christl A Donnelly, Simon Cauchemez, William P Hanage, Maria D Van Kerkhove, T Déirdre Hollingsworth, Jamie Griffin, Rebecca F Baggaley, Helen E Jenkins, Emily J Lyons, Thibaut Jombart, Wes R Hinsley, Nicholas C Grassly, Francois Balloux, Azra C Ghani, Neil M Ferguson, Andrew Rambaut, Oliver G Pybus, Hugo Lopez-Gatell, Celia M Alpuche-Aranda, Ietza Bojorquez Chapela, Ethel Palacios Zavala, Dulce Ma Espejo Guevara, Francesco Checchi, Erika Garcia, Stephane Hugonnet, and Cathy Roth. Pandemic potential of a strain of influenza A (H1N1): early findings. *Science (New York, N.Y.)*, 324(5934):1557–1561, jun 2009. ISSN 1095-9203 (Electronic). doi: 10.1126/science.1176062.

85. Laura Fumanelli, Marco Ajelli, Piero Manfredi, Alessandro Vespignani, and Stefano Merler. Inferring the Structure of Social Contacts from Demographic Data in the Analysis of Infectious Diseases Spread. *PLoS Computational Biology*, 8(9):35–39, 2012. ISSN 1553734X. doi: 10.1371/journal.pcbi.1002673.

86. Sylvain Gandon, Troy Day, C. Jessica E. Metcalf, and Bryan T. Grenfell. Forecasting Epidemiological and Evolutionary Dynamics of Infectious Diseases. *Trends in Ecology and Evolution*, 31(10):776–788, 2016. ISSN 01695347. doi: 10.1016/j.tree.2016.07.010. URL http://dx.doi.org/10.1016/j.tree.2016.07.010.

87. Rebecca Garten, Lenee Blanton, Anwar Isa, Abd Elal, Noreen Alabi, John Barnes, Matthew Biggerstaff, Lynnette Brammer, Alicia P Budd, Erin Burns, Charisse N Cummings, Todd Davis, Shikha Garg, Larisa Gubareva, Yunho Jang, Krista Kniss, Natalie Kramer, Stephen Lindstrom, Desiree Mustaquim, Alissa O Halloran, Wendy Sessions, Calli Taylor, Xiyan Xu, Vivien G Dugan, Alicia M Fry, David E Wentworth, Jacqueline Katz, and Daniel Jernigan. Update : Influenza Activity in the United States During the 2017 – 18 Season and Composition of the 2018 – 19 Influenza Vaccine Antigenic and Genetic Characterization of Influenza Viruses. 67(22), 2018.

88. Kellie Gavin, Rhonda Owen, and Ian G Barr. Communicable Diseases Intelligence Volume 41 Issue 4; December 2017 - Annual report of the National Influenza Surveillance Scheme, 2010. *Cdi*, 41(4):348–368, 2017.

89. Nicholas Geard, Kathryn Glass, James M. McCaw, Emma S. McBryde, Kevin B. Korb, Matt J. Keeling, and Jodie McVernon. The effects of demographic change on disease transmission and vaccine impact in a household structured population. *Epidemics*, 13:56–64, 2015. ISSN 18780067. doi: 10.1016/j.epidem.2015.08.002. URL http://dx.doi.org/10.1016/j.epidem.2015.08.002.

90. Andrew Gelman. Scaling regression inputs by dividing by two standard deviations. *Statistics in Medicine*, (27):2865–2873, 2007. doi: 10.1002/sim.3107. URL http://www.stat.columbia.edu/{~}gelman/research/published/standardizing7.pdf.

91. Andrew Gelman, Aleks Jakulin, Maria Grazia Pittau, and Yu-Sung Su. A weakly informative default prior distribution for logistic and other regression models. *The Annals of Applied Statistics*, 2(4):1360–1383, dec 2008. ISSN 1932-6157. doi: 10.1214/08-AOAS191. URL http://projecteuclid.org/euclid.aoas/1231424214.

92. Jemma L. Geoghegan, Aldo F. Saavedra, Sebastián Duchêne, Sheena Sullivan, Ian Barr, and Edward C. Holmes. Continental synchronicity of human influenza virus epidemics despite climactic variation. *PLoS Pathogens*, 14(1):1–16, 2018. ISSN 15537374. doi: 10.1371/journal.ppat.1006780.

93. V. E. GILBERT, C. T. CARAWAY, J. M. BRUCE, and W. J. MOGABGAB. Epidemic recurrence of Asian Influenza in Louisiana, 1959-1960. *American journal of public health and the nation's health*, 52(9):1432–1443, 1962. ISSN 00029572. doi: 10.2105/AJPH.52.9.1432.

94. Peter W Gill and Alan M Murphy. Naturally acquired immunity to influenza type A: a further prospective study. *Medical Journal of Australia*, 2(23):761–765, 1977.

95. W. Paul Glezen, Jordana K. Schmier, Carrie M. Kuehn, Kellie J. Ryan, and John Oxford. The burden of influenza B: A structured literature review. *American Journal of Public Health*, 103(3):43–51, 2013. ISSN 00900036. doi: 10.2105/AJPH.2012.301137.

96. Nele Goeyvaerts, Eva Santermans, Gail Potter, Andrea Torneri, Kim Van Kerckhove, Lander Willem, Marc Aerts, Philippe Beutels, and Niel Hens. Household members do not contact each other at random: Implications for infectious disease modelling. *Proceedings of the Royal Society B: Biological Sciences*, 285(1893), 2018. ISSN 14712954. doi: 10.1098/rspb.2018.2201.

97. J R Gog and J Swinton. A status based approach to multiple strain dynamics. *J Math Biol*, 44:169–184, 2002. ISSN 0303-6812. doi: 10.1007/s002850100120.

98. Julia R Gog and Bryan T Grenfell. Dynamics and selection of many-strain pathogens. *Proceedings of the National Academy of Sciences of the United States of America*, 99(26):17209–14, 2002. ISSN 0027-8424. doi: 10.1073/pnas.252512799. URL http://www.pnas.org/content/99/26/17209.short.

99. Edward Goldstein, Sarah Cobey, Saki Takahashi, Joel C. Miller, and Marc Lipsitch. Predicting the epidemic sizes of influenza A/H1N1, A/H3N2, and B: A statistical method. *PLoS Medicine*, 8(7):1–12, 2011. ISSN 15491277. doi: 10.1371/journal.pmed.1001051.

100. Sigrid Gouma, Kangchon Kim, Madison Weirick, Megan E. Gumina, Angela Branche, David J Topham, Emily Toth Martin, Arnold Monto, Sarah Cobey, and Scott E. Hensley. Middle-aged individuals may be in a perpetual state of H3N2 influenza virus susceptibility. *medRxiv*, (215):2020.01.09.20017038, 2020. doi: 10.1101/2020.01.09.20017038. URL https://www.medrxiv.org/content/10.1101/2020.01.09.20017038v1.

101. Stephen E Graham and Thomas McCurdy. Developing meaningful cohorts for human exposure models. *Journal of exposure analysis and environmental epidemiology*, 14(1):23–43, jan 2004. ISSN 1053-4245 (Print). doi: 10.1038/sj.jea.7500293.

102. Sharon K. Greene, Edward L. Ionides, and Mark L. Wilson. Patterns of influenza-associated mortality among US elderly by geographic region and virus subtype, 1968-1998. *American Journal of Epidemiology*, 163(4):316–326, 2006. ISSN 00029262. doi: 10.1093/aje/kwj040.

103. B. T. Grenfell and R. M. Anderson. The estimation of age-related rates of infection from case notifications and serological data. *Journal of Hygiene*, 95(2):419–436, 1985. ISSN 00221724. doi: 10.1017/S0022172400062859.

104. Bryan T Grenfell. Unifying the Epidemiological and Evolutionary Dynamics of Pathogens. *Science*, 303(5656):327–332, 2004. ISSN 0036-8075. doi: 10.1126/science.1090727. URL http://www.ncbi.nlm.nih.gov/pubmed/14726583{%}5Cnhttp://www.sciencemag.org/cgi/doi/10.1126/science.1090727.

105. Upma Gulati, Wenxin Wu, Shelly Gulati, Kshama Kumari, Joseph L. Waner, and Gillian M. Air. Mismatched hemagglutinin and neuraminidase specificities in recent human H3N2 influenza viruses. *Virology*, 339(1):12–20, 2005. ISSN 00426822. doi: 10.1016/j.virol.2005.05.009.

106. Sunetra Gupta, Neil Ferguson, and Roy Anderson. Chaos, persistence, and evolution of strain structure in antigenically diverse infectious agents. *Science*, 280(5365):912–915, 1998. ISSN 00368075. doi: 10.1126/science.280.5365.912.

107. Brian P. Hanley and Birthe Borup. Aerosol influenza transmission risk contours: A study of humid tropics versus winter temperate zone. *Virology Journal*, 7:1–18, 2010. ISSN 1743422X. doi: 10.1186/1743-422X-7-98.

108. David J. Haw, Derek A.T. T. Cummings, Justin Lessler, Henrik Salje, Jonathan M. Read, and Steven Riley. Differential mobility and local variation in infection attack rate. *PLOS Computational Biology*, 15(1):e1006600, jan 2019. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1006600. URL http://dx.plos.org/10.1371/journal.pcbi.1006600.

109. J. H. Hemmes, K. C. Winkler, and S. M. Kool. Virus survival as a seasonal factor in influenza and poliomyelitis. *Nature*, 188(1):430–431, 1960. ISSN 00036072. doi: 10.1007/BF02538737.

110. Carole Henry, Nai Ying Zheng, Min Huang, Alexandra Cabanov, Karla Thatcher Rojas, Kaval Kaur, Sarah F. Andrews, Anna Karin E. Palm, Yao Qing Chen, Yang Li, Katerina Hoskova, Henry A. Utset, Marcos C. Vieira, Jens Wrammert, Rafi Ahmed, Jeanne Holden-Wiltse, David J. Topham, John J. Treanor, Hildegund C. Ertl, Kenneth E. Schmader, Sarah Cobey, Florian Krammer, Scott E. Hensley, Harry Greenberg, Xiao Song He, and Patrick C. Wilson. Influenza Virus Vaccination Elicits Poorly Adapted B Cell Responses in Elderly Individuals. *Cell Host and Microbe*, 25(3):357–366.e6, 2019. ISSN 19346069. doi: 10.1016/j.chom.2019.01.002. URL https://doi.org/10.1016/j.chom.2019.01.002.

111. Scott E. Hensley. Challenges of selecting seasonal influenza vaccine strains for humans with diverse pre-exposure histories. *Current Opinion in Virology*, 8:85–89, 2014. ISSN 18796265. doi: 10.1016/j.coviro.2014.07.007. URL http://dx.doi.org/10.1016/j.coviro.2014.07.007.

112. Herbert W. Hethcote. Qualitative analyses of communicable disease models. *Mathematical Biosciences*, 28(3-4):335–356, 1976. ISSN 00255564. doi: 10.1016/0025-5564(76)90132-2.

113. Herbert W. Hethcote. Mathematics of infectious diseases. *SIAM Review*, 42(4):599–653, 2000. ISSN 00361445. doi: 10.1137/S0036144500371907.

114. Siddhivinayak Hirve, Laura P Newman, John Paget, Eduardo Azziz-Baumgartner, Julia Fitzner, Niranjan Bhat, Katelijn Vandemaele, and Wenqing Zhang. Influenza Seasonality in the Tropics and Subtropics – When to Vaccinate? *PLOS ONE*, 11(4): 1–12, 2016. doi: 10.1371/journal.pone.0153003. URL https://doi.org/10.1371/journal.pone.0153003.

115. D Hobson, R L Curry, A S Beare, and A Ward-Gardner. The role of serum haemagglutination-inhibiting antibody in protection against challenge infection with influenza A2 and B viruses. *The Journal of hygiene*, 70(4):767–777, dec 1972. ISSN 0022-1724 (Print). doi: 10.1017/s0022172400022610.

116. Michael Höhle, Andrea Riebler, and Michaela Paul. Getting started with outbreak detection. 2007.

117. R. E. Hope-Simpson and D. B. Golubev. A new concept of the epidemic process of influenza A virus. *Epidemiology and Infection*, 99(1):5–54, 1987. ISSN 14694409. doi: 10.1017/S0950268800066851.

118. Katja Hoschler, Catherine Thompson, Nick Andrews, Monica Galiano, Richard Pebody, Joanna Ellis, Elaine Stanford, Marc Baguelin, Elizabeth Miller, Maria Zambon, England East, East Midlands, North East, South East, South West, West Midlands, North West, and South West. Seroprevalence of Influenza A (H1N1) pdm09 Virus. *Emerging Infectious Diseases*, 18(11):2010–2013, nov 2012. ISSN 10806040. doi: 10.3201/eid1811.120720.

119. Thomas House and Matt J. Keeling. Insights from unifying modern approximations to infections on networks. *Journal of the Royal Society Interface*, 8(54):67–73, 2011. ISSN 17425662. doi: 10.1098/rsif.2010.0179.

120. Su Jhen Hung, Yin Mei Hsu, Sheng Wen Huang, Huey Pin Tsai, Leo Yi Yang Lee, Aeron C. Hurt, Ian G. Barr, Shin Ru Shih, and Jen Ren Wang. Genetic variations on 31 and 450 residues of influenza A nucleoprotein affect viral replication and translation. *Journal of Biomedical Science*, 27(1):1–13, 2020. ISSN 14230127. doi: 10.1186/s12929-019-0612-z.

121. Lori Hutwagner, William Thompson, G. Matthew Seeman, and Tracee Treadwell. The Bioterrorism Preparedness and Response Early Aberration Reporting System (EARS). *Journal of Urban Health*, 80(1):89–96, 2003. ISSN 10993460. doi: 10.1007/pl00022319.

122. Lori Hutwagner, Timothy Browne, G. Matthew Seeman, and Aaron T. Fleischauer. Comparing aberration detection methods with stimulated data. *Emerging Infectious Diseases*, 11(2):314–316, 2005. ISSN 10806040. doi: 10.3201/eid1102.040587.

123. Charlotte Jackson, Emilia Vynnycky, Jeremy Hawker, Babatunde Olowokure, and Punam Mangtani. School closures and influenza: Systematic review of epidemiological studies. *BMJ Open*, 3(2):1–10, 2013. ISSN 20446055. doi: 10.1136/bmjopen-2012-002149.

124. J. M. Jaramillo, Junling Ma, P. van den Driessche, and Sanling Yuan. Host contact structure is important for the recurrence of Influenza A. *Journal of Mathematical Biology*, 77(5):1563–1588, 2018. ISSN 14321416. doi: 10.1007/s00285-018-1263-5. URL https://doi.org/10.1007/s00285-018-1263-5.

125. Brian Karrer and M. E.J. Newman. Competing epidemics on complex networks. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, 84(3):1–14, 2011. ISSN 15393755. doi: 10.1103/PhysRevE.84.036106.

126. Steven M Kay. *Fundamentals of statistical signal processing: Practical algorithm development*, volume 3. Pearson Education, 2013. ISBN 9780132808033.

127. Matt J. Keeling, Leon Danon, Matthew C. Vernon, and Thomas A. House. Individual identity and movement networks for disease metapopulations. *Proceedings of the National Academy of Sciences of the United States of America*, 107(19):8866–8870, 2010. ISSN 00278424. doi: 10.1073/pnas.1000416107.

128. William Ogilvy Kermack, A G McKendrick, and Gilbert Thomas Walker. A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, 115(772):700–721, 1927. ISSN 00029890. doi: 10.1098/rspa.1927.0118. URL https://royalsocietypublishing.org/doi/abs/10.1098/rspa.1927.0118.

129. Björn F Koel, David F Burke, Theo M Bestebroer, Stefan Van Der Vliet, Gerben C M Zondag, Gaby Vervaet, Eugene Skepner, Nicola S Lewis, Monique I J Spronken, Colin A Russell, Mikhail Y Eropkin, Aeron C Hurt, Ian G Barr, Jan C De Jong, Guus F Rimmelzwaan, Albert D M E Osterhaus, Ron A M Fouchier, and Derek J Smith. Substitutions Near the Receptor Binding. *Science*, 2006(November):976–980, 2013. ISSN 0036-8075. doi: 10.1126/science.1244730. URL http://www.ncbi.nlm.nih.gov/pubmed/24264991{%}5Cnhttp://www.sciencemag.org/cgi/doi/10.1126/science.1244730.

130. Katia Koelle, Sarah Cobey, Bryan Grenfell, and Mercedes Pascual. Epochal Evolution Shapes the Phylodynamics of Interpandemic Influenza A (H3N2) in Humans. *Science*, 6829(December):1898–1904, 2006.

131. Katia Koelle, Meredith Kamradt, and Mercedes Pascual. Understanding the dynamics of rapidly evolving pathogens through modeling the tempo of antigenic change: Influenza as a case study. *Epidemics*, 1(2):129–137, 2009. ISSN 17554365. doi: 10.1016/j.epidem.2009.05.003.

132. Katia Koelle, Priya Khatri, Meredith Kamradt, and Thomas B Kepler. A two-tiered model for simulating the ecological and evolutionary dynamics of rapidly evolving viruses, with an application to influenza. *Journal of the Royal Society, Interface / the Royal Society*, 7(50):1257–1274, 2010. ISSN 1742-5689. doi: 10.1098/rsif.2010.0007.

133. Karen A. Kormuth, Kaisen Lin, Zhihong Qian, Michael M. Myerburg, Linsey C. Marr, and Seema S. Lakdawala. Environmental Persistence of Influenza Viruses Is Dependent upon Virus Type and Host Origin. *mSphere*, 4(4):1–14, 2019. ISSN 2379-5042. doi: 10.1128/msphere.00552-19.

134. F. Krammer, R. Hai, M. Yondola, G. S. Tan, V. H. Leyva-Grado, A. B. Ryder, M. S. Miller, J. K. Rose, P. Palese, A. Garcia-Sastre, and R. A. Albrecht. Assessment of Influenza Virus Hemagglutinin Stalk-Based Immunity in Ferrets. *Journal of Virology*, 88(6):3432–3442, 2014. ISSN 0022-538X. doi: 10.1128/jvi.03004-13.

135. Florian Krammer and Peter Palese. Influenza virus hemagglutinin stalk-based antibodies and vaccines. *Current opinion in virology*, 3(5):521–530, oct 2013. ISSN 1879-6265 (Electronic). doi: 10.1016/j.coviro.2013.07.007.Influenza.

136. Florian Krammer and Peter Palese. Universal Influenza Virus Vaccines That Target the Conserved Hemagglutinin Stalk and Conserved Sites in the Head Domain. *Journal of Infectious Diseases*, 219(Suppl 1):S62–S67, 2019. ISSN 15376613. doi: 10.1093/infdis/jiy711.

137. Sergey Kryazhimskiy, Ulf Dieckmann, Simon A. Levin, and Jonathan Dushoff. On state-space reduction in multi-strain pathogen models, with an application to antigenic drift in influenza A. *PLoS Computational Biology*, 3(8):1513–1525, 2007. ISSN 1553734X. doi: 10.1371/journal.pcbi.0030159.

138. Adam J. Kucharski and Julia R. Gog. The Role of Social Contacts and Original Antigenic Sin in Shaping the Age Pattern of Immunity to Seasonal Influenza. *PLoS Computational Biology*, 8(10), 2012. ISSN 1553734X. doi: 10.1371/journal.pcbi.1002741.

139. Adam J. Kucharski, Kin O. Kwok, Vivian W I Wei, Benjamin J. Cowling, Jonathan M. Read, Justin Lessler, Derek A. Cummings, and Steven Riley. The Contribution of Social Behaviour to the Transmission of Influenza A in a Human Population. *PLoS Pathogens*, 10(6), 2014. ISSN 15537374. doi: 10.1371/journal.ppat.1004206.

140. Adam J. Kucharski, Justin Lessler, Jonathan M. Read, Huachen Zhu, Chao Qiang Jiang, Yi Guan, Derek A.T. Cummings, and Steven Riley. Estimating the Life Course of Influenza A(H3N2) Antibody Responses from Cross-Sectional Data. *PLoS Biology*, 13 (3):1–16, 2015. ISSN 15457885. doi: 10.1371/journal.pbio.1002082.

141. Adam J. Kucharski, Viggo Andreasen, and Julia R. Gog. Capturing the dynamics of pathogens with many strains. *Journal of Mathematical Biology*, 72(1-2):1–24, 2016. ISSN 14321416. doi: 10.1007/s00285-015-0873-4. URL http://dx.doi.org/10.1007/s00285-015-0873-4.

142. Seema S. Lakdawala and Kanta Subbarao. The challenge of flu transmission. *Nature Medicine*, 18(10):1468–1471, 2012. ISSN 10788956. doi: 10.1038/nm.2953. URL http://dx.doi.org/10.1038/nm.2953.

143. Edward K.S. Lam, Dylan H Morris, Aeron C Hurt, Ian G Barr, and Colin A Russell. The impact of climate and antigenic evolution on seasonal influenza virus epidemics in Australia. *Nature Communications*, 11(1), 2020. ISSN 20411723. doi: 10.1038/s41467-020-16545-6. URL http://dx.doi.org/10.1038/s41467-020-16545-6.

144. Ha Minh Lam, Amy Wesolowski, Nguyen Thanh Hung, Tran Dang Nguyen, Nguyen Thi Duy Nhat, Stacy Todd, Dao Nguyen Vinh, Nguyen Ha Thao Vy, Tran Thi Nhu Thao, Nguyen Thi Le Thanh, Phan Tri Tin, Ngo Ngoc Quang Minh, Juliet E. Bryant,

Caroline O. Buckee, Tran Van Ngoc, Nguyen Van Vinh Chau, Guy E. Thwaites, Jeremy Farrar, Dong Thi Hoai Tam, Ha Vinh, and Maciej F. Boni. Nonannual seasonality of influenza-like illness in a tropical urban setting. *Influenza and other Respiratory Viruses*, 12(6):742–754, 2018. ISSN 17502659. doi: 10.1111/irv.12595.

145. Jonathan H. Lam and Nicole Baumgarth. The Multifaceted B Cell Response to Influenza Virus. *The Journal of Immunology*, 202(2):351–359, 2019. ISSN 0022-1767. doi: 10.4049/jimmunol.1801208.

146. Pinky Langat, Jayna Raghwani, Gytis Dudas, Thomas A. Bowden, Stephanie Edwards, Astrid Gall, Trevor Bedford, Andrew Rambaut, Rodney S. Daniels, Colin A. Russell, Oliver G. Pybus, John McCauley, Paul Kellam, and Simon J. Watson. Genome-wide evolutionary dynamics of influenza B viruses on a global scale. *PLoS pathogens*, 13 (12):e1006749, dec 2017. ISSN 15537374. doi: 10.1371/journal.ppat.1006749.

147. Lincoln L. H. Lau, Benjamin J. Cowling, Vicky J. Fang, Kwok-Hung Chan, Eric H. Y. Lau, Marc Lipsitch, Calvin K. Y. Cheng, Peter M. Houck, Timothy M. Uyeki, J. S. Malik Peiris, and Gabriel M. Leung. Viral Shedding and Clinical Illness in Naturally Acquired Influenza Virus Infections. *The Journal of Infectious Diseases*, 201(10):1509–1516, 2010. ISSN 0022-1899. doi: 10.1086/652241.

148. Juhye M Lee, Rachel Eguia, Seth J Zost, Saket Choudhary, Patrick C Wilson, Trevor Bedford, Terry Stevens-Ayers, Michael Boeckh, Aeron C Hurt, Seema S Lakdawala, Scott E Hensley, and Jesse D Bloom. Mapping person-to-person variation in viral mutations that escape polyclonal serum targeting influenza hemagglutinin. *eLife*, 8, aug 2019. ISSN 2050-084X. doi: 10.7554/elife.49324. URL https://elifesciences.org/articles/49324.

149. Juhye M Lee, Rachel Eguia, Seth J Zost, Saket Choudhary, Patrick C Wilson, Trevor Bedford, Terry Stevens-Ayers, Michael Boeckh, Aeron C Hurt, Seema S Lakdawala, Scott E Hensley, and Jesse D Bloom. Mapping person-to-person variation in viral mutations that escape polyclonal serum targeting influenza hemagglutinin. *eLife*, 8, aug 2019. ISSN 2050-084X. doi: 10.7554/elife.49324.

150. Magali Lemaitre and Fabrice Carrat. Comparative age distribution of influenza morbidity and mortality during seasonal influenza epidemics and the 2009 H1N1 pandemic. *BMC infectious diseases*, 10(April 2009):162, 2010. ISSN 1471-2334. doi: 10.1186/1471-2334-10-162.

151. Justin Lessler, Steven Riley, Jonathan M. Read, Shuying Wang, Huachen Zhu, Gavin J D Smith, Yi Guan, Chao Qiang Jiang, and Derek A T Cummings. Evidence for antigenic seniority in influenza A (H3N2) antibody responses in southern China. *PLoS Pathogens*, 8(7):26, 2012. ISSN 15537366. doi: 10.1371/journal.ppat.1002802.

152. Kathy Leung, Mark Jit, Eric H.Y. Lau, and Joseph T. Wu. Social contact patterns relevant to the spread of respiratory infectious diseases in Hong Kong. *Scientific Reports*, 7(July):1–12, 2017. ISSN 20452322. doi: 10.1038/s41598-017-08241-1. URL http://dx.doi.org/10.1038/s41598-017-08241-1.

153. Gabriel E. Leventhal, Alison L. Hill, Martin A. Nowak, and Sebastian Bonhoeffer. Evolution and emergence of infectious diseases in theoretical and real-world networks.

Nature Communications, 6:1–11, 2015. ISSN 20411723. doi: 10.1038/ncomms7101. URL http://dx.doi.org/10.1038/ncomms7101.

154. Richard Levins. Some demographic and genetic consequences of environmental heterogeneity for biological control. *American Entomologist*, 15(3):237–240, 1969.

155. Fredrik Liljeros, Christofer R Edling, Luís A Nunes Amaral, H Eugene Stanley, and Yvonne Åberg. The web of human sexual contacts. *Nature*, 411(6840):907–908, 2001. ISSN 1476-4687. doi: 10.1038/35082140. URL https://doi.org/10.1038/35082140.

156. Susanne L. Linderman, Benjamin S. Chambers, Seth J. Zost, Kaela Parkhousea, Yang Li, Christin Herrmann, Ali H. Ellebedy, Donald M. Carter, Sarah F. Andrews, Nai Ying Zheng, Min Huang, Yunping Huang, Donna Strauss, Beth H. Shaz, Richard L. Hodinka, Gustavo Reyes-Terán, Ted M. Ross, Patrick C. Wilson, Rafi Ahmed, Jesse D. Bloom, and Scott E. Hensley. Potential antigenic explanation for atypical h1n1 infections among middle-aged adults during the 2013-2014 influenza season. *Proceedings of the National Academy of Sciences of the United States of America*, 111(44):15798–15803, 2014. ISSN 10916490. doi: 10.1073/pnas.1409171111.

157. Jennifer Lindquist, Junling Ma, P. van den Driessche, and Frederick H. Willeboordse. Effective degree network disease models. *Journal of Mathematical Biology*, 62(2): 143–164, feb 2011. ISSN 0303-6812. doi: 10.1007/s00285-010-0331-2. URL http://link.springer.com/10.1007/s00285-010-0331-2.

158. Marc Lipsitch and Cécile Viboud. Influenza seasonality: Lifting the fog. *Proceedings of the National Academy of Sciences of the United States of America*, 106(10):3645–3646, 2009. ISSN 00278424. doi: 10.1073/pnas.0900933106.

159. J. O. Lloyd-Smith, S. J. Schreiber, P. E. Kopp, and W. M. Getz. Superspreading and the effect of individual variation on disease emergence. *Nature*, 438(7066):355–359, 2005. ISSN 14764687. doi: 10.1038/nature04153.

160. Eric Lofgren, N. H. Fefferman, Y. N. Naumov, J. Gorski, and E. N. Naumova. Influenza Seasonality: Underlying Causes and Modeling Theories. *Journal of Virology*, 81(11): 5429–5436, 2007. ISSN 0022-538X. doi: 10.1128/jvi.01680-06.

161. Anice C. Lowen and John Steel. Roles of Humidity and Temperature in Shaping Influenza Seasonality. *Journal of Virology*, 88(14):7692–7695, 2014. ISSN 0022-538X. doi: 10.1128/JVI.03544-13. URL http://jvi.asm.org/lookup/doi/10.1128/JVI.03544-13.

162. Anice C. Lowen, Samira Mubareka, John Steel, and Peter Palese. Influenza virus transmission is dependent on relative humidity and temperature. *PLoS Pathogens*, 3 (10):1470–1476, 2007. ISSN 15537366. doi: 10.1371/journal.ppat.0030151.

163. Anice C. Lowen, John Steel, Samira Mubareka, and Peter Palese. High Temperature (30°C) Blocks Aerosol but Not Contact Transmission of Influenza Virus. *Journal of Virology*, 82(11):5650–5652, 2008. ISSN 0022-538X. doi: 10.1128/jvi.00325-08.

164. Giancarlo De Luca, Kim Van Kerckhove, Pietro Coletti, Chiara Poletto, Nathalie Bossuyt, Niel Hens, and Vittoria Colizza. The impact of regular school closure on seasonal influenza epidemics: A data-driven spatial transmission model for

Belgium. *BMC Infectious Diseases*, 18(1):1–16, 2018. ISSN 14712334. doi: 10.1186/s12879-017-2934-3.

165. Junling Ma and David J.D. Earn. Generality of the final size formula for an epidemic of a newly invading infectious disease. *Bulletin of Mathematical Biology*, 68(3):679–702, 2006. ISSN 00928240. doi: 10.1007/s11538-005-9047-7.

166. Pierre Magal, Ousmane Seydi, and Glenn Webb. Final Size of an Epidemic for a Two-Group SIR Model. *SIAM Journal on Applied Mathematics*, 76(5):2042–2059, 2016. doi: 10.1137/16M1065392. URL https://doi.org/10.1137/16M1065392.

167. Pierre Magal, Ousmane Seydi, and Glenn Webb. Final size of a multi-group SIR epidemic model: Irreducible and non-irreducible modes of transmission. *Mathematical Biosciences*, 301(June 2017):59–67, jul 2018. ISSN 18793134. doi: 10.1016/j.mbs.2018.03.020. URL https://linkinghub.elsevier.com/retrieve/pii/S0025556417303474https://doi.org/10.1016/j.mbs.2018.03.020.

168. Robert M. May and Alun L. Lloyd. Infection dynamics on scale-free networks. *Physical Review E - Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics*, 64(6):4, 2001. ISSN 1063651X. doi: 10.1103/PhysRevE.64.066112.

169. Jonathan A. McCullers and Victor C. Huber. Correlates of vaccine protection from influenza and its complications. *Human Vaccines and Immunotherapeutics*, 8(1):34–44, 2012. ISSN 2164554X. doi: 10.4161/hv.8.1.18214.

170. James McDevitt, Stephen Rudnick, Melvin First, and John Spengler. Role of absolute humidity in the inactivation of influenza viruses on stainless steel surfaces at elevated temperatures. *Applied and Environmental Microbiology*, 76(12):3943–3947, 2010. ISSN 00992240. doi: 10.1128/AEM.02674-09.

171. Andrew J McMichael, Frances M Gotch, Gary R Noble, and Paul A S Beare. Cytotoxic T-Cell Immunity to Influenza. *New England Journal of Medicine*, 309(1): 13–17, 1983. doi: 10.1056/NEJM198307073090103. URL https://doi.org/10.1056/NEJM198307073090103.

172. Philip Meade, Guillermina Kuan, Shirin Strohmeier, Hannah E. Maier, Fatima Amanat, Angel Balmaseda, Kimihito Ito, Ericka Kirkpatrick, Andres Javier, Lionel Gresh, Raffael Nachbagauer, Aubree Gordon, and Florian Krammer. Influenza Virus Infection Induces a Narrow Antibody Response in Children but a Broad Recall Response in Adults. *mBio*, 11(1):1–15, 2020. ISSN 21507511. doi: 10.1128/mBio.03243-19.

173. Alessia Melegaro, Mark Jit, Nigel Gay, Emilio Zagheni, and W. John Edmunds. What types of contacts are important for the spread of infections? Using contact survey data to explore European mixing patterns. *Epidemics*, 3(3-4):143–151, 2011. ISSN 17554365. doi: 10.1016/j.epidem.2011.04.001. URL http://dx.doi.org/10.1016/j.epidem.2011.04.001.

174. Alessia Melegaro, Mark Jit, Nigel Gay, Emilio Zagheni, and W. John Edmunds. What types of contacts are important for the spread of infections? Using contact survey data to explore European mixing patterns. *Epidemics*, 3(3-4):143–151, 2011. ISSN 17554365. doi: 10.1016/j.epidem.2011.04.001. URL http://dx.doi.org/10.1016/j.epidem.2011.04.001.

175. Matthew J Memoli, Brett W Jagger, Vivien G Dugan, Li Qi, Jadon P Jackson, and Jeffery K Taubenberger. Recent human influenza A/H3N2 virus evolution driven by novel selection factors in addition to antigenic drift. *The Journal of infectious diseases*, 200(8):1232–1241, oct 2009. ISSN 1537-6613 (Electronic). doi: 10.1086/605893.

176. Matthew J Memoli, Alison Han, Kathie-Anne Walters, Lindsay Czajkowski, Susan Reed, Rani Athota, Luz Angela Rosas, Adriana Cervantes-Medina, Jae-Keun Park, David M Morens, John C Kash, and Jeffery K Taubenberger. Influenza A Reinfection in Sequential Human Challenge: Implications for Protective Immunity and "Universal" Vaccine Development. *Clinical Infectious Diseases*, 3203(Xx):1–6, 2019. ISSN 1058-4838. doi: 10.1093/cid/ciz281.

177. Lauren Ancel Meyers, Babak Pourbohloul, M.E.J. E.J. Newman, Danuta M. Skowronski, and Robert C. Brunham. Network theory and SARS: predicting outbreak diversity. *Journal of Theoretical Biology*, 232(1):71–81, jan 2005. ISSN 00225193. doi: 10.1016/j.jtbi.2004.07.026. URL https://linkinghub.elsevier.com/retrieve/pii/S0022519304003510.

178. Matthew S Miller, Thomas J Gardner, Florian Krammer, Lauren C Aguado, Christopher F Basler, and Peter Palese. Neutralizing Antibodies Against Previously Encountered Influenza Virus Strains Increase over Time: A Longitudinal Analysis. *Science Translational Medicine*, 5(198), 2013. doi: 10.1126/scitranslmed.3006637. Neutralizing.

179. Pavlo Minayev and Neil Ferguson. Improving the realism of deterministic multi-strain models: Implications for modelling influenza A. *Journal of the Royal Society Interface*, 6(35):509–518, 2009. ISSN 17425662. doi: 10.1098/rsif.2008.0333.

180. S. M. Minhaz Ud-Dean. Structural explanation for the effect of humidity on persistence of airborne virus: Seasonality of influenza. *Journal of Theoretical Biology*, 264 (3):822–829, 2010. ISSN 00225193. doi: 10.1016/j.jtbi.2010.03.013. URL http://dx.doi.org/10.1016/j.jtbi.2010.03.013.

181. Noelle Angelique M. Molinari, Ismael R. Ortega-Sanchez, Mark L. Messonnier, William W. Thompson, Pascale M. Wortley, Eric Weintraub, and Carolyn B. Bridges. The annual impact of seasonal influenza in the US: Measuring disease burden and costs. *Vaccine*, 25(27):5086–5096, 2007. ISSN 0264410X. doi: 10.1016/j.vaccine.2007.03.046.

182. Arnold S. Monto. Interrupting the transmission of respiratory tract infections: Theory and practice. *Clinical Infectious Diseases*, 28(2):200–204, 1999. ISSN 10584838. doi: 10.1086/515113.

183. ARNOLD S MONTO and FARIDEH KIOUMEHR. THE TECUMSEH STUDY OF RESPIRATORY ILLNESS: IX. OCCURRENCE OF INFLUENZA IN THE COMMUNITY, 1966–19711. *American Journal of Epidemiology*, 102(6):553–563, 1975. ISSN 0002-9262. doi: 10.1093/oxfordjournals.aje.a112193. URL https://doi.org/10.1093/oxfordjournals.aje.a112193.

184. E. Q. Mooring and S. Bansal. Increasing herd immunity with influenza revaccination. *Epidemiology and Infection*, 144(6):1267–1277, apr 2016. ISSN 14694409. doi: 10.1017/S0950268815002253.

185. Dylan H. Morris, Velislava N. Petrova, Fernando W. Rossine, Edyth Parker, Bryan T. Grenfell, Richard A. Neher, Simon A. Levin, and Colin A. Russell. Asynchrony between virus diversity and antibody selection limits influenza virus evolution. *bioRxiv*, page 2020.04.27.064915, 2020. doi: 10.1101/2020.04.27.064915.

186. Joël Mossong, Niel Hens, Mark Jit, Philippe Beutels, Kari Auranen, Rafael Mikolajczyk, Marco Massari, Stefania Salmaso, Gianpaolo Scalia Tomba, Jacco Wallinga, Janneke Heijne, Malgorzata Sadkowska-Todys, Magdalena Rosinska, and W. John Edmunds. Social contacts and mixing patterns relevant to the spread of infectious diseases. *PLoS Medicine*, 5(3):0381–0391, 2008. ISSN 15491277. doi: 10.1371/journal.pmed.0050074.

187. Fernanda E.A. Moura. Influenza in the tropics. *Current Opinion in Infectious Diseases*, 23(5):415–420, 2010. ISSN 09517375. doi: 10.1097/QCO.0b013e32833cc955.

188. Vito M.R. Muggeo. Estimating regression models with unknown break-points. *Statistics in Medicine*, 22(19):3055–3071, 2003. ISSN 02776715. doi: 10.1002/sim.1545.

189. Raffael Nachbagauer, Angela Choi, Ruvim Izikson, Manon M. Cox, Peter Palese, and Florian Krammer. Age dependence and isotype specificity of influenza virus hemagglutinin stalk-reactive antibodies in humans. *mBio*, 7(1):1–10, 2016. ISSN 21507511. doi: 10.1128/mBio.01996-15.

190. Martha I. Nelson, Lone Simonsen, Cecile Viboud, Mark A. Miller, and Edward C. Holmes. Phylogenetic analysis reveals the global migration of seasonal influenza A viruses. *PLoS Pathogens*, 3(9):1220–1228, 2007. ISSN 15537366. doi: 10.1371/journal.ppat.0030131.

191. Kathleen M. M. M. Neuzil, Lisa A. A. A. Jackson, Jennifer Nelson, Alexander Klimov, Nancy Cox, Carolyn B. B. B. Bridges, John Dunn, Frank DeStefano, and David Shay. Immunogenicity and Reactogenicity of 1 versus 2 Doses of Trivalent Inactivated Influenza Vaccine in Vaccine-Naive 5–8-Year-Old Children. *The Journal of Infectious Diseases*, 194(8):1032–1039, oct 2006. ISSN 0022-1899. doi: 10.1086/507309. URL https://academic.oup.com/jid/article-lookup/doi/10.1086/507309.

192. M E J Newman. Spread of epidemic disease on networks. *Physical Review E*, 66 (1):016128, jul 2002. ISSN 1063-651X. doi: 10.1103/PhysRevE.66.016128. URL https://link.aps.org/doi/10.1103/PhysRevE.66.016128.

193. M E J Newman, S H Strogatz, and D J Watts. Random graphs with arbitrary degree distributions and their applications. 64:1–17, 2001. doi: 10.1103/PhysRevE.64.026118.

194. M. E.J. Newman. Threshold effects for two pathogens spreading on a network. *Physical Review Letters*, 95(10), 2005. ISSN 00319007. doi: 10.1103/PhysRevLett.95.108701.

195. Eri Nobusawa and Katsuhiko Sato. Comparison of the Mutation Rates of Human Influenza A and B Viruses. *J. Virol.*, 80(7):3675–3678, 2006. ISSN 0022-538X. doi: 10.1128/JVI.80.7.3675. URL http://jvi.asm.org/cgi/content/abstract/80/7/3675.

196. Angela Noufaily, Roger A. Morbey, Felipe J. Colón-González, Alex J. Elliot, Gillian E. Smith, Iain R. Lake, and Noel McCarthy. Comparison of statistical algorithms for daily syndromic surveillance aberration detection. *Bioinformatics*, 35(17):3110–3118, 2019. ISSN 14602059. doi: 10.1093/bioinformatics/bty997.

197. Romualdo Pastor-Satorras and Alessandro Vespignani. Epidemic spreading in scale-free networks. *The Structure and Dynamics of Networks*, 9781400841:493–496, 2001. ISSN 00319007. doi: 10.1103/PhysRevLett.86.3200.

198. C M Pease. An evolutionary epidemiological mechanism, with applications to type A influenza. *Theoretical population biology*, 31(3):422–452, jun 1987. ISSN 0040-5809 (Print). doi: 10.1016/0040-5809(87)90014-1.

199. Lorenzo Pellis, Frank Ball, Shweta Bansal, Ken Eames, Thomas House, Valerie Isham, and Pieter Trapman. Eight challenges for network epidemic models. *Epidemics*, 10:58–62, 2015. ISSN 18780067. doi: 10.1016/j.epidem.2014.07.003. URL http://dx.doi.org/10.1016/j.epidem.2014.07.003.

200. Velislava N. Petrova and Colin A. Russell. The evolution of seasonal influenza viruses. *Nature Reviews Microbiology*, 16(1):47–60, 2018. ISSN 17401534. doi: 10.1038/nrmicro.2017.118. URL http://dx.doi.org/10.1038/nrmicro.2017.118.

201. C W POTTER and J S OXFORD. DETERMINANTS OF IMMUNITY TO INFLUENZA INFECTION IN MAN. *British Medical Bulletin*, 35(1):69–75, jan 1979. ISSN 0007-1420. doi: 10.1093/oxfordjournals.bmb.a071545. URL https://doi.org/10.1093/oxfordjournals.bmb.a071545.

202. Gail E. Potter, Mark S. Handcock, Ira M. Longini, and M. Elizabeth Halloran. Estimating within-household contact networks from egocentric data. *Annals of Applied Statistics*, 5(3):1816–1838, 2011. ISSN 19326157. doi: 10.1214/11-AOAS474.

203. Public Health England. Flu Plan Winter 2016/17. *Public Health England*, 2016. doi: 10.1037/e500942012-001. URL https://www.gov.uk/government/organisations/public-health-england/about.

204. Jayna Raghwani, Robin N Thompson, and Katia Koelle. Selection on non-antigenic gene segments of seasonal influenza A virus and its impact on adaptive evolution. *Virus Evolution*, 3(2):1–12, 2017. ISSN 2057-1577. doi: 10.1093/ve/vex034.

205. Andrew Rambaut, Oliver G. Pybus, Martha I. Nelson, Cecile Viboud, Jeffery K. Taubenberger, and Edward C. Holmes. The genomic and epidemiological dynamics of human influenza A virus. *Nature*, 453(7195):615–619, 2008. ISSN 14764687. doi: 10.1038/nature06945.

206. Sylvia Ranjeva, Rahul Subramanian, Vicky J Fang, Gabriel M Leung, Dennis K M Ip, Ranawaka A P M Perera, J S Malik Peiris, Benjamin J Cowling, and Sarah Cobey. Age-specific differences in the dynamics of protective immunity to influenza. *Nature Communications*, 10(1):1660, 2019. ISSN 2041-1723. doi: 10.1038/s41467-019-09652-6. URL https://doi.org/10.1038/s41467-019-09652-6.

207. O Ratmann, G Donker, A Meijer, C Fraser, and K Koelle. Phylodynamic inference and model assessment with approximate bayesian computation: influenza as a case study. *PLoS Comput Biol*, 8(12):e1002835, 2012. ISSN 1553-734x. doi: 10.1371/journal.pcbi.1002835PCOMPBIOL-D-12-00439[pii]. URL http://www.ncbi.nlm.nih.gov/pubmed/23300420.

208. Ahmed Rguig, Imad Cherkaoui, Margaret McCarron, Hicham Oumzil, Soumia Triki, Houria Elmbarki, Abderrahman Bimouhen, Fatima El Falaki, Zakia Regragui, Hassan Ihazmad, Chakib Nejjari, and Mohammed Youbi. Establishing seasonal and alert influenza thresholds in Morocco. *BMC Public Health*, 20(1):1029, 2020. ISSN 1471-2458. doi: 10.1186/s12889-020-09145-y. URL https://bmcpublichealth.biomedcentral.com/articles/10.1186/s12889-020-09145-y.

209. Pejman Rohani, Matthew J. Keeling, and Bryan T. Grenfell. The interplay between determinism and stochasticity in childhood diseases. *American Naturalist*, 159(5): 469–481, 2002. ISSN 00030147. doi: 10.1086/339467.

210. Pejman Rohani, Xue Zhong, and Aaron A. King. Contact network structure explains the changing epidemiology of pertussis. *Science*, 330(6006):982–985, 2010. ISSN 00368075. doi: 10.1126/science.1194134.

211. Ronald Ross. *The Prevention of malaria*. J. Murray, 1910.

212. C A Russell, T C Jones, I G Barr, N J Cox, R J Garten, V Gregory, I D Gust, A W Hampson, A J Hay, A C Hurt, J C de Jong, A Kelso, A I Klimov, T Kageyama, N Komadina, A S Lapedes, Y P Lin, A Mosterin, M Obuchi, T Odagiri, A D Osterhaus, G F Rimmelzwaan, M W Shaw, E Skepner, K Stohr, M Tashiro, R A Fouchier, and D J Smith. The global circulation of seasonal influenza A (H3N2) viruses. *Science*, 320(5874):340–346, 2008. ISSN 0036-8075. doi: 10.1126/science.1154137. URL http://www.ncbi.nlm.nih.gov/pubmed/18420927.

213. Takatsugu Sakai, Hiroshi Suzuki, Asami Sasaki, Reiko Saito, Naohito Tanabe, and Kiyosu Taniguchi. Geographic and temporal trends in influenzalike illness, Japan, 1992-1999. *Emerging Infectious Diseases*, 10(10):1822–1826, 2004. ISSN 10806040. doi: 10.3201/eid1010.040147.

214. B. Salah, A. T. Dinh Xuan, J. L. Fouilladieu, A. Lockhart, and J. Regnard. Nasal mucociliary transport in healthy subjects is slower when breathing dry air. *European Respiratory Journal*, 1(9):852–855, 1988. ISSN 09031936.

215. Marcel Salathé, Maria Kazandjieva, Jung Woo Lee, Philip Levis, Marcus W. Feldman, and James H. Jones. A high-resolution human contact network for infectious disease transmission. *Proceedings of the National Academy of Sciences of the United States of America*, 107(51):22020–22025, 2010. ISSN 00278424. doi: 10.1073/pnas.1009094108.

216. F. L. Schaffer, M. E. Soergel, and D. C. Straube. Survival of airborne influenza virus: Effects of propagating host, relative humidity, and composition of spray fluids. *Archives of Virology*, 51(4):263–273, 1976. ISSN 03048608. doi: 10.1007/BF01317930.

217. Nathalie Schnepf, Matthieu Resche-Rigon, Antoine Chaillon, Anne Scemla, Guillaume Gras, Oren Semoun, Pierre Taboulet, Jean Michel Molina, François Simon, Alain Goudeau, and Jérôme LeGoff. High burden of non-influenza viruses in influenza-like illness in the early weeks of H1N1v epidemic in France. *PLoS ONE*, 6(8), 2011. ISSN 19326203. doi: 10.1371/journal.pone.0023514.

218. Bethany K Sederdahl and John V Williams. Epidemiology and Clinical Characteristics of Influenza C Virus. *Viruses*, 12(1):1–11, jan 2020. ISSN 1999-4915 (Electronic). doi: 10.3390/v12010089. URL https://viralzone.expasy.org/81.

219. Nikita E. Seleznev and Vasiliy N. Leonenko. Absolute humidity anomalies and the influenza onsets in Russia: A computational study. *Procedia Computer Science*, 119: 224–233, 2017. ISSN 18770509. doi: 10.1016/j.procs.2017.11.180. URL https://doi.org/10.1016/j.procs.2017.11.180.

220. Robert E. Serfling. Methods for Current Statistical Analysis of Excess Pneumonia-Influenza Deaths. *Public Health Reports (1896-1970)*, 78(6):494, 1963. ISSN 00946214. doi: 10.2307/4591848.

221. J. Shaman and M. Kohn. Absolute humidity modulates influenza survival, transmission, and seasonality. *Proceedings of the National Academy of Sciences*, 106(9):3243–3248, 2009. ISSN 0027-8424. doi: 10.1073/pnas.0806852106. URL http://www.pnas.org/cgi/doi/10.1073/pnas.0806852106.

222. Jeffrey Shaman, Virginia E. Pitzer, Cécile Viboud, Bryan T. Grenfell, and Marc Lipsitch. Absolute humidity and the seasonal onset of influenza in the continental United States. *PLoS Biology*, 8(2), 2010. ISSN 15449173. doi: 10.1371/journal.pbio.1000316.

223. I. L. Shechmeister. Studies on the Experimental Epidemiology of Respiratory Infections: III. Certain Aspects of the Behavior of Type A Influenza Virus as an Air-Borne Cloud. *Journal of Infectious Diseases*, 87(2):128–132, 1950. ISSN 0022-1899. doi: 10.1093/infdis/87.2.128.

224. Lynette Pei-Chi Shek and Bee-Wah Lee. Epidemiology and seasonality of respiratory tract virus infections in the tropics. *Paediatric respiratory reviews*, 4(2):105–111, jun 2003. ISSN 1526-0542 (Print). doi: 10.1016/s1526-0542(03)00024-1.

225. Makoto Shoji, Kouki Katayama, and Kunio Sano. Absolute Humidity as a Deterministic Factor Affecting Seasonal Influenza Epidemics in Japan. *The Tohoku Journal of Experimental Medicine*, 224(4):251–256, 2011. ISSN 0040-8727. doi: 10.1620/tjem.224.251.

226. James Mark Simmerman, Pranee Thawatsupha, Darika Kingnate, Keiji Fukuda, Arunee Chaising, and Scott F. Dowell. Influenza in Thailand: A case study for middle income countries. *Vaccine*, 23(2):182–187, 2004. ISSN 0264410X. doi: 10.1016/j.vaccine.2004.05.025.

227. Lone Simonsen, Thomas A. Reichert, Cecile Viboud, William C. Blackwelder, Robert J. Taylor, and Mark A. Miller. Impact of Influenza Vaccination on Seasonal Mortality in the US Elderly Population. *Arch Intern Med.*, 165(3):265–272, 2005. doi: 10.1001/archinte.165.3.265.

228. J J Skehel, D J Stevens, R S Daniels, A R Douglas, M Knossow, I A Wilson, and D C Wiley. A carbohydrate side chain on hemagglutinins of Hong Kong influenza viruses inhibits recognition by a monoclonal antibody. *Cell*, 81(March):1779–1783, 1984. ISSN 0027-8424. doi: 10.1073/pnas.81.6.1779. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=345004{&}tool=pmcentrez{&}rendertype=abstract.

229. Daniel Smilkov, Cesar A. Hidalgo, and Ljupco Kocarev. Beyond network structure: How heterogeneous susceptibility modulates the spread of epidemics. *Scientific Reports*, 4:1–13, 2014. ISSN 20452322. doi: 10.1038/srep04795.

230. D J Smith, S Forrest, D H Ackley, and A S Perelson. Variable efficacy of repeated annual influenza vaccination. *Proceedings of the National Academy of Sciences of the United States of America*, 96(24):14001–6, 1999. ISSN 0027-8424. doi: 10.1073/pnas.96.24.14001. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=24180{&}tool=pmcentrez{&}rendertype=abstract.

231. D. J. Smith, A. S. Lapedes, J. C. de Jong, T. M. Bestebroer, G. F. Rimmelzwaan, A. D. Osterhaus, and R. A. Fouchier. Mapping the Antigenic and Genetic Evolution of Influenza Virus. *Science*, 305(5682):371–376, 2004. ISSN 0036-8075. doi: 10.1126/science.1097211. URL http://www.sciencemag.org/content/305/5682/371{%}5Cnhttp://www.sciencemag.org/cgi/doi/10.1126/science.1097211.

232. Tadao Sonoguchi, H. Naito, Masaru Hara, Yasue Takeuchi, and Hideo Fukumi. Cross-subtype protection in humans during sequential, overlapping, and/or concurrent epidemics caused by h3n2 and h1n1 influenza viruses. *Journal of Infectious Diseases*, 151 (1):81–88, 1985. ISSN 15376613. doi: 10.1093/infdis/151.1.81.

233. Stan Development Team. Stan Modeling Language Users Guide and Reference Manual, 2018. URL http://mc-stan.org.

234. Stan Development Team. {RStan}: the {R} interface to {Stan}, 2020. URL http://mc-stan.org/.

235. John Steel, Anice C. Lowen, Taia T. Wang, Mark Yondola, Qinshan Gao, Kester Haye, Adolfo García-Sastre, and Peter Palesea. Influenza virus vaccine based on the conserved hemagglutinin stalk domain. *mBio*, 1(1):1–9, 2010. ISSN 21507511. doi: 10.1128/mBio.00018-10.

236. Yvonne C.F. Su, Justin Bahl, Udayan Joseph, Ka Man Butt, Heidi A. Peck, Evelyn S.C. Koay, Lynette L.E. Oon, Ian G. Barr, Dhanasekaran Vijaykrishna, and Gavin J.D. Smith. Phylodynamics of H1N1/2009 influenza reveals the transition from host adaptation to immune-driven selection. *Nature Communications*, 6:1–13, 2015. ISSN 20411723. doi: 10.1038/ncomms8952. URL http://dx.doi.org/10.1038/ncomms8952.

237. Jianhua Sui, William C Hwang, Sandra Perez, Ge Wei, Daniel Aird, Li-mei Chen, Eugenio Santelli, Boguslaw Stec, Greg Cadwell, Maryam Ali, Hongquan Wan, Akikazu Murakami, Anuradha Yammanuru, Thomas Han, Nancy J Cox, Laurie a Bankston, Ruben O Donis, Robert C Liddington, and Wayne a Marasco. Structural and functional bases for broad-spectrum neutralization of avian and human influenza A viruses. *Nature structural & molecular biology*, 16(3):265–273, 2009. ISSN 1545-9993. doi: 10.1038/nsmb.1566.

238. Sheena G. Sullivan, Monique B.N. Chilver, Geoff Higgins, Allen C. Cheng, and Nigel P. Stocks. Influenza vaccine effectiveness in Australia: Results from the Australian Sentinel Practices Research Network. *Medical Journal of Australia*, 201(2):109–111, 2014. ISSN 13265377. doi: 10.5694/mja14.00106.

239. James Tamerius, Martha I. Nelson, Steven Z. Zhou, Cécile Viboud, Mark A. Miller, and Wladimir J. Alonso. Global influenza seasonality: Reconciling patterns across temperate and tropical regions. *Environmental Health Perspectives*, 119(4):439–445, 2011. ISSN 00916765. doi: 10.1289/ehp.1002383.

240. James D. Tamerius, Jeffrey Shaman, Wladmir J. Alonso, Kimberly Bloom-Feshbach, Christopher K. Uejio, Andrew Comrie, and Cécile Viboud. Environmental Predictors of Seasonal Influenza Epidemics across Temperate and Tropical Climates. *PLoS Pathogens*, 9(3), 2013. ISSN 15537366. doi: 10.1371/journal.ppat.1003194.

241. Ee Laine Tay, Kristina Grant, Martyn Kirk, Anthony Mounts, and Heath Kelly. Exploring a Proposed WHO Method to Determine Thresholds for Seasonal Influenza Surveillance. *PLoS ONE*, 8(10):1–10, 2013. ISSN 19326203. doi: 10.1371/journal.pone.0077244.

242. Raymond Tellier, Yuguo Li, Benjamin J. Cowling, and Julian W. Tang. Recognition of aerosol transmission of infectious agents: A commentary. *BMC Infectious Diseases*, 19(1):1–9, 2019. ISSN 14712334. doi: 10.1186/s12879-019-3707-y.

243. William W. Thompson, David K. Shay, Eric Weintraub, Lynnette Brammer, Nancy J. Cox, and Larry J Anderson. Mortality Associated With Influenza and Respiratory Syncytial Virus in the United States. *American Medical Association*, 289(2):179–186, 2003.

244. Mark Throsby, Edward van den Brink, Mandy Jongeneelen, Leo L M Poon, Philippe Alard, Lisette Cornelissen, Arjen Bakker, Freek Cox, Els van Deventer, Yi Guan, Jindrich Cinatl, Jan ter Meulen, Ignace Lasters, Rita Carsetti, Malik Peiris, John de Kruif, and Jaap Goudsmit. Heterosubtypic neutralizing monoclonal antibodies cross-protective against H5N1 and H1N1 recovered from human IgM+ memory B cells. *PLoS ONE*, 3(12), 2008. ISSN 19326203. doi: 10.1371/journal.pone.0003942.

245. Jerome I Tokars, Sonja J Olsen, and Carrie Reed. Seasonal Incidence of Symptomatic Influenza in the United States. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*, 66(10):1511–1518, may 2018. ISSN 1537-6591 (Electronic). doi: 10.1093/cid/cix1060.

246. Damon J.A. Toth, Molly Leecaster, Warren B.P. Pettey, Adi V Gundlapalli, Hongjiang Gao, Jeanette J Rainey, Amra Uzicanin, and Matthew H Samore. The role of heterogeneity in contact timing and duration in network models of influenza spread in schools. *Journal of the Royal Society Interface*, 12(108), 2015. ISSN 17425662. doi: 10.1098/rsif.2015.0279.

247. Steffen Unkel, C. Paddy Farrington, Paul H Garthwaite, Chris Robertson, and Nick Andrews. Statistical methods for the prospective detection of infectious disease outbreaks: A review, 2012. ISSN 09641998.

248. Sara Y.Del Valle, James M. Hyman, and Nakul Chitnis. Mathematical models of contact patterns between age groups for predicting the spread of infectious diseases. *Mathematical Biosciences and Engineering*, 10(5-6):1475–1497, 2013. ISSN 15471063. doi: 10.3934/mbe.2013.10.1475.

249. Pauline van den Driessche. Reproduction numbers of infectious disease models. *Infectious Disease Modelling*, 2(3):288–303, 2017. ISSN 24680427. doi: 10.1016/j.idm.2017.06.002. URL http://dx.doi.org/10.1016/j.idm.2017.06.002.

250. Johannes van der Pol. *Introduction to Network Modeling Using Exponential Random Graph Models (ERGM): Theory and an Application Using R-Project*, volume 54. Springer US, 2019. ISBN 1061401898. doi: 10.1007/s10614-018-9853-2. URL https://doi.org/10.1007/s10614-018-9853-2.

251. Erik Van Nimwegen, James P. Crutchfield, and Melanie Mitchell. Finite populations induce metastability in evolutionary search. *Physics Letters, Section A: General, Atomic and Solid State Physics*, 229(3):144–150, 1997. ISSN 03759601. doi: 10.1016/S0375-9601(97)00192-8.

252. Tomás Vega, Jose Eugenio Lozano, Tamara Meerhoff, René Snacken, Joshua Mott, Raul Ortiz de Lejarazu, and Baltazar Nunes. Influenza surveillance in Europe: Establishing epidemic thresholds by the Moving Epidemic Method. *Influenza and other Respiratory Viruses*, 7(4):546–558, 2013. ISSN 17502640. doi: 10.1111/j.1750-2659.2012.00422.x.

253. Tomás Vega, José E. Lozano, Tamara Meerhoff, René Snacken, Julien Beauté, Pernille Jorgensen, Raúl Ortiz de Lejarazu, Lisa Domegan, Joël Mossong, Jens Nielsen, Rita Born, Amparo Larrauri, and Caroline Brown. Influenza surveillance in Europe: Comparing intensity levels calculated using the moving epidemic method. *Influenza and other Respiratory Viruses*, 9(5):234–246, 2015. ISSN 17502659. doi: 10.1111/irv.12330.

254. Paul F. Velleman. Definition and Comparison of Robust Nonlinear Data Smoothing Algorithms. *Journal of the American Statistical Association*, 75(371):609–615, sep 1980. ISSN 0162-1459. doi: 10.1080/01621459.1980.10477521. URL http://www.tandfonline.com/doi/abs/10.1080/01621459.1980.10477521.

255. Cécile Viboud, Pierre Yves Boëlle, Khashayar Pakdaman, Fabrice Carrat, Alain Jacques Valleron, and Antoine Flahault. Influenza Epidemics in the United States, France, and Australia, 1972-1997. *Emerging Infectious Diseases*, 10(1):32–39, 2004. ISSN 10806040. doi: 10.3201/eid1001.020705.

256. Cécile Viboud, O. N. Bjornstad, David L. Smith, Lone Simonsen, Mark A. Miller, Bryan T. Grenfell, Ottar N Bjørnstad, David L. Smith, Lone Simonsen, Mark A. Miller, and Bryan T. Grenfell. Synchrony, waves, and spatial hierarchies in the spread of influenza. *Science*, 312(5772):447–451, 2006. ISSN 0036-8075. doi: 10.1126/science.1125237. URL http://www.sciencemag.org/cgi/doi/10.1126/science.1125237.

257. Jean-louis L Virelizier. Host defenses against influenza virus: the role of anti-hemagglutinin antibody. *Journal of immunology (Baltimore, Md. : 1950)*, 115(2):434–439, aug 1975. ISSN 0022-1767 (Print).

258. Erik Volz. SIR dynamics in random networks with heterogeneous connectivity. *Journal of Mathematical Biology*, 56(3):293–310, 2008. ISSN 03036812. doi: 10.1007/s00285-007-0116-4.

259. Erik Volz and Lauren Ancel Meyers. Epidemic thresholds in dynamic contact networks. *Journal of the Royal Society Interface*, 6(32):233–241, mar 2009. ISSN 17425662. doi: 10.1098/rsif.2008.0218.

260. Erik M. Volz, Katia Koelle, and Trevor Bedford. Viral Phylodynamics. *PLoS Computational Biology*, 9(3), 2013. ISSN 1553734X. doi: 10.1371/journal.pcbi.1002947.

261. Michael M. Wagner, Fu Chiang Tsui, Jeremy U. Espino, Virginia M. Dato, Dean F. Sitting, Richard A. Caruana, Laura F. McGinnis, David W. Deerfield, Marek J. Druzdzel, and Douglas B. Fridsma. The emerging science of very early detection of disease outbreaks. *Journal of Public Health Management and Practice*, 7(6):51–59, 2001. ISSN 15505022. doi: 10.1097/00124784-200107060-00006.

262. Xiaoli Wang, Shuangsheng Wu, C Raina MacIntyre, Hongbin Zhang, Weixian Shi, Xiaomin Peng, Wei Duan, Peng Yang, Yi Zhang, and Quanyi Wang. Using an Adjusted Serfling Regression Model to Improve the Early Warning at the Arrival of Peak Timing of Influenza in Beijing. *PLOS ONE*, 10(3):1–14, 2015. doi: 10.1371/journal.pone.0119923. URL https://doi.org/10.1371/journal.pone.0119923.

263. Caroline Watts and Heath Kelly. Fragmentation of influenza surveillance in Australia. *Communicable diseases intelligence*, 26(1):8–13, 2002. ISSN 07253141.

264. Caroline G. Watts, Ross M. Andrews, Julian D. Druce, and Heath A. Kelly. Establishing thresholds for influenza surveillance in Victoria. *Australian and New Zealand Journal of Public Health*, 27(4):409–412, 2003. ISSN 13260200. doi: 10.1111/j.1467-842X.2003.tb00418.x.

265. D C Wiley, I A Wilson, and J J Skehel. Structural identification of the antibody-binding sites of Hong Kong influenza haemagglutinin and their involvement in antigenic variation. *Nature*, 289(5796):373–378, 1981. ISSN 0028-0836. doi: 10.1038/289373a0. URL http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=pubmed{&}cmd=Retrieve{&}dopt=AbstractPlus{&}list{_}uids=6162101.

266. I A Wilson and N J Cox. Structural basis of immune recognition of influenza virus hemagglutinin. *Annual review of immunology*, 8:737–771, 1990. ISSN 0732-0582 (Print). doi: 10.1146/annurev.iy.08.040190.003513.

267. Simon N Wood. Statistical inference for noisy nonlinear ecological dynamic systems. *Nature*, 466(August):1102–1104, 2010. ISSN 0028-0836. doi: 10.1038/nature09319. URL http://www.ncbi.nlm.nih.gov/pubmed/20703226.

268. Daniel John Woods. *An interactive approach for solving multi-objective optimization problems (interactive computer, nelder-mead simplex algorithm, graphics)*. PhD thesis, 1985.

269. World Health Organization. Influenza (Seasonal) Fact sheet, 2018. URL https://www.who.int/news-room/fact-sheets/detail/influenza-(seasonal).

270. Aiping Wu, Yousong Peng, Xiangjun Du, Yuelong Shu, and Taijiao Jiang. Correlation of influenza virus excess mortality with antigenic variation: Application to rapid estimation of influenza mortality burden. *PLoS Computational Biology*, 6(8), 2010. ISSN 1553734X. doi: 10.1371/journal.pcbi.1000882.

271. Nicholas C. Wu, Jakub Otwinowski, Andrew J. Thompson, Corwin M. Nycholat, Armita Nourmohammad, and Ian A. Wilson. Major antigenic site B of human influenza H3N2 viruses has an evolving local fitness landscape. *Nature Communications*, 11(1), 2020. ISSN 20411723. doi: 10.1038/s41467-020-15102-5.

272. Katherine S Xue, Terry Stevens-ayers, Angela P Campbell, Janet A Englund, Steven A Pergam, Michael Boeckh, and Jesse D Bloom. Parallel evolution of influenza across multiple spatiotemporal scales. *eLife*, pages 1–16, 2017. ISSN 2050-084X. doi: 10.7554/eLife.26875.

273. Wan Yang, Subbiah Elankumaran, and Linsey C. Marr. Relationship between Humidity and Influenza A Viability in Droplets and Implications for Influenza's Seasonality. *PLoS ONE*, 7(10):1–8, 2012. ISSN 19326203. doi: 10.1371/journal.pone.0046789.

274. Wan Yang, Marc Lipsitch, and Jeffrey Shaman. Inference of seasonal and pandemic influenza transmission dynamics. *Proceedings of the National Academy of Sciences of the United States of America*, 112(9):2723–2728, 2015. ISSN 10916490. doi: 10.1073/pnas.1415012112.

275. Wan Yang, Eric H Y Lau, and Benjamin J Cowling. Dynamic interactions of influenza viruses in Hong Kong during 1998-2018. *PLoS computational biology*, 16(6):e1007989, 2020. doi: 10.1371/journal.pcbi.1007989. URL http://www.ncbi.nlm.nih.gov/pubmed/32542015.

276. Sanling Yuan, P. van den Driessche, Frederick H. Willeboordse, Zhisheng Shuai, and Junling Ma. Disease invasion risk in a growing population. *Journal of Mathematical Biology*, 73(3):665–681, 2016. ISSN 14321416. doi: 10.1007/s00285-015-0962-4.

277. Xueying Zheng, Zhengyu Song, Yapeng Li, Juanjuan Zhang, and Xi Ling Wang. Possible interference between seasonal epidemics of influenza and other respiratory viruses in Hong Kong, 2014-2017. *BMC Infectious Diseases*, 17(1):1–7, 2017. ISSN 14712334. doi: 10.1186/s12879-017-2888-5.

278. Daniel Zinder, Trevor Bedford, Sunetra Gupta, and Mercedes Pascual. The Roles of Competition and Mutation in Shaping Antigenic and Genetic Diversity in Influenza. *PLoS Pathogens*, 9(1), 2013. ISSN 15537366. doi: 10.1371/journal.ppat.1003104.

279. Casey M Zipfel and Shweta Bansal. Health inequities in influenza transmission and surveillance. *MedRxiv*, page 2020.03.30.20048017, 2020. doi: 10.1101/2020.03.30.20048017. URL http://medrxiv.org/cgi/content/short/2020.03.30.20048017v1?rss=1{&}utm{_}source=researcher{_}app{&}utm{_}medium=referral{&}utm{_}campaign=RESR{_}MRKT{_}Researcher{_}inbound.

Appendix A

Publications by the candidate during the course of this PhD

Papers published from this thesis:

1. Edward K.S. Lam, Dylan H Morris, Aeron C Hurt, Ian G Barr, and Colin A Russell. The impact of climate and antigenic evolution on seasonal influenza virus epidemics in Australia. *Nature Communications*, 11(1), 2020. ISSN 20411723. doi: 10.1038/s41467-020-16545-6. URL http://dx.doi.org/10.1038/s41467-020-16545-6

nature COMMUNICATIONS

ARTICLE

https://doi.org/10.1038/s41467-020-16545-6 OPEN

The impact of climate and antigenic evolution on seasonal influenza virus epidemics in Australia

Edward K. S. Lam [1], Dylan H. Morris [2], Aeron C. Hurt[3,4], Ian G. Barr[3,4,5] & Colin A. Russell [6]✉

Although seasonal influenza viruses circulate globally, prevention and treatment occur at the level of regions, cities, and communities. At these scales, the timing, duration and magnitude of epidemics vary substantially, but the underlying causes of this variation are poorly understood. Here, based on analyses of a 15-year city-level dataset of 18,250 laboratory-confirmed and antigenically-characterised influenza virus infections from Australia, we investigate the effects of previously hypothesised environmental and virological drivers of influenza epidemics. We find that anomalous fluctuations in temperature and humidity do not predict local epidemic onset timings. We also find that virus antigenic change has no consistent effect on epidemic size. In contrast, epidemic onset time and heterosubtypic competition have substantial effects on epidemic size and composition. Our findings suggest that the relationship between influenza population immunity and epidemiology is more complex than previously supposed and that the strong influence of short-term processes may hinder long-term epidemiological forecasts.

[1] Department of Veterinary Medicine, University of Cambridge, Cambridge, UK. [2] Department of Ecology and Evolutionary Biology, Princeton University, Princeton, NJ, USA. [3] WHO Collaborating Centre for Reference and Research on Influenza, VIDRL, Peter Doherty Institute for Infection and Immunity, Melbourne, VIC, Australia. [4] Department of Microbiology and Immunology, University of Melbourne, Parkville, VIC, Australia. [5] School of Applied Biomedical Sciences, Federation University, Churchill, VIC, Australia. [6] Department of Medical Microbiology, Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands. ✉email: c.a.russell@amsterdamumc.nl

Seasonal influenza virus epidemics are a substantial source of disease burden and result in ~650,000 deaths each year[1]. Four co-circulating subtypes/lineages of influenza viruses currently cause disease in humans: A/H3N2 (A/H3), A/H1N1 (currently A/H1pdm09, previously A/H1seasonal (A/H1sea)), B/Victoria/2/87-like (B/Vic) and B/Yamagata/16/88-like (B/Yam) viruses. The timing, duration and size of local influenza virus epidemics can vary substantially from year to year[2,3], but the underlying causes of this variation are poorly understood. Better understanding of the factors that govern epidemic onset and magnitude could allow for accurate and timely epidemiological forecasts[4] and more efficient allocation of public health resources[5].

In temperate regions of the Northern and Southern Hemispheres, influenza virus activity is most common in winter months, but the mechanistic basis of this seasonality remains unclear. Experimental studies demonstrated that reductions in temperature and absolute humidity enhance viral stability and aerosol transmission[6–8]. However, epidemics in tropical and subtropical regions often occur during periods of high temperature and humidity[9].

Climatic fluctuations have been implicated as triggers for influenza epidemics in temperate regions. A study of state-level epidemiological data from the United States found that influenza epidemics sometimes follow 2-week periods of anomalously low absolute humidity[10]. Subsequent studies of epidemiological activity have found similar results using prefecture-level data from Japan[11], city-level data from the New York Metropolitan Area[12] and region-level data from France[13].

Influenza virus evolutionary dynamics are another theorised driver of influenza virus epidemiology. Within each type and subtype of seasonal influenza virus, new major antigenic variants arise every 3–8 years[14,15]. New variants partially escape the immunity induced by prior infections and vaccinations, rendering a higher fraction of individuals susceptible to infection. Epidemiological theory predicts that epidemics caused by a new antigenic variant should therefore be larger than epidemics of previously circulated variants[16,17].

Antigenic change could also produce earlier and more spatiotemporally synchronous epidemics. When more individuals are susceptible, fewer transmission chains go stochastically extinct, so each new introduction of a virus into a population has a higher chance of causing an epidemic. Consistent with this, studies have suggested that antigenic change is associated with earlier epidemics in Israel[18], and with more synchronous epidemics among cities in the United States[19–21], Japan[22] and Australia[23].

Studies of environmental and virological drivers of influenza virus epidemiology, including the studies referenced above, have been limited by three factors: (1) the reliance on influenza-like illness (ILI) data, (2) the aggregation of ILI or virologically confirmed data over large geographical scales (state/province/country) and (3) where virologically confirmed data are available, the use of data without subtype and antigenic variant-level resolution.

ILI data frequently include a wide variety of respiratory infections[24], and limited laboratory characterisation obscures influenza virus type/subtype- and antigenic variant-specific patterns. These patterns become superimposed upon each other due to aggregation of ILI or virologically confirmed data to ecological scales (state/province/country) that sum over multiple local epidemics (county/city/town), which can individually vary substantially in timing, magnitude and influenza virus composition. Altogether, these sources of obfuscation make it difficult to disentangle local-level, antigenic variant-specific patterns, and critically investigate the impact of virus antigenic change.

Here, we use a 15-year data set of 18,250 typed, subtyped and antigenically characterised seasonal influenza viruses from the five most populous cities in Australia to investigate the impact of environmental and virological factors on the timing and magnitude of city-level influenza virus epidemics. We find that climatic fluctuations and virus antigenic change have no consistent effects on epidemic onset timing or size, while epidemic onset timing itself and heterosubtypic competition have substantial impacts on epidemic size and virus subtype composition. The lack of consistent effect of easily measured climatic and virus antigenic properties, and seeming dominance of noisy short-term transmission processes likely diminishes the feasibility of meaningful long-term influenza epidemic forecasting at local scales.

Results

Australia laboratory-confirmed influenza. We aggregated 18,250 laboratory-confirmed and antigenically characterised cases of seasonal influenza viruses from 2000 to 2015 by 2-week (14-day) periods, creating a set of subtype- and antigenic variant-specific time series for the five most populous cities in Australia: Sydney (~5.5 million people), Melbourne (~5.0 million), Brisbane (~2.4 million people), Perth (~2.3 million) and Adelaide (~1.4 million) (Fig. 1). We excluded all virus cases from the 2009 season from all analyses because the 2009 A/H1N1 virus pandemic was atypical compared with seasonal epidemics and likely to be driven by different processes, affecting both epidemic dynamics and data collection of A/H1pdm09, as well as the other subtypes. Using a Poisson count detection method (see 'Methods'), we identified periods of sustained, above-baseline levels of epidemic activity for each antigenic variant in each city. To facilitate comparisons among cities, we calculated the laboratory-confirmed incidence per 10^6 individuals using the annual estimated resident population values of each city[25].

Epidemic magnitude and the most common virus subtype varied substantially among cities (Fig. 1). For example, during the 2002 season, A/H3 and B/Vic viruses were the most common strains in both Brisbane and Sydney. Absolute A/H3 virus incidence in Brisbane was much higher than in Sydney (186 vs 38.0 cases per 10^6 individuals), as was absolute B/Vic incidence (40.3 vs 22.7 cases per 10^6 individuals). But B/Vic had a substantially higher relative incidence in Sydney than in Brisbane (37% of all cases, vs only 18%). In some seasons, a virus antigenic variant caused a major epidemic in one or more cities, but failed to produce any observable above-baseline activity in another city. For example, in 2006, the A/Wisconsin/67/2005 (H3N2) virus variant caused epidemics in Brisbane, Perth and Melbourne, while above-baseline levels of activity were completely absent in Adelaide.

Effect of climatic factors. Epidemic onset timing varied substantially within and among cities and virus subtypes (Supplementary Fig. 1). Previously, Shaman et al.[10] showed that the 2-week period preceding the onset of state-level ILI epidemics in the United States was often marked by unusually low temperatures (T) or absolute humidities (AH). Fluctuations in these climatic factors from the historic averages expected for that specific day of the year (T′ and AH′, respectively) were anomalously large and negative when compared against a bootstrapped distribution of random samples from the historical records of observed daily climatic fluctuations recorded over wintertime (defined as 1 October–28 February). Following the same bootstrap-sampling method (see 'Methods') and aggregating epidemics across all five Australian cities, there were no statistically significant differences (all $P > 0.05$, see Supplementary Table 1) between the bootstrapped distribution of random samples of typical wintertime fluctuations (1 April–31 August for Australia) and the observed fluctuations in anomalous temperature and absolute humidity

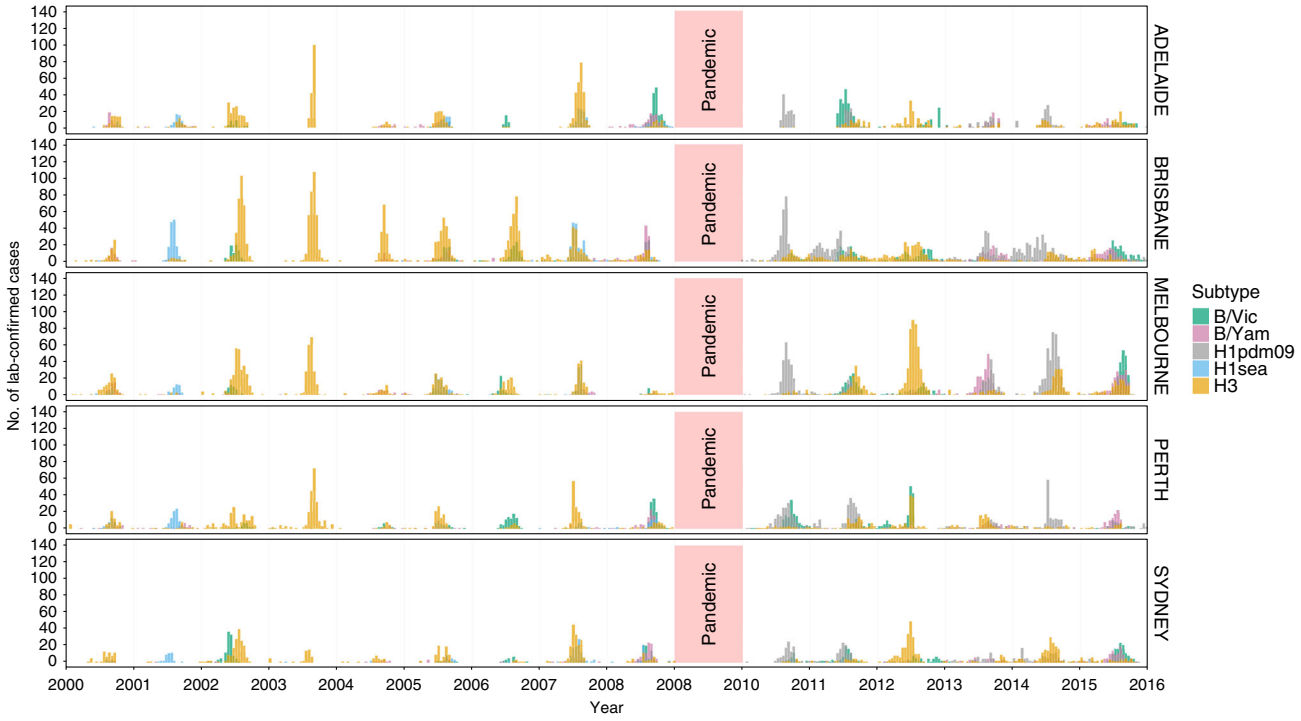


Fig. 1 Number of laboratory-confirmed seasonal influenza virus infections from 2000 to 2015 for the five largest cities in Australia. Cases are aggregated by 2-week periods, stratified by city and coloured by subtype/lineage.

over the 2-, 4- and 6-week periods immediately prior to the onset of the earliest epidemics from 2000 to 2015 (excluding 2009, 15 years × 5 cities = 75 epidemics in total) (Fig. 2). Individual city-by-city analyses (Supplementary Fig. 2 and Supplementary Table 2) showed that there was substantial local variation but no consistent patterns. Epidemic onset times coincided with both high and low temperature and absolute humidity periods, and there were no statistically significant patterns in four of the five cities.

Even if anomalous fluctuations in temperature and humidity do not necessarily affect epidemic onset, climatic factors could have an impact on virus transmission[7] and overall epidemic size: for example, influenza mortality in New York Metropolitan Area was shown to be negatively associated with temperature and humidity[12]. Overall, epidemic incidence should depend strongly on the initial exponential growth phase of the epidemic, where transmission may be facilitated by favourable climatic conditions. We therefore investigated the impact of mean temperature and mean absolute humidity during each epidemic, as well as just the period from epidemic onset to the peak, on that epidemic's size. For both time periods considered, epidemic incidence was not associated with mean absolute humidity (Supplementary Fig. 3). We found that epidemic incidence was weakly negatively associated with the mean temperature during the epidemic and the period from start to the peak, but this relationship appears to be primarily driven by two instances, where small epidemics occurred during the early and warmer part of the season; on balance, the highly variable epidemic sizes observed over a range of climatic conditions, suggest that climatic factors have limited and noisy effects (Supplementary Fig. 3).

A recent study by Geoghegan et al.[23] estimated epidemic onset timings for influenza A virus epidemics in Australian postcodes for the seasons from 2007 to 2016. Despite the lack of subtype-level resolution, their data set is substantially larger (450,000 entries) than the one used here, and offers an opportunity to compare findings. We repeated our anomalous temperature and absolute humidity analyses on the Geoghegan et al.[23] data set. As with our original data set, there were no consistent statistically significant relationships between climate anomalies and epidemic onset (Supplementary Discussion, Supplementary Tables 3, 4, Supplementary Figs. 4 and 5).

Other climatic factors have been proposed as drivers of influenza dynamics, notably relative humidity and rainfall[6,9]. We repeated the above analyses for relative humidity and rainfall. There were some city-level associations, but no consistent pattern and no pattern when aggregating across cities. Epidemic onset was not associated with statistically significant fluctuations in anomalous relative humidity and rainfall.

Effect of antigenic change. We next examined the effect of antigenic evolution on epidemic dynamics. Between 2000 and 2015, 7A/H3, 3A/H1sea, 1A/H1pdm09, 3 B/Vic and 5 B/Yam virus antigenic variants circulated in Australia. All A/H1pdm09 virus epidemics from 2009 to 2015 were excluded for this set of analyses for two reasons. First, we could not accurately estimate the size of the 2009 pandemic. Second, there was no subsequent, detectable antigenic change observed for A/H1pdm09 viruses during the study period. We normalised epidemic sizes (see 'Methods') to enable comparisons between cities. Stratifying by subtype/lineage, we compared the size of the first epidemic caused by an antigenic variant against the sizes of epidemics of the same antigenic variant in subsequent years (Wilcoxon two-sample test, Fig. 3). Contrary to the predictions of previous theoretical studies[16,17], newly emerged antigenic variants caused epidemics, both larger and smaller than city-specific mean epidemic sizes, and there was no evidence of a consistent effect of antigenic change on epidemic size.

We also compared the timing of the first epidemic caused by an antigenic variant against the timings of subsequent epidemics (Supplementary Fig. 6) to test the hypothesis that new variants cause earlier epidemics. The range of onset timings was very

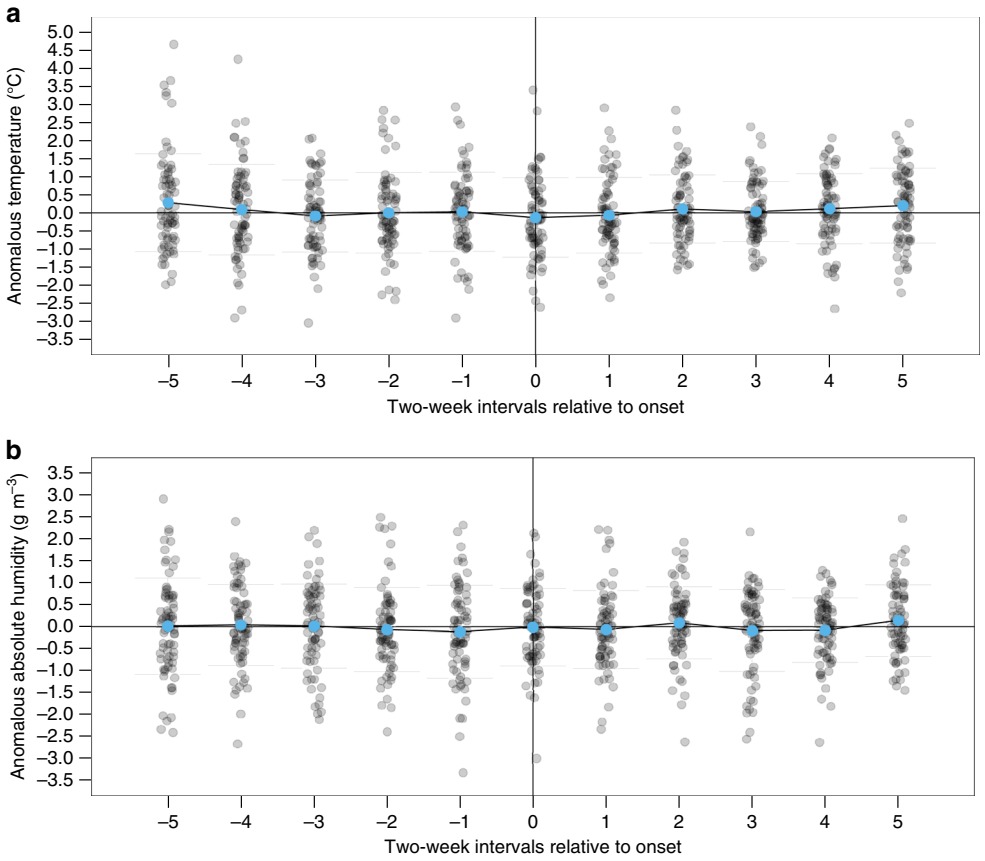


Fig. 2 Climatic conditions around epidemic onset. a Anomalous temperature T′ and **b** absolute humidity AH′ prior to and after epidemic onset across all five cities. Epidemic onset is marked by the vertical line at 0. For the earliest-onset epidemic in each season and city (15 years × 5 cities = 75 epidemics), T′ and AH′ for each time point are represented by grey points: a point below the horizontal line denotes that the value is lower than the 31-year city-specific mean. Blue points show the mean T′ and AH′ for that 2-week period for all epidemics across all cities within the study period. There were no periods with statistically significant reductions ($P < 0.05$; Non-parametric bootstrapping) in T′ or AH′ from the 31-year averages.

broad, with epidemics starting from very early to late into the season, and there were no statistically significant differences in epidemic onset timing between new and extant variant epidemics.

To investigate the impact of antigenic change on the spatiotemporal synchrony of epidemics, we examined the timing of epidemic activity across cities for years when a new major antigenic variant circulated in all five cities. New antigenic variants often failed to initiate epidemics across all five cities in a given year. We compared the synchrony of epidemics (defined as the reciprocal of the variance in epidemic onset timings) in the season in which an antigenic variant first emerges to the synchrony in subsequent seasons. There were no statistically significant differences in epidemic synchrony associated with antigenic novelty (Supplementary Fig. 7).

To check the robustness of this result, we repeated these analyses using estimated-onset timings from Geoghegan et al.[23]. There was again no discernible effect of antigenic change on the timing or synchrony of epidemics (Supplementary Discussion).

Effect of prior immunity. After an antigenic variant causes an epidemic in a city for the first time, the accumulated population immunity to that variant should lead to smaller subsequent epidemics, and eventually render further epidemics of that variant less likely. For each epidemic caused by a given antigenic variant, we investigated the relationship between that epidemic's size and the cumulative number of cases caused by that antigenic variant in preceding seasons. To account for differences in population size and surveillance intensity among cities, we normalised epidemic and cumulative case counts by the city-specific mean epidemic size. Antigenic variants that emerged prior to the start of the study period, such as A/Moscow/10/99 (A/H3) and A/New Caledonia/20/99 (A/H1sea) and all A/H1pdm09 epidemics from 2009 to 2015 were excluded from this analysis, since it was not possible to calculate cumulative case counts for them. Specific B/Yam antigenic variants rarely caused more than one epidemic in a given city, but specific antigenic variants of A/H3 and B/Vic viruses caused repeated epidemics in the same city. For A/H3 and B/Vic viruses, epidemic size and cumulative prior incidence were not correlated (Pearson's correlation test, Fig. 4).

The accumulation of population immunity should also reduce the probability of successful epidemic initiation, making epidemics, regardless of size, less likely to start after an antigenic variant has already caused an epidemic in that city. For B/Vic and A/H1sea viruses, binary logistic regression showed non-significant associations between the cumulative incidence over prior seasons and the probability of successful epidemic initiation (all OR < 1; all $P > 0.05$, Supplementary Fig. 8 and Supplementary Table 5). This partially resulted from the small number of A/H1sea epidemics during the study period, most of which were caused by newly emerged antigenic variants. However, B/Yam and H3 viruses showed significant negative relationships between cumulative prior incidence and epidemic probability, suggesting that prior incidence may have a substantial impact on the probability of successful epidemic initiation.

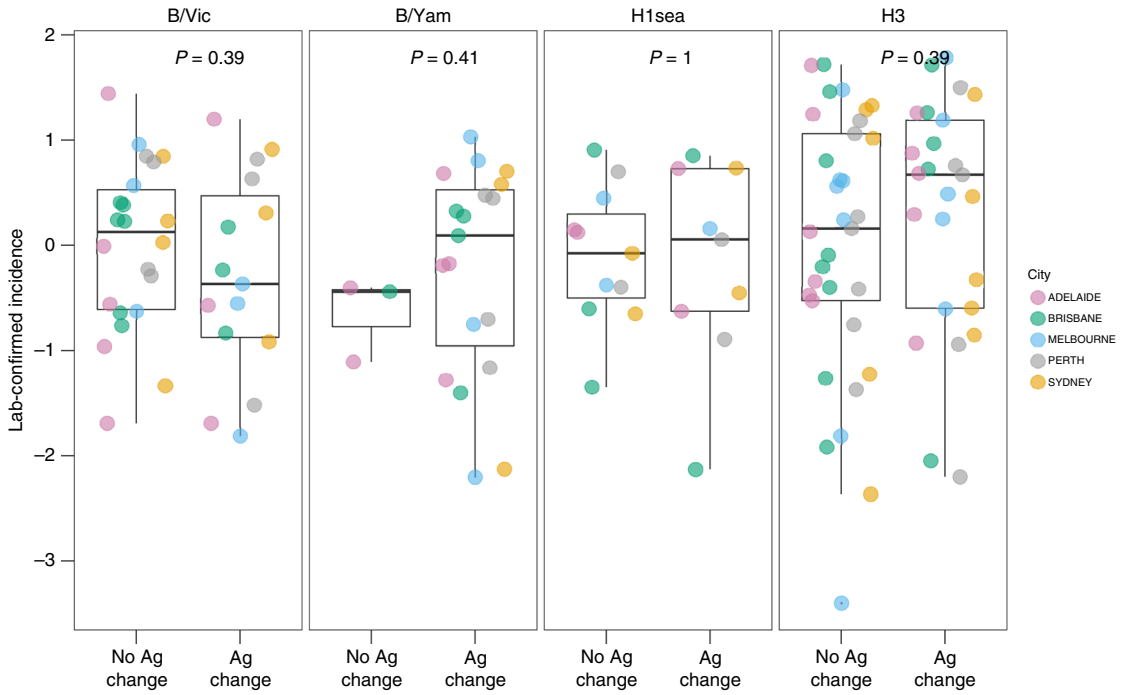


Fig. 3 Effect of antigenic change on epidemic incidence. Epidemic incidence was compared between seasons associated with and without the epidemic-level circulation of a new major antigenic variant. Within each subtype, the incidence for individual epidemics was log-transformed and subtracted by the city-specific mean of log incidence, to allow for comparison between cities. *P* values are from Wilcoxon two-sample tests ($n = 37$, 26, 22 and 63 for B/Vic, B/Yam, A/H1sea and A/H3, respectively). Each point corresponds to one epidemic in a city, and the box plots show the median, first and third quartile of the transformed values, and range.

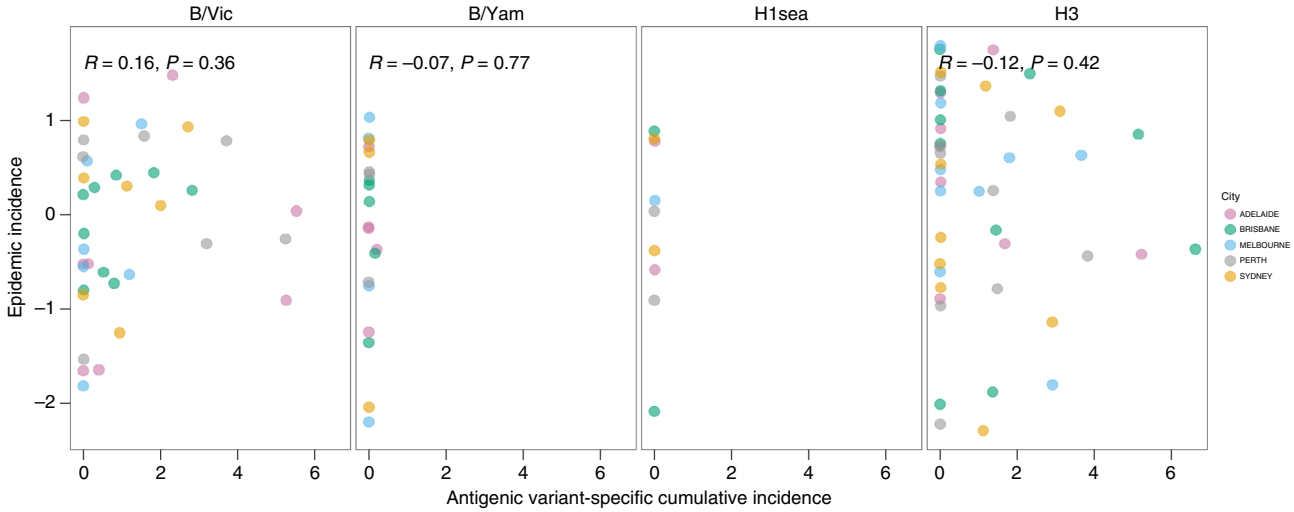


Fig. 4 Effect of prior immunity on epidemic incidence. Within each subtype, the incidence for individual epidemics was log-transformed and subtracted by the city-specific mean of log incidence, to allow for comparison between cities. Antigenic variant-specific cumulative incidence was measured relative to the city-specific mean epidemic size, where 1 is equivalent to the mean epidemic incidence. *r*- and *P* values are from Pearson's correlation tests ($n = 37$, 20, 9 and 45 for B/Vic, B/Yam, A/H1sea and A/H3, respectively). Antigenic variants of B/Yam rarely initiated multiple epidemics during the study period, and it was not possible to calculate a correlation coefficient for A/H1sea because the one new antigenic variant to emerge during the study period caused only a single epidemic per city.

Aggregating across subtypes. There may be subtype/lineage-specific differences in the effect of antigenic change and prior immunity. Notably, B/Yam antigenic variants typically cause only one epidemic per city. We repeated these analyses with epidemics aggregated together, across all subtypes and cities to increase statistical power (see the project Github repository for the analyses and code). As before, there were no statistically significant differences in the magnitude of epidemics between the first and subsequent epidemics of an antigenic variant, or any association between epidemic size and the cumulative incidence over prior seasons. Binary logistic regression showed that the probability of successful epidemic initiation may be moderately reduced by the cumulative incidence over prior seasons. Our findings were robust to the method of normalisation used to allow for comparison between cities and subtypes/lineages (see 'Methods').

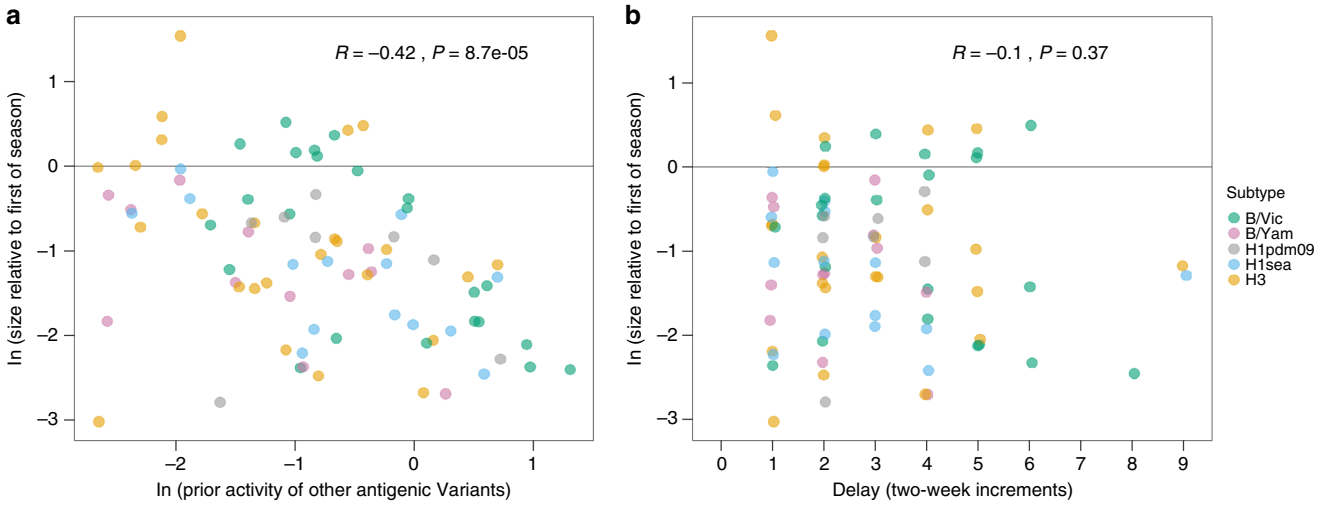


Fig. 5 Effect of competition among subtypes on epidemic incidence. The relationship between the size of an epidemic and **a** the amount of prior activity of all other antigenic variants and subtypes, and **b** the delay in epidemic onset. The size of each epidemic, relative to the earliest epidemic of that season, was log-transformed: the horizontal line at 0 denotes that the size of an epidemic is equal to that of the earliest epidemic of that season and city. In panel (**a**), prior activity by other subtypes within the same season was measured relative to the city-specific mean epidemic size. In panel **b**, delay in epidemic onset was measured relative to the onset timing of the earliest epidemic of that season. *r*- and *P* values are from Pearson's correlation tests ($n = 82$).

Effect of competition among subtypes. Competition among virus subtypes for hosts should create a first-mover advantage for the first subtype to sustain above-baseline epidemic activity in a city in a given season. Subsequent epidemics of other subtypes within that same season should therefore be reduced in size. We considered two proxies for this kind of intersubtypic interference: the cumulative amount of epidemic activity prior to the onset of a subtype's epidemic and the lag between the focal epidemic and the season's earliest epidemic. To allow for comparisons across cities and subtypes, we normalised log-epidemic case counts by subtracting off the city- and subtype-specific mean log- epidemic case count. There was a strongly negative and statistically significant correlation between prior epidemic activity and epidemic size (Pearson's correlation test, $r = -0.420$; $P = 8.7e{-}5$, Fig. 5). An important caveat is that seasonality in the transmission rate could result in epidemics that start later in a season being smaller than those that started earlier, regardless of intersubtypic competition.

Joint contributions of climatic and virological factors. Whilst the magnitude of the effects of the climatological and virological factors may be individually subtle, it could be the case that they are only able to affect observable changes on the magnitude and timing of epidemics when acting in concert, or that large effects in opposing directions mask each other. We used a Bayesian multilevel regression model to identify which putative predictor variables affected epidemic incidence, and estimate posterior distributions for their effects on epidemic size. The model included the following variables: antigenic change, cumulative prior cases of the antigenic variant, mean absolute humidity during the epidemic, activity by other subtypes earlier in the season, epidemic start date and rainfall during the epidemic. Mean temperature during the epidemic was omitted as a predictor, since it was highly collinear with absolute humidity; analyses were subsequently repeated using mean temperature, and omitting absolute humidity with no substantial changes in the overall results.

The model suggested that epidemics that were the first of the season or had early start dates should be modestly larger (Fig. 6). Start date had the largest estimated effect and the clearest posterior support for a non-trivial effect size. Posterior modes for the mean effects of antigenic change and absolute humidity across subtypes were near zero (Fig. 6), with tight credible intervals (95% credible intervals: (−0.56, 0.27) for absolute humidity, (−0.50, 0.30) for antigenic change). Prior cases of an old variant given no antigenic change (95% credible intervals (−0.26, 0.85)), prior cases of all variants for non-first epidemics (95% credible intervals (−0.84, 0.33)) and rainfall during the epidemic (95% credible intervals (−0.68, 0.20)) also showed no strongly discernible effects, though with less posterior certainty. The model could not explain much of the variation in the data: the median-estimated standard deviation of log epidemic size about the expected log size is 0.77 (95% credible intervals (0.67, 0.90)). Since exp(0.77) is ~2.15, this implies that it is not unusual to see epidemics half or twice the expected incidence. The model estimated that the effects were very similar across subtypes (Supplementary Fig. 14, median-estimated SDs for the distribution of subtype-specific effect sizes about the overall mean effect size near zero, Supplementary Fig. 15). Only the effect of whether an epidemic was the first of the season showed meaningful heterogeneity: the model estimated that it is somewhat weaker for B/Vic than for other subtypes (Supplementary Fig. 14).

Discussion

Based on city-level analyses of a subtyped and antigenically characterised influenza virus data set covering the five largest cities in Australia, we find that climate and antigenic novelty have limited effects on epidemic sizes. The results presented here suggest that, at least in temperate areas, epidemics are governed by factors other than host immunity at local scales, where global fitness advantages for new antigenic variants may not be realised. Conversely, competition for hosts among influenza virus types and subtypes has strong effects on local dynamics. The first virus subtype to establish above-baseline epidemic activity in a city and season typically dominates.

A recent study of fine-scale influenza epidemiology in Australia[23] showed that there was substantial heterogeneity among Australian cities in the activity of influenza A and B viruses. Our subtyped and antigenically characterised data set allowed us to confirm that further heterogeneity exists at the level of antigenic

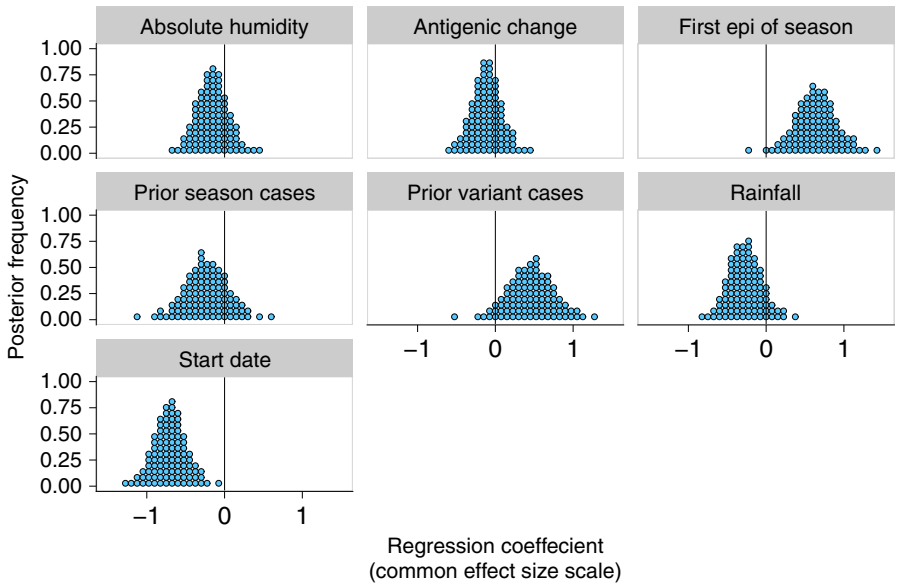


Fig. 6 Joint contributions of climatic and virological factors on epidemic incidence. The mean effects across all subtypes were estimated using the Bayesian multilevel model. Predictors were mean-centred and scaled, so effect sizes are shown on a common scale.

variants. In particular, specific antigenic variants often cause large epidemics in some cities while not causing detectable activity in others.

While prior studies found that the onset of epidemics in the United States and France was preceded by a 2-week period of anomalously low absolute humidity[10,13], we found no evidence for climatic effects when aggregating across the five Australian cities. Anomalous fluctuations in temperature and absolute humidity were sometimes positive, sometimes negative, but on average approximately zero. Importantly, the overall effect size reported by Shaman et al.[10], after aggregating across all 48 contiguous states of the United States, was very small (with mean AH′ being ~ $-0.25\,\mathrm{kg\,kg^{-1}}$ or $-0.21\,\mathrm{g\,m^{-3}}$, compared against $0\,\mathrm{g\,m^{-3}}$, the mean of the null distribution of historic wintertime values). About 55–60% of epidemics were preceded by negative AH′ values: a moderate increase upon the null hypothesis being a baseline of 50%.

Shaman et al.[10] also found regional differences in the associations between fluctuations in absolute humidity and epidemic onset. Strong associations were found in the Southeastern United States but not in Western states. In Australia, there does not appear to be an aggregate effect at the country level, and there were no consistent patterns at the level of individual cities (Supplementary Fig. 2 and Supplementary Table 2). The small effect sizes and lack of consistency in climatic patterns across regions and cities in the United States and Australia may reflect the fact that climatic factors alone are unlikely to account for the differences in the patterns of influenza seasonality between temperate and tropical regions[26].

Seasonal epidemic waves in the United States appear to begin in the Southern states, which have warmer and more humid climates[21,27], casting some doubt on the role of low humidity as a trigger for influenza epidemics. Rather than acting as specific triggers, it is plausible that climatic factors are acting on longer timescales than the anomalous fluctuations reported by Shaman et al.[10] to more generally enhance transmission and increase incidence[28]. However, in Australia, epidemic size does not appear to be strongly associated with the mean temperature or absolute humidity over the epidemic period.

Given the interest in influenza virus as a model system for phylodynamics of a pathogen that consists of multiple co-existing antigenic variants[29], there is interest in understanding how competition between these related variants, typified by cross-immunity, shapes epidemiological dynamics. Studies have hypothesised that antigenic change should result in larger[16,30–32] and earlier[18] local epidemics, which exhibit greater spatiotemporal synchrony at the national level[19–23]. The sequential replacement of old antigenic variants by new ones is indicative that antigenicity and population immunity are important for the global-level phylodynamics of influenza viruses. In contrast, at the local level, we find for A/H3 and B/Vic viruses that neither antigenic change nor the accumulation of antigenic variant-specific immunity are strong drivers of epidemic size, though accumulating variant-specific immunity may moderately reduce the probability of successful epidemic initiation.

It is striking that individual antigenic variants of A/H3 and B/Vic viruses are capable of re-invading the same city multiple times over consecutive years, despite a lack of substantial antigenic change. A/H1pdm09 viruses had previously been shown to cause repeat epidemics without antigenic change[33,34], but our study establishes that this occurs for multiple types and subtypes of human influenza. One possible explanation for the lack of evidence for the year-on-year depletion of susceptible hosts is that influenza virus infection often fails to confer strain-specific and effective immunity. In some individuals, antigenic seniority and existing immunity against previously encountered antigenic variants may suppress novel strain-specific antibody responses, leading to only modest specific protection against reinfection[35,36]. Similarly, vaccine trials suggest that multiple exposures can be required in order for children to become seropositive sufficiently to protect themselves[37]. Potentially, multiple natural infections may also be needed to confer protective immunity[38], particularly in children[39], who effectively form a non-depleting pool of susceptibles.

There may also be substantial and previously unaccounted heterogeneity in individual susceptibility towards the same virus strain. The notion that population-level strain-specific immunity to influenza viruses is monolithic may be an artefact of the single-infection ferret models typically used to estimate antibody-mediated protection. In humans, there is substantial individual-to-individual variation in the antigenicity of amino acid escape mutations for influenza haemagglutinin[40]. Such heterogeneity

between individuals stems from their varied exposure histories to different influenza viruses. Unfortunately, age records for our data set were too incomplete to allow us to study age-specific heterogeneities in demographics, and attack rates between cities, and whether such patterns change over seasons.

Spatial and social connectivity structures among hosts in a city may also limit the spread of epidemics. Heterogeneous contact patterns between hosts can have a substantial impact on the resulting epidemiological dynamics[41,42]. Epidemics may be inherently frail processes: relatively minor human behavioural or environmental perturbations could prematurely terminate epidemics before they exhaust the pool of susceptible hosts, preserving a substantial number of susceptibles, and permitting subsequent epidemics of the same antigenic variant.

While our data set is substantially smaller (>450,000 vs 18,250 cases) than the one analysed by Geoghegan et al.[23], and is thus more likely to be affected by noise in epidemic and surveillance processes, the differences between our findings and theirs highlight the importance of subtyping and antigenic characterisation, particularly for drawing conclusions about the effects of antigenic change. Geoghegan et al.[23] had cautiously suggested, given only virus-type data, that the 2009, 2012 and 2014 influenza A virus epidemics in Australia exhibited greater spatiotemporal synchrony potentially due to the emergence of the novel A/H1pdm09 subtype in 2009 and novel A/H3 antigenic variants in 2012 and 2014. However, with further subtype resolution and antigenic characterisation, we find that the majority of influenza A activity in Adelaide and Melbourne in 2014 was attributable to A/H1pdm09, rather than the (antigenically novel) A/H3; in fact, there was no above-baseline A/H3 activity in Perth. The fact that different virus subtypes caused these apparently synchronous epidemics implies that the epidemic synchrony described by Geoghegan et al.[23] was not due to the antigenic evolution or regional spread of a single virus strain.

Apart from competition between antigenic variants, previous epidemiological studies have hypothesised the existence of heterosubtypic competition where prior infection by a virus of one subtype is negatively associated with subsequent infection by a virus of another subtype[43,44]. In agreement with a previous US study of national-level ILI activity augmented with limited virus subtyping[45], we also find evidence for a first-mover advantage and competition to infect hosts within a city, where the subtype or type that initiates above-baseline levels of activity first is most likely to have the largest epidemic of that season.

There are multiple caveats to our study that merit explicit consideration. The most important ones derive from our use of passive surveillance data that might not accurately reflect true underlying influenza virus activity. For example, surveillance intensity could plausibly vary between cities and years. While variation in surveillance efforts is evident among cities, there was no evidence of systematic increases or decreases in the number of laboratory-confirmed cases, or changes to surveillance practices within each city during the study period. Despite this, the longer duration of epidemics recorded after 2009 could be indicative of enhanced surveillance in the post-pandemic era: to mitigate this possibility, we repeated our analyses on the effect of antigenic change on epidemic size, splitting between pre- and post-pandemic eras and epidemic sizes normalising by their respective era-specific means. In either era, there was no consistent effect of antigenic change on epidemic size, with the caveat that splitting across eras reduced the number of observations in each era and thus our statistical power (Supplementary Fig. 18).

The intensity of surveillance could also vary over the course of an epidemic. For example, sentinel physicians could become more likely to submit samples for further testing as an epidemic unfolds, or conversely, testing could prematurely cease as facilities become overwhelmed with samples. Despite being unable to definitively rule out the former scenario, the latter is unlikely to affect our data. If reporting ceased after a certain number of samples had been tested, the distribution of epidemic sizes would be truncated, and each epidemic would be unlikely to have an exponentially declining tail. No such patterns exist in our data.

The intensity of surveillance could also potentially vary across subtypes and lineages. The mean age of infection for A/H3 is greater than influenza B[46] viruses, and healthcare-seeking behaviour may differ between adults, parents with children and children. Furthermore, it is commonly thought that A/H3 virus infections result in more severe clinical presentations and greater risk of mortality[47] than influenza B viruses, potentially resulting in differences in the likelihood of detection by a sentinel health practitioner, though this may not be the case (see ref. [48]).

Another important caveat is that while we were able to include antigenic data in this study, these data were all derived from haemagglutination inhibition (HI) assays. HI assays do not measure virus antigenic changes that occur away from the receptor-binding site, and thus likely represent an incomplete picture of antigenic change. Reference viruses and sera used in the haemagglutination inhibition assays can also impact the interpretation of the assay readout, and the HI data used in this study were therefore treated with caution (see 'Methods').

In this study, we attempted to identify associations between population susceptibility and epidemic incidence. Accurately quantifying the former is a complex challenge, so cumulative antigenic variant-specific epidemic incidence was used as a proxy, but that itself is subject to the limitations listed above. Besides natural infection, immunity can also be derived from vaccination, the contribution and effectiveness of which could not be determined due to a lack of temporally and geographically complete vaccination records over the study period. Regardless, we hypothesise that the impact of seasonal vaccination would be limited, particularly in the context of Australia, given the low uptake of vaccination[49]. Crucially, the uptake by children, who are important in driving local community transmission, is often below 10%[50].

While our Bayesian multilevel model estimated negligible effects on epidemic size stemming from climatic factors and prior cases attributed to the same antigenic variant, the estimated credible intervals were not tight enough to rule out these effects conclusively (Fig. 6). However, our study suggests that climatic and antigenic factors are unlikely to be strong drivers of local influenza epidemiological dynamics. Indeed, the effects of these specific factors are dwarfed in magnitude by more generic epidemiological drivers: seasonality not directly captured by climate (measured by start date) and competition for hosts among subtypes (measured by whether an epidemic is the first of the season) (Fig. 6). We also find that even with all generic and specific factors considered, precise predictions of epidemic size remain difficult because of substantial noise in the local epidemic process.

Our Bayesian multilevel model for epidemic size avoids explicitly modelling underlying transmission processes, and may fail to fully capture the nature of the relationship (linear vs nonlinear) between transmission rates/R_0 and the total cases in an epidemic. However, based on previous virus-transmissibility studies[7], if climatic factors are strong drivers of epidemiological dynamics, we would expect the climatic variabilities observed in Australia to have a substantial impact on transmission rates, and produce detectable differences in epidemic size, but this is not the case.

Climatic drivers of seasonality and homosubtypic competition between virus antigenic variants are thought to be strong drivers of seasonal influenza epidemiology, but seasonal influenza virus epidemiological dynamics in major Australian cities appear to be

more substantially shaped by other factors, particularly the establishment of sustained virus-transmission activity, and subsequent competition among virus types and subtypes. This implies that the time horizon for meaningful forecasting of epidemic subtype composition is very short (days to weeks), and forecasting efforts aimed at longer-term predictions will require further insights into the dynamics of virus introduction and epidemic establishment, and into the accumulation of population immunity to seasonal influenza viruses.

Methods

Australian surveillance data. Influenza viruses from Australia were collected by the WHO Collaborating Centre (WHOCC) for Reference and Research on Influenza in Melbourne, Australia. The Melbourne WHOCC receives a subset of influenza-positive clinical samples collected by various sentinel surveillance systems across Australia throughout the year. The samples in this study were typed, subtyped and antigenically characterised by haemagglutination inhibition assay to the vaccine reference vaccine strain in use at the time of sample collection.

The data set consists of 18,250 influenza-positive cases, collected between 2000 and 2015 in the city of Brisbane, the city of Perth, the state of South Australia, the city of Sydney and the state of Victoria. The breakdown at the subtype/lineage level is as follows: A/H3 (7661), A/H1sea (1410), A/H1pdm09 (3987), B/Vic (3021) and B/Yam (2171). All of these correspond specifically to individual cities, except for the data from Victoria and South Australia. As of June 2015, 75 and 78% of the inhabitants of the states of Victoria and South Australia resided in the cities of Melbourne and Adelaide, respectively. We therefore treated the Victoria and South Australia data as representative of city-level patterns in those two major cities.

All epidemic activity of all subtypes for the 2009 season was excluded from all analyses because of the 2009 A/H1N1 pandemic. Unsurprisingly, patterns of virus circulation during the pandemic were anomalous compared with typical seasonal influenza virus epidemics, and potentially distortive of the patterns we sought to characterise.

Estimation of epidemic timing. The exact timing of interseasonal periods of sporadic activity and epidemic onset for each subtype is highly variable between years, even for individual cities, so it is necessary to determine the onset and end of each epidemic independently for each antigenic variant, season and city.

For each individual antigenic variant-specific time series, we used a Poisson count detection algorithm implemented in the Surveillance package in R[51,52] to distinguish periods of sustained epidemic activity from a background of sporadic interseasonal activity. We assume that the start of the calendar year falls sometime within the interseasonal period, which is justified by the scarce number of cases observed during this time of the year, and the fact that it is summertime in Australia. Making no further assumptions on the exact duration and timing for the interseasonal period or epidemic onset, starting at the beginning of the year, successive 2-week periods y_t are evaluated using the number of cases in each of the n-preceding 2-week periods $\{y_{t-n}, y_{t-n+1}, \cdots, y_{t-2}, y_{t-1}\}$ as reference values for sporadic activity. These reference values are used to predict a threshold value y_α: if the observed number of cases y_t exceeds the threshold y_α, the focal 2-week period is marked as the 2-week period of epidemic onset.

The Poisson count model assumes that the reference values y_t are identically and independently Poisson distributed with a mean of λ (Eq. (1)). λ itself has a Gamma distribution as a prior (Eq. (2)). From Eqs. (1) and (2), the posterior predictive distribution is a negative binomial distribution (Eq. (3)).

$$y_i \sim Po(\lambda) \tag{1}$$

$$\lambda \sim Ga(\alpha, \beta) \tag{2}$$

$$z|y_{t-n}, y_{t-n+1}, \cdots, y_{t-2}, y_{t-1} \sim NegBin\left(\alpha + \sum_{i=1}^{n} y_i, \frac{\beta + n}{\beta + n + 1}\right) \tag{3}$$

The threshold value y_α can then be calculated using quantile parameter α, where y_α is the smallest value that satisfies Eq. (4).

$$p(y \le y_\alpha) \ge 1 - \alpha \tag{4}$$

We used the same algorithm to identify the end of an epidemic. Starting at the end of the year, successive 2-week periods, in the backward direction, are evaluated using the number of cases in each of the n following the 2-week period as reference values $\{y_{t+1}, y_{t+2}, \cdots, y_{t+n-1}, y_{t+n}\}$.

During interseasonal periods, where there were often many 2-week periods reporting no cases, an isolated 2-week period with sporadic activity can be misconstrued as the onset of an epidemic. To reduce the impact of outliers in the time series and increase specificity of the detection algorithm, we first applied the 4253H, twice nonlinear data-smoothing algorithm[53], which is a compound smoother consisting of multiple running medians.

We tested a variety of n- and α-parameter values, and chose $n = 3$ and $\alpha = 0.12$ for the analyses presented in the text as a good compromise between sensitivity and specificity in the identification of all of the epidemics within the time series and their individual onset and end timings, which were confirmed by visual inspection. The results of these analyses are also robust to alternative parameter values and corresponding changes to the sensitivity and specificity of the Poisson count detection algorithm (see Supplementary Discussion for sensitivity analyses).

Aggregation of cases by 2-week periods was deemed necessary, to smoothen the time series in light of the relatively low number of cases within the data set; this relatively long timescale could however potentially obscure fluctuations in weather that occur at shorter scales. Whilst weekly time series were appreciably noisier, we found a high degree of correspondence in the estimated epidemic onset and end timings with values calculated from data aggregated by 2-week periods: indeed, our results were robust to aggregation by week (see Supplementary Discussion for sensitivity analyses).

We deemed an antigenic variant to have failed to cause an epidemic if, within a season, the algorithm was unable to define an epidemic period; we confirmed all putative failures by inspection of the raw time series. Once the epidemic period was defined, the size of an epidemic per antigenic variant was calculated using the estimated resident population for that particular year and city.

Normalisation of epidemic incidence. For each epidemic, the incidence of laboratory-confirmed cases per million people was calculated from the number of raw counts. Given the positive skew in the distribution of epidemic incidences, individual incidence values were log-transformed. To enable comparisons within subtypes, we needed to account for potential differences in surveillance intensity, and normalise values between cities: we subtracted off the overall city-specific mean log-transformed incidence from each individual value. Although the apparent heterogeneity in the effect of antigenic change and prior immunity between subtypes suggests that data should be stratified by subtype, we repeated our analyses with data aggregated and normalised across subtypes in order to increase statistical power. Individual log-transformed values for each epidemic were instead transformed by subtracting off the overall city- and subtype-specific mean of the log-transformed values.

Virus antigenic characterisation by haemagglutination inhibition assay. For our analyses, we defined an antigenic variant as in Smith et al.[14], where an antigenic variant is sufficiently different from preceding variants to warrant an update of the seasonal influenza virus vaccine. To this end, our analyses only accounted for major antigenic changes, and did not account for the possibility of small or gradual antigenic changes (neither of which are well studied for seasonal influenza viruses).

The haemagglutination inhibition (HI) assay data used in this study only compared the test virus and the then current reference vaccine strain to assess whether or not viruses had changed antigenically. However, this comparison to a single reference point is potentially problematic, given that new Southern Hemisphere's influenza vaccine composition recommendations are made every September. This is usually after the end of the influenza season in Australia, and may lead to misidentification during antigenic characterisation of submitted samples during the preceding season where samples containing a novel antigenic variant may have been tested with sera raised against its predecessor variant. To ameliorate this potential source of bias, we compared the antigenic characterisation data against phylogenetic data. This comparison revealed two instances for A/H3 viruses where the reference strain comparison by HI was misleading regarding the antigenic composition of an epidemic. There were a substantial number of laboratory-confirmed cases attributable to A/H3/Fujian/411/2002-like viruses in 2004, but phylogenetic analyses of sequences dated 2004 show that the Fujian/411/2002-like viruses had already been replaced by the novel California/7/2004 variant viruses. Similarly, in 2005, a substantial number of samples initially identified as A/H3 California/7/2004-like viruses were phylogenetically in the new A/Wisconsin/67/2005 variant group.

To account for the likelihood of misidentification due to delays in updating nomenclature, we assumed that all A/H3 cases in 2004 were California/7/2004-like, and in 2005 were Wisconsin/67/2005-like antigenic variants. Additional analyses were also carried out with the raw data set without these corrections (see Supplementary Figs. 9–13 and Supplementary Table 6), and lead to no significant or substantive differences to our findings.

Demographic data. We retrieved estimated resident populations for Adelaide, Brisbane, Melbourne, Perth and Sydney on 30 June of each year from 2000 to 2015 from the Australian Bureau of Statistics (http://stat.abs.gov.au/).

Climate data. For each of the five cities, we compiled the mean temperature (°C) and relative humidity (%) from TuTiempo (https://en.tutiempo.net/), and calculated the mean absolute humidity (g m^{-3}) for each 2-week period from 1985 to 2015. For each of the 26 2-week periods of the calendar year, we calculated 31-year mean temperature $\bar{T}$ and absolute humidity $\overline{AH}$ values (see Eqs. (5) and (6) below).

Testing the statistical significance of anomalous absolute humidity and temperature. Following the method presented by Shaman et al.[10], we calculated local anomalous T' and absolute humidity AH' values for each city, and 2-week period of the year from 2000 to 2015. For each 2-week period, T' and AH' are

defined as the deviation in observed temperature T and actual absolute humidity AH from their 31-year mean values, $\bar{T}$ and $\overline{AH}$, respectively (Eqs. (5) and (6)).

$$T' = T - \bar{T} \tag{5}$$

$$AH' = AH - \overline{AH} \tag{6}$$

Following Shaman et al.[10], we generated a synthetic distribution of wintertime climatic values by bootstrap sampling. In order to maintain the sampling structure and control for anomaly variability among the cities, 15 n-week continuous blocks were randomly sampled from 1 April–31 August, 1985–2015 for each of the five cities. These 75 samples were then averaged to produce a mean T' and AH' value. This was repeated 100,000 times to produce a bootstrapped distribution of average values. The statistical significance for the mean T' and AH' values derived from the 75 empirically observed earliest-in-the-season epidemics was then calculated non-parametrically, by determining the quantile for the observed values within the bootstrap distributions. This bootstrap was repeated at the city level to see if there were geographical differences with individual bootstrap distributions that were created for each city.

We also evaluated whether or not epidemic onset is associated with climatic fluctuations that are anomalous for that particular time of the year. By definition, for any given 2-week period of the year, the 31-year mean for T' and AH' is 0. We used a Wilcoxon one-sample test to assess whether there were reductions in climatic values in the observed set of T' and AH' values in each of the 2-week blocks preceding the onset of the earliest epidemic of the season.

Bayesian hierarchical regression. To estimate reasonable bounds on the possible effects of climate and antigenicity on epidemic size, we used a Bayesian hierarchical model that partially pooled effect-size estimates across subtypes, increasing the capacity to detect any potential effects without assuming a priori that effects should be the same across different subtypes. We fit the model using Markov Chain Monte Carlo (MCMC) with Stan[54] and its R interface rstan[55]; Stan implements a no-u-turn sampler (NUTS)[54]. All data and code needed to reproduce the analysis and figures are provided in the project Github repository, along with directions in a README file.

In the model notation that follows, the symbol "~" is a "sampling statement"; it denotes that a random variable is distributed according to the given distribution. Normal distributions are parameterised as Normal(mean, standard deviation), generalised Student-T distributions are parameterised as Student-T(degrees of freedom, location and scale). Positive-constrained normal distributions (Half-Normal) are parameterised as Half-Normal(mode, standard deviation).

We predicted log incidence minus city- and subtype-specific mean log incidence as a function of the following predictor variables:

X_1: whether the epidemic was the first epidemic for an antigenic variant in the city (binary, yes or no)
X_2: cumulative prior incidence of the antigenic variant (measured as log(total prior cases/city- and subtype-specific mean cases per epidemic))
X_3: mean absolute humidity during the epidemic, from the start to end date of the epidemic (measured as fortnight of the year)
X_4: start date of the epidemic (measured as fortnight of the year)
X_5: whether the epidemic was the earliest epidemic (of any subtype) in the city that year (binary, yes or no)
X_6: the cumulative amount of influenza activity (of any subtype) in the city that year prior to the epidemic
X_7: mean rainfall during the epidemic, from the start to end date of the epidemic (measured as fortnight of the year).

We omitted mean epidemic temperature as a predictor as it was highly collinear with absolute humidity. Any observed large effect of absolute humidity could therefore theoretically have been attributable to temperature, though in practice we estimated an effect near zero for absolute humidity.

We made a linear prediction of an epidemic's normalised size given its values for $\boldsymbol{X} = (X_1,\ldots,X_7)$. Effect sizes b_i for each predictor X_i were subtype-specific, with b_{ij} denoting the effect of variable i for subtype j. We also estimated subtype-specific intercepts a_j.

We included cumulative antigenic variant activity and prior activity in the year only for old antigenic variants and epidemics that were not first of the year, respectively, that is, as interaction terms with one minus the corresponding binary variables. So the predicted mean- centred log size $<y_k>$ of an epidemic of subtype j is given by Eq. (7), where X_{ik} denotes the value of X_i for epidemic k. Following Gelman[51], we mean-centred and scaled continuous predictors so that effect sizes b would be directly comparable between binary and continuous predictors.

$$\begin{aligned}\langle y_k \rangle &= a_j + b_{1j}X_{1k} + b_{2j}X_{2k}(1 - X_{1k}) + b_{3j}X_{3k} + b_{4j}X_{4k} + b_{5j}X_{5k} \\ &\quad + b_{6j}X_{6k}(1 - X_{5k}) + b_{7j}X_{7k}\end{aligned} \tag{7}$$

We assumed that observed normalized log epidemic sizes y_k were normally distributed about their predicted log sizes $<y_k>$ with an unknown, estimated standard deviation σ_y (Eq. (8)):

$$y_k \sim \text{Normal}(\langle y_k \rangle, \sigma_y) \tag{8}$$

We assumed that subtype effect sizes b_{ij} for each predictor i and subtype j were normally distributed about a general mean effect size $<b_i>$, with an unknown, estimated predictor-specific standard deviation σ_{bi} (Eq. (9)):

$$b_{ij} \sim \text{Normal}(\langle b_i \rangle, \sigma_{bi}) \tag{9}$$

Likewise, we assumed that intercepts a_i were normally distributed about a mean intercept $<a>$ with an unknown, estimated standard deviation σ_a (Eq. (10)).

$$a_i \sim \text{Normal}(\langle a \rangle, \sigma_a) \tag{10}$$

We assumed that predictor-specific effect-size standard deviations σ_{bi} were half-normally distributed with mode 0 and an unknown, estimated standard deviation σ_b (Eq. (11)).

$$\sigma_{bi} \sim \text{Half} - \text{Normal}(0, \sigma_b) \tag{11}$$

We placed weakly informative[56] positive-constrained half-normal priors on the intercept, effect size and error-term standard deviations σ_a, σ_b and σ_y (Eqs. (12–14)). Weakly informative priors rule out biologically or mathematically implausible parameter values while allowing data rather than assumptions to inform inferences regarding plausible values.

$$\sigma_a \sim \text{Half} - \text{Normal}(0, 0.5) \tag{12}$$

$$\sigma_b \sim \text{Half} - \text{Normal}(0, 1) \tag{13}$$

$$\sigma_y \sim \text{Half} - \text{Normal}(0, 1) \tag{14}$$

We placed a weakly informative Gaussian prior on the mean intercept $<a>$ (Eq. (15)) and a weakly informative Student-T prior on the mean effect sizes $<b_i>$ (Eq. (16)):

$$\langle a \rangle \sim \text{Normal}(0, 1) \tag{15}$$

$$\langle b_j \rangle \sim \text{Student} - \text{T}(3, 0, 2.5) \tag{16}$$

The intercept prior was based on the degree of variation in the normed outcome variable to cover it while ruling out intercepts much larger or smaller than the largest and smallest observations. The effect-size prior was based on a recommendation for weakly informative regression effect-size priors (for scaled predictors) from the Stan prior recommendation wiki (https://github.com/stan-dev/stan/wiki/Prior-Choice-Recommendations).

We ran four MCMC chains, each with a 1000-step sample warmup period followed by 1000 saved posterior samples, for a total of 4000 posterior draws. We verified convergence by inspecting trace plots, and confirming that all parameters had sufficiently low Rhat values (all Rhat < 1.005) and sufficiently large effective sample sizes (all neff >16% of total sample size). We visualised posteriors as quantile dotplots[57] to aid in visual estimation of distributions.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

All of the data for these statistical analyses and models are available at the following Github repository: https://github.com/edwardkslam/australian_seasonal_flu.

Code availability

Code developed for these statistical analyses and models is available at the following Github repository: https://github.com/edwardkslam/australian_seasonal_flu.

Received: 15 August 2019; Accepted: 9 May 2020;
Published online: 02 June 2020

References

1. World Health Organization. Influenza (Seasonal) Fact sheet. https://www.who.int/news-room/fact-sheets/detail/influenza-(seasonal) (2018).
2. Denoeud, L. et al. Predicting pneumonia and influenza mortality from morbidity data. *PLoS ONE* **2**, 3–6 (2007).
3. Finkelman, B. S. et al. Global patterns in seasonal activity of influenza A/H3N2, A/H1N1, and B from 1997 to 2005: Viral coexistence and latitudinal gradients. *PLoS One* **2**, e1296 (2007).
4. Gandon, S., Day, T., Metcalf, C. J. E. & Grenfell, B. T. Forecasting epidemiological and evolutionary dynamics of infectious diseases. *Trends Ecol. Evol.* **31**, 776–788 (2016).
5. Public Health England. Flu plan winter 2016/17. *Public Heal. Engl.* https://doi.org/10.1037/e500942012-001 (2016).

6. Lowen, A. C., Mubareka, S., Steel, J. & Palese, P. Influenza virus transmission is dependent on relative humidity and temperature. *PLoS Pathog.* **3**, 1470–1476 (2007).
7. Shaman, J. & Kohn, M. Absolute humidity modulates influenza survival, transmission, and seasonality. *Proc. Natl Acad. Sci. USA* **106**, 3243–3248 (2009).
8. Lowen, A. C. & Steel, J. Roles of humidity and temperature in shaping influenza seasonality. *J. Virol.* **88**, 7692–7695 (2014).
9. Tamerius, J. D. et al. Environmental predictors of seasonal influenza epidemics across temperate and tropical climates. *PLoS Pathog.* **9**, e1003194 (2013).
10. Shaman, J., Pitzer, V. E., Viboud, C., Grenfell, B. T. & Lipsitch, M. Absolute humidity and the seasonal onset of influenza in the continental United States. *PLoS Biol.* **8**, e1000316 (2010).
11. Shoji, M., Katayama, K. & Sano, K. Absolute humidity as a deterministic factor affecting seasonal influenza epidemics in Japan. *Tohoku J. Exp. Med.* **224**, 251–256 (2011).
12. Davis, R. E., Rossier, C. E. & Enfield, K. B. The impact of weather on influenza and pneumonia mortality in New York City, 1975–2002: a retrospective study. *PLoS ONE* **7**, e34091 (2012).
13. Seleznev, N. E. & Leonenko, V. N. Absolute humidity anomalies and the influenza onsets in Russia: a computational study. *Procedia Comput. Sci.* **119**, 224–233 (2017).
14. Smith, D. J. Mapping the antigenic and genetic evolution of influenza virus. *Science* **305**, 371–376 (2004).
15. Langat, P. et al. Genome-wide evolutionary dynamics of influenza B viruses on a global scale. *PLoS Pathog.* **13**, e1006749 (2017).
16. Koelle, K., Cobey, S., Grenfell, B. & Pascual, M. Epochal evolution shapes the phylodynamics of interpandemic influenza A (H3N2) in humans. *Science* **6829**, 1898–1904 (2006).
17. Koelle, K., Kamradt, M. & Pascual, M. Understanding the dynamics of rapidly evolving pathogens through modeling the tempo of antigenic change: influenza as a case study. *Epidemics* **1**, 129–137 (2009).
18. Bock Axelsen, J., Yaari, R., Grenfell, B. T. & Stone, L. Multiannual forecasting of seasonal influenza dynamics reveals climatic and evolutionary drivers. *Proc. Natl Acad. Sci. USA* **111**, 9538–9542 (2014).
19. Greene, S. K., Ionides, E. L. & Wilson, M. L. Patterns of influenza-associated mortality among US elderly by geographic region and virus subtype, 1968-1998. *Am. J. Epidemiol.* **163**, 316–326 (2006).
20. Viboud, C. et al. Synchrony, waves, and spatial hierarchies in the spread of influenza. *Science* **312**, 447–451 (2006).
21. Charu, V. et al. Human mobility and the spatial transmission of influenza in the United States. *PLoS Comput. Biol.* **13**, 1–23 (2017).
22. Sakai, T. et al. Geographic and temporal trends in influenzalike illness, Japan, 1992-1999. *Emerg. Infect. Dis.* **10**, 1822–1826 (2004).
23. Geoghegan, J. L. et al. Continental synchronicity of human influenza virus epidemics despite climactic variation. *PLoS Pathog.* **14**, 1–16 (2018).
24. Schnepf, N. et al. High burden of non-influenza viruses in influenza-like illness in the early weeks of H1N1v epidemic in France. *PLoS ONE* **6**, e23514 (2011).
25. Australian Bureau of Statistics. ERP by SA2 and above (ASGS 2001), 1991 to 2016. http://www.abs.gov.au (2017).
26. Tamerius, J. et al. Global influenza seasonality: reconciling patterns across temperate and tropical regions. *Environ. Health Perspect.* **119**, 439–445 (2011).
27. Chattopadhyay, I., Kiciman, E., Elliott, J. W. & Shaman, J. L. Conjunction of factors triggering waves of seasonal influenza. *eLife* **7**, e30756 (2018).
28. Deyle, E. R., Maher, M. C., Hernandez, R. D., Basu, S. & Sugihara, G. Global environmental drivers of influenza. *Proc. Natl Acad. Sci. USA* **113**, 13081–13086 (2016).
29. Kucharski, A. J., Andreasen, V. & Gog, J. R. Capturing the dynamics of pathogens with many strains. *J. Math. Biol.* **72**, 1–24 (2016).
30. Wu, A., Peng, Y., Du, X., Shu, Y. & Jiang, T. Correlation of influenza virus excess mortality with antigenic variation: application to rapid estimation of influenza mortality burden. *PLoS Comput. Biol.* **6**, e1000882 (2010).
31. Bedford, T. et al. Integrating influenza antigenic dynamics with molecular evolution. *eLife* **3**, 1–26 (2014).
32. Yang, W., Lipsitch, M. & Shaman, J. Inference of seasonal and pandemic influenza transmission dynamics. *PLoS Natl Acad. Sci. USA* **112**, 2723–2729 (2015).
33. Barr, I. G. et al. A new pandemic influenza A(H1N1) genetic variant predominated in the winter 2010 influenza season in Australia, New Zealand and Singapore. *Eurosurveillance* **15**, 1–6 (2010).
34. Hoschler, K. et al. Seroprevalence of influenza A (H1N1) pdm09 Virus. *Emerg. Infect. Dis.* **18**, 2010–2013 (2012).
35. Lessler, J. et al. Evidence for antigenic seniority in influenza A (H3N2) antibody responses in southern China. *PLoS Pathog.* **8**, 26 (2012).
36. Kucharski, A. J. et al. Estimating the life course of influenza A(H3N2) antibody responses from cross-sectional data. *PLoS Biol.* **13**, 1–16 (2015).
37. Neuzil, K. M. M. M. et al. Immunogenicity and reactogenicity of 1 versus 2 doses of trivalent inactivated influenza vaccine in vaccine-naive 5–8-year-old children. *J. Infect. Dis.* **194**, 1032–1039 (2006).
38. Gouma, S., Kim, K., Weirick, M., Gumina, M. E. & Branche, A. Middle-aged individuals may be in a perpetual state of H3N2 influenza virus susceptibility. Preprint at https://www.medrxiv.org/content/10.1101/2020.01.09.20017038v1.full.pdf (2020).
39. Meade, P. et al. Influenza virus infection induces a narrow antibody response in children but a broad recall response in adults. *mBio* **11**, 1–15 (2020).
40. Lee, J. M. et al. Mapping person-to-person variation in viral mutations that escape polyclonal serum targeting influenza hemagglutinin. *eLife* **8**, e49324 (2019).
41. Dalziel, B. D. et al. Contact heterogeneity, rather than transmission efficiency, limits the emergence and spread of canine influenza virus. *PLoS Pathog.* **10**, e1004455 (2014).
42. Dalziel, B. D. et al. Urbanization and humidity shape the intensity of influenza epidemics in U.S. cities. *Science.* **362**, 75–79 (2018).
43. Sonoguchi, T., Naito, H., Hara, M., Takeuchi, Y. & Fukumi, H. Cross-subtype protection in humans during sequential, overlapping, and/or concurrent epidemics caused by h3n2 and h1n1 influenza viruses. *J. Infect. Dis.* **151**, 81–88 (1985).
44. Cowling, B. J. et al. Protective efficacy of seasonal influenza vaccination against seasonal and pandemic influenza virus infection during 2009 in Hong Kong. *Clin. Infect. Dis.* **51**, 1370–1379 (2010).
45. Goldstein, E., Cobey, S., Takahashi, S., Miller, J. C. & Lipsitch, M. Predicting the epidemic sizes of influenza A/H1N1, A/H3N2, and B: a statistical method. *PLoS Med.* **8**, 1–12 (2011).
46. Bedford, T. et al. Global circulation patterns of seasonal influenza viruses vary with antigenic drift. *Nature* **523**, 217–220 (2015).
47. Thompson, W. W. et al. Mortality associated with influenza and respiratory syncytial virus in the United States. *Am. Med. Assoc.* **289**, 179–186 (2003).
48. Caini, S., Kroneman, M., Wiegers, T., El Guerche-Séblain, C. & Paget, J. Clinical characteristics and severity of influenza infections by virus type, subtype, and lineage: a systematic literature review. *Influenza Other Respir. Viruses* **12**, 780–792 (2018).
49. Barr, I. G., Vijaykrishna, D. & Sullivan, S. G. Differential age susceptibility to influenza B/victoria lineage viruses in the 2015 Australian influenza season. *Eurosurveillance* **21**, 1–9 (2016).
50. Sullivan, S. G., Chilver, M. B. N., Higgins, G., Cheng, A. C. & Stocks, N. P. Influenza vaccine effectiveness in Australia: results from the Australian Sentinel Practices Research Network. *Med. J. Aust.* **201**, 109–111 (2014).
51. Höhle, M., Riebler, A. & Paul, M. Getting started with outbreak detection. https://cran.r-project.org/web/packages/surveillance/vignettes/surveillance.pdf (2007).
52. Salmon, M., Schumacher, D. & Höhle, M. Monitoring count time series in *R*: aberration detection in public health surveillance. *J. Stat. Softw.* **70**, 10 (2016).
53. Velleman, P. F. Definition and comparison of robust nonlinear data smoothing algorithms. *J. Am. Stat. Assoc.* **75**, 609 (1980).
54. Stan Development Team. Stan Modeling Language Users Guide and Reference Manual. https://mc-stan.org/users/documentation/ (2018).
55. Stan Development Team. RStan: the R interface to Stan. http://mc-stan.org/ (2018).
56. Gelman, A., Jakulin, A., Pittau, M. G. & Su, Y.-S. A weakly informative default prior distribution for logistic and other regression models. *Ann. Appl. Stat.* **2**, 1360–1383 (2008).
57. Fernandes, M., Walls, L., Munson, S., Hullman, J. & Kay, M. Uncertainty displays using quantile dotplots or CDFs improve transit decision-making. In *Proc. 2018 CHI Conference on Human Factors in Computing Systems* 1–12, https://doi.org/10.1145/3173574.3173718 (ACM Press, 2018).

Acknowledgements

The authors thank the WHO Collaborating Centre for Reference and Research on Influenza (Victorian Infectious Diseases Reference Laboratory, Melbourne, Australia), which is supported by the Australian Government Department of Health, for the virological analysis of samples submitted by sentinel health practitioners and surveillance systems.

Author contributions

E.K.S.L., D.H.M. and C.A.R. designed the research; E.K.S.L. and D.H.M. performed the data analyses; A.C.H. and I.G.B. generated the influenza virus data; E.K.S.L., D.H.M. and C.A.R. wrote the first draft of the paper. All authors contributed to the critical review and revision of the paper.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41467-020-16545-6.

Correspondence and requests for materials should be addressed to C.A.R.

Peer review information *Nature Communications* thanks the anonymous reviewer(s) for their contribution to the peer review of this work.

Reprints and permission information is available at http://www.nature.com/reprints

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

Appendix B

Appendix to Chapter 4

B.1 Robustness of inferences derived from my epidemic onset timings

Our antigenically characterised data set is relatively small: 18,250 cases. Especially in seasons with fewer cases, it can be difficult to differentiate epidemic from baseline activity. This limits the accuracy with which the timing of epidemic onset can be estimated. To check the robustness of our results to errors in estimated onset, I re-ran our analysis using estimated influenza A epidemic onset timings from a large-scale study of >450,000 Australian influenza cases[92].

Whilst the data set utilised by Geoghegan et al.[92] has many more cases than our dataset and thus might produce more accurate timing estimates, the lack of subtype level resolution means that the city-level epidemic activity recorded was the summation of underlying A/H3, A/H1sea and A/H1pdm09 virus specific activity. Nevertheless, I investigated whether or not, more generally, the onset of influenza A epidemic activity from 2007 to 2015 was preceded by periods of anomalous climatic conditions (Tables B.3–B.4 and Figs. B.4–B.5). There were no biologically significant effect sizes or consistent patterns in T' or AH' prior to epidemic onset across the five cities, whether be it comparing against broader wintertime conditions (Tables B.3–B.4), as in[222], or against long term average conditions for that particular time of the year (Figs. B.4–B.5).

To further assess the robustness of our analyses, towards potential inaccuracies in our estimates of epidemic onset timings arising from our relatively small data set, I augmented our estimated epidemic onset timings with those of Geoghegan et al.[92] and repeated our climatic factor analyses using several different methods to impute epidemic activity within a season to a particular virus subtype. For each city and season, 1. I assumed that our

timing estimate for the dominant influenza A subtype was incorrect and replaced it with the Geoghegan et al.[92] estimate; 2. I assumed that our timing estimate for the influenza A subtype that initiates epidemic activity earliest was incorrect and replaced it with the Geoghegan et al.[92] estimate; 3. in years in which the number of cases for the dominant influenza A subtype were small or it was difficult to discern the period of epidemic from background activity, I assumed that our timing estimate was incorrect and replaced it with the Geoghegan et al.[92] estimate.

After substituting our estimates with the Geoghegan et al.[92] timings, in concordance with the above assumptions, I proceeded to identify the subtype/lineage (A/H3, A/H1sea, A/H1pdm09, B/Victoria, B/Yamagata) that initiated epidemic activity earliest within a season and re-ran our climatic analyses, comparing the climatic conditions in the two week periods preceding the earliest epidemic against average wintertime and average climatic values for that time of year (see Section 4.2; these additional analyses can be reproduced by code included in the project Github repository). These assumed scenarios had little impact on the set of epidemic timings used for downstream analyses, due to the limited number of seasons considered by Geoghegan et al.[92]: 8 seasons from 2007-2015, since 2009 was omitted. Furthermore, discrepancies between our estimates and the Geoghegan et al.[92] values are inconsequential whenever epidemic activity is first initiated by an influenza B virus lineage during a season.

I also repeated our analyses on the effect of antigenic change on the local timing of epidemics and temporal synchrony of epidemics between cities. Again, I substituted a subset of our estimated epidemic timings with timings from Geoghegan et al.[92], based on the same 3 sets of assumptions mentioned above. Across all replicates and set of assumptions, there was no evidence that antigenic change resulted in earlier local epidemics or more temporally synchronous epidemic activity across cities (these additional analyses can be reproduced by code included in the project Github repository). Overall, after adjusting for multiple testing (Holm correction), the findings presented in the main text are robust against potential inaccuracies in our estimates of epidemic onset timings arising from our relatively small data set.

B.2 Sensitivity analysis of epidemic onset and end detection of algorithm

The threshold value y_α, which if exceeded marks the onset of an epidemic, and thus the sensitivity of the detection algorithm are determined by the quantile parameter α (see

Eq. (3.16); Section 4.2). In the main text, I chose $\alpha = 0.12$, since it identified epidemic onset and end timings that corresponded well with visual inspection of the raw time series. I repeated the estimation of onset and end timings using $\alpha = 0.05\,\&\,0.2$, which increased and reduced threshold values respectively. Overall, these alternative timings were similar to those originally estimated with $\alpha = 0.12$. However, timing estimates were found to be systematically earlier when utilising lower threshold values due to an increase in sensitivity. At the same time, this also reduced the specificity of the detection algorithm: spurious non-epidemic activity early in the calendar year were conflated as periods of above-baseline levels of epidemic activity and recorded as small sized epidemics on multiple occasions, (Fig. B.16).

I reran our bootstrap analyses on the effects of climatic factors with these alternative epidemic timings. The estimated epidemic onset and end timings remained largely invariant to changes in α so it was unsurprising that I did not identify any fluctuations in anomalous temperature and absolute humidity in the two, four and six week periods immediately prior to the onset of the earliest epidemics from 2000-2015.

At a lower value of $\alpha = 0.05$, which increased the threshold value for detection and the specificity of the algorithm, I similarly found no evidence of consistent effects of antigenic change on epidemic size (Wilcoxon two-sample test) . In contrast, when $\alpha = 0.2$, it appeared that for B/Yam and A/H3, the epidemics were of greater size in seasons associated with the emergence of a new antigenic variant (Wilcoxon two-sample test; Fig. B.17). However, this is likely to be an artefact of the reduced specificity of the detection algorithm, which inflated the number of seasons in which small so-called epidemics were detected.

I aggregated the data by week and by two-week periods and found that the latter produced smoother time series: this reduced the effect of stochastic noise and made it more amenable for use with our detection algorithm. Aggregation by two-week periods could however obscure fluctuations in local weather, which are likely to occur at shorter timescales. Reassuringly however, I found that the detection of epidemics and estimated timings corresponded well between values calculated from data aggregated by two-week periods and by week: 239/320 instances had identical results, whilst in only 43/320 instances did timing estimates differ by more than 14 days. These relatively minor differences in timing estimates did not impact our results: I did not identify any fluctuations in anomalous temperature and absolute humidity in the two, four and six week periods (two-week aggregation) or in the one, two and three week periods (weekly aggregation) immediately prior to the onset of the earliest epidemics from 2000-2015.

Overall, our detection algorithm and downstream results from the analyses on the effects of climatic factors and antigenic change remain robust to choice of time period for the

aggregation of case counts and the selection of alternative parameters, which alter the sensitivity and specificity of the algorithm (analyses can be reproduced from code included in the project Github repository).

B.3 Appendix B Figures

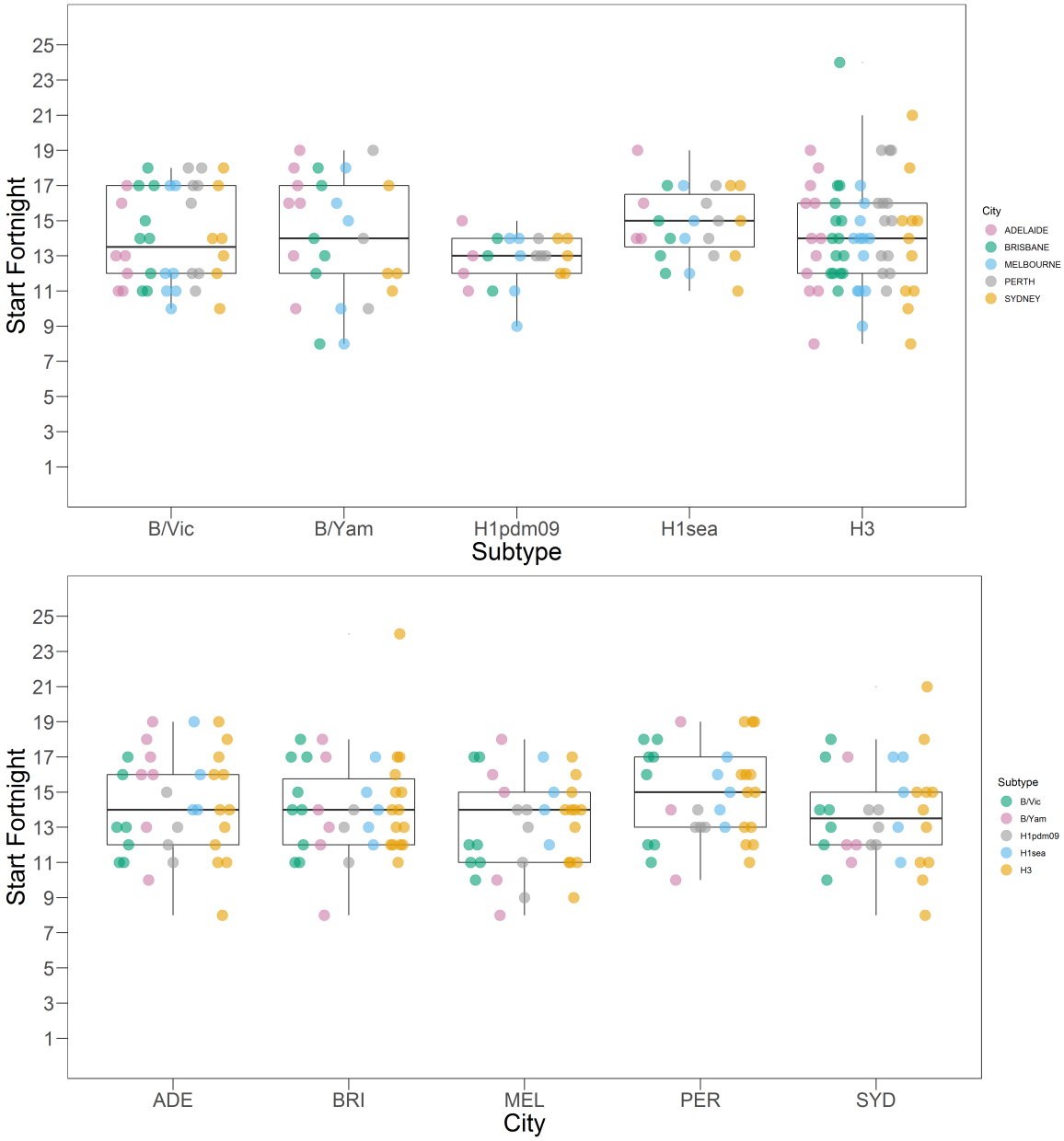


Fig. B.1 Comparison of epidemic onset timing among subtypes and among cities.

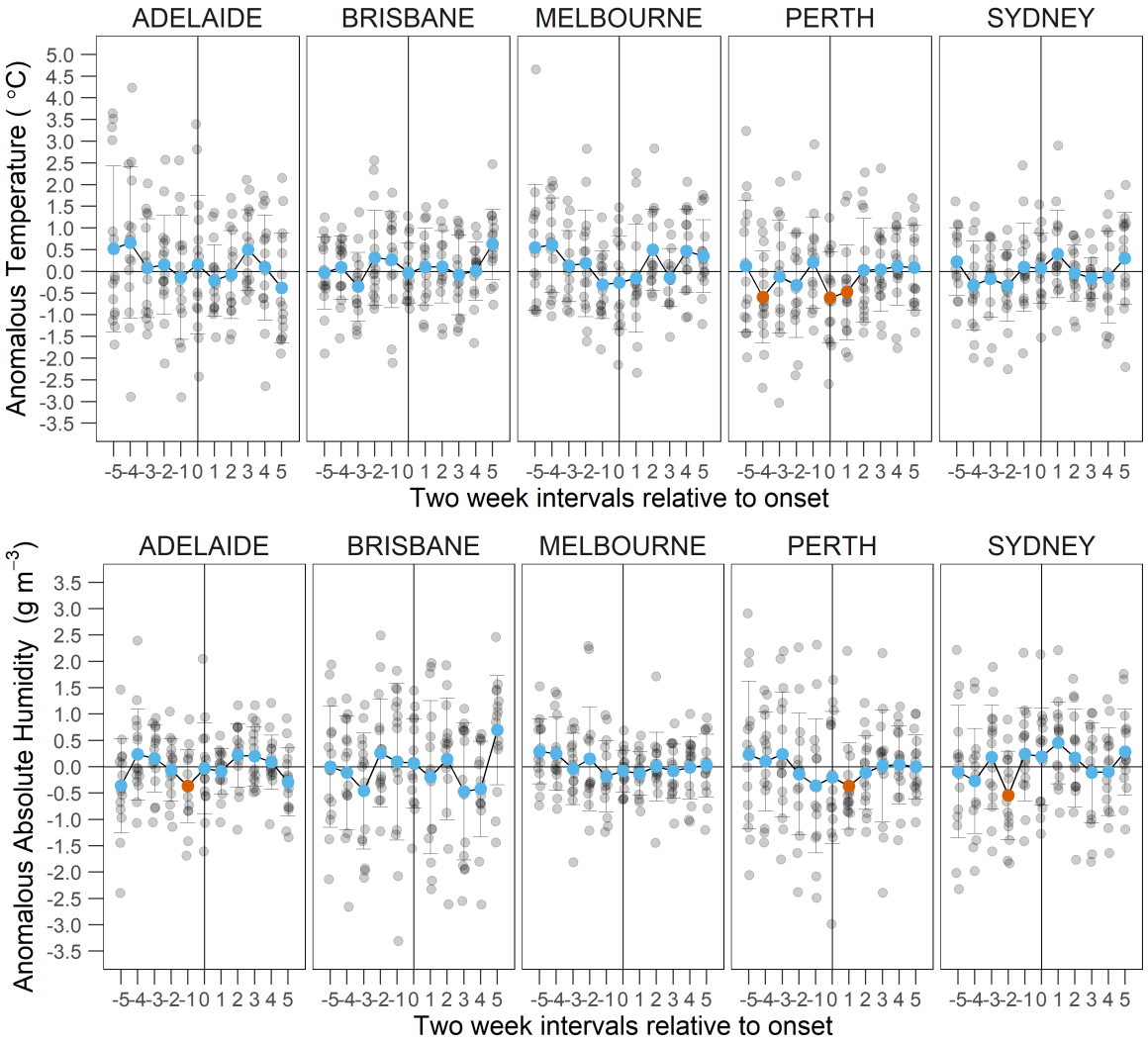


Fig. B.2 Climatic conditions around epidemic onset. (a) Anomalous temperature T' and (b) absolute humidity AH' prior to and after epidemic onset for each of all five cities. Epidemic onset is marked by the vertical line at 0. For the earliest onset epidemic in each season and city (15 epidemics per city), T' and AH' for each time point are represented by grey points: a point below the horizontal line denotes that the value is lower than the 31 year city-specific mean. Blue points show the mean T' and AH' for that two week period for all epidemics within the study period in a particular city. Time periods with statistically significantly ($p<0.05$) reductions in mean T' or AH' from the 31-year average are shown in orange. In the two-week period immediately prior to epidemic onset, there is a statistically significant reduction in AH' of $0.366gm^{-3}$ in Adelaide ($p = 0.021$, Wilcoxon one-sample test), which is roughly equivalent to a 3.11% reduction in relative humidity. This result was not statistically significant after correcting for multiple testing (Holm correction).

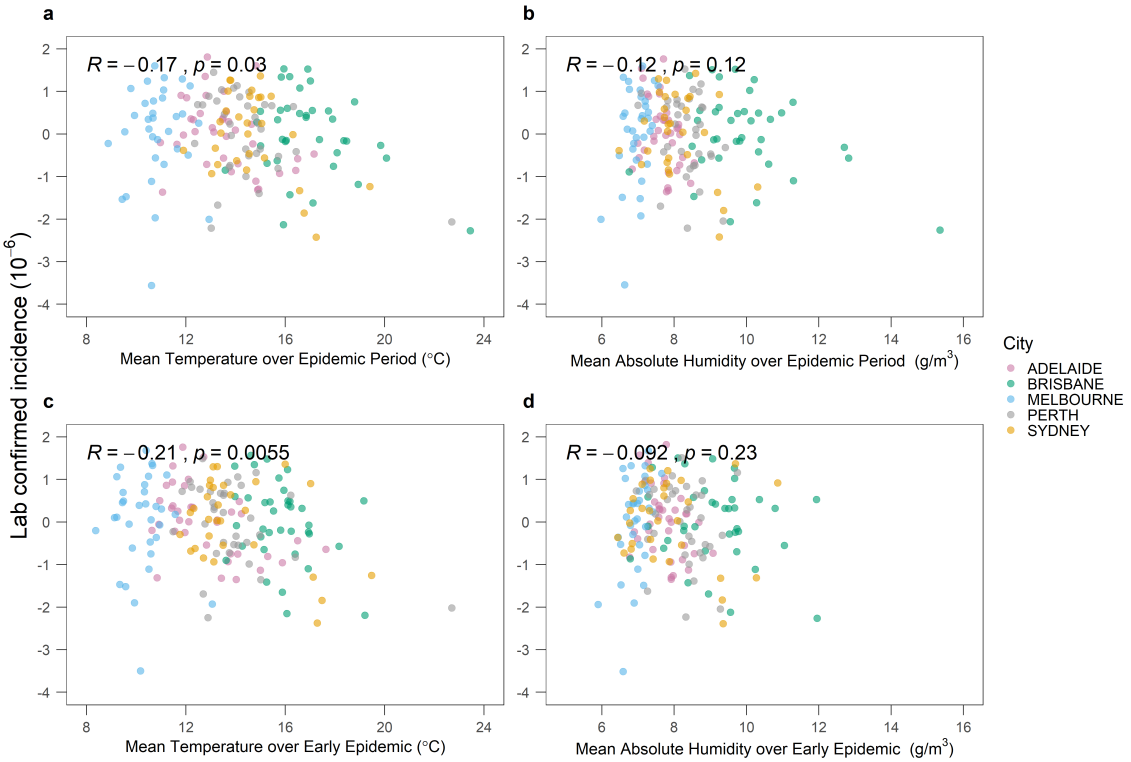


Fig. B.3 Effect of climatic factors on epidemic incidence. The relationship between epidemic incidence and (a,c) the mean temperature and (b,d) absolute humidity values over the entire epidemic period (a,b) and the early epidemic (defined as period from the onset to the peak of an epidemic; Panels c,d). Incidence for individual epidemics were log transformed and subtracted by the city- and subtype-specific mean of log incidence, to allow for comparison across cities and subtypes. The distribution of mean climatic values displays segregation by city, reflecting underlying differences between climatic regions. Whilst there were weak negative associations between the size of an epidemic and mean temperature over the entire epidemic period or just the period from epidemic onset to peak, the high variability in epidemic sizes observed over a large range of climatic conditions and lack of overall, as well as within-city, trends suggest that climatic factors have limited and noisy effects on epidemic size.

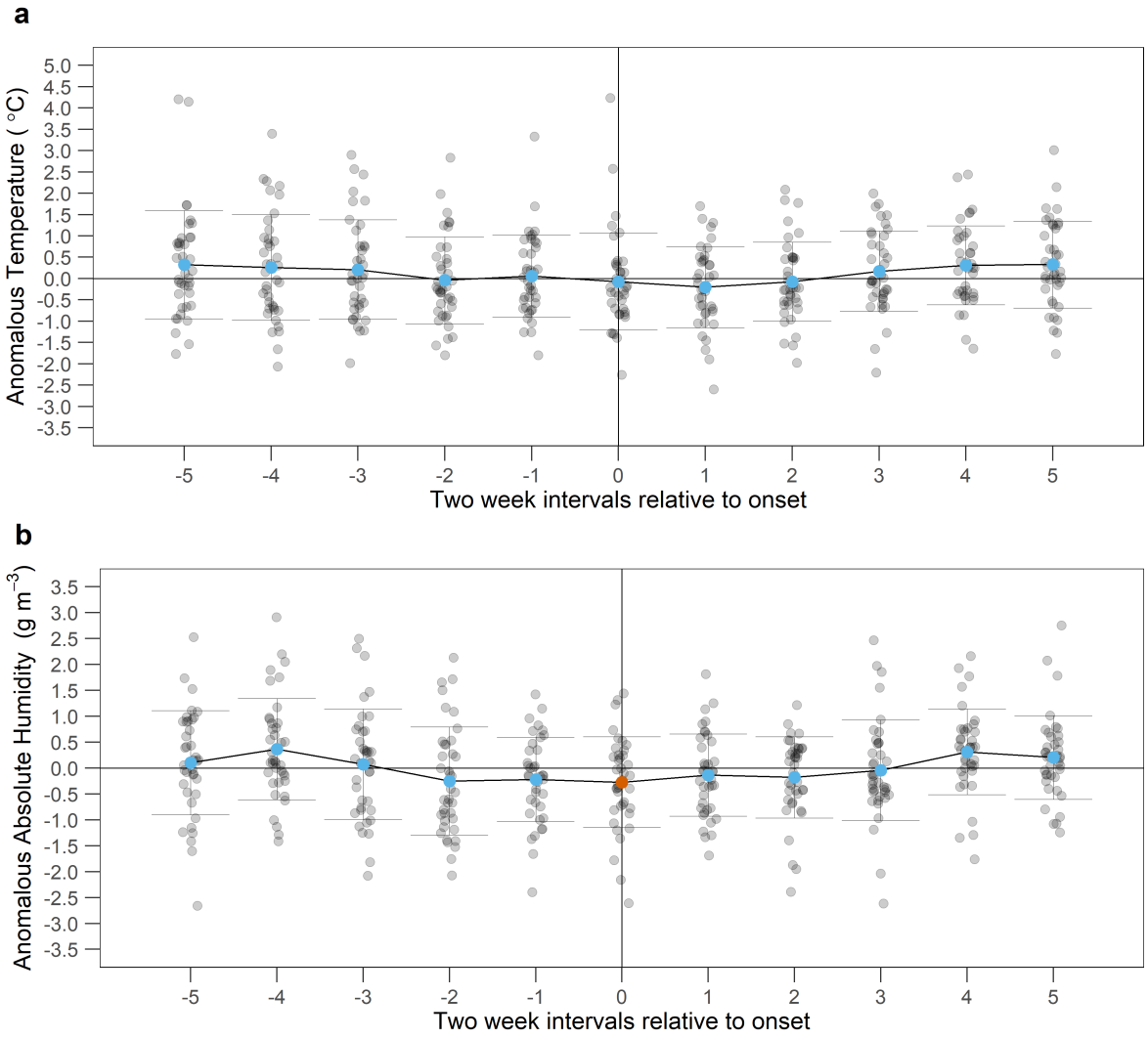


Fig. B.4 Robustness of climatic factor analyses. Utilising only the timing estimates by Geoghegan et al.[92], I assessed if more generally, the onset of influenza A epidemic activity in the seasons from 2007 to 2015 was preceded by periods of anomalous climatic conditions; see Appendix B.1. (a) Anomalous temperature T' and (b) absolute humidity AH' prior to and after epidemic onset across all five cities. Epidemic onset is marked by the vertical line at 0. For the earliest onset epidemic in each season and city (8 years x 5 cities = 40 epidemics), T' and AH' for each time point are represented by grey points: a point below the horizontal line denotes that the value is lower than the 31 year city-specific mean. Blue points show the mean T' and AH' for that two week period for all epidemics within the study period. Time periods with statistically significantly ($p < 0.05$; Wilcoxon one-sample test) reductions in mean T' or AH' from the 31-year average are shown in orange.

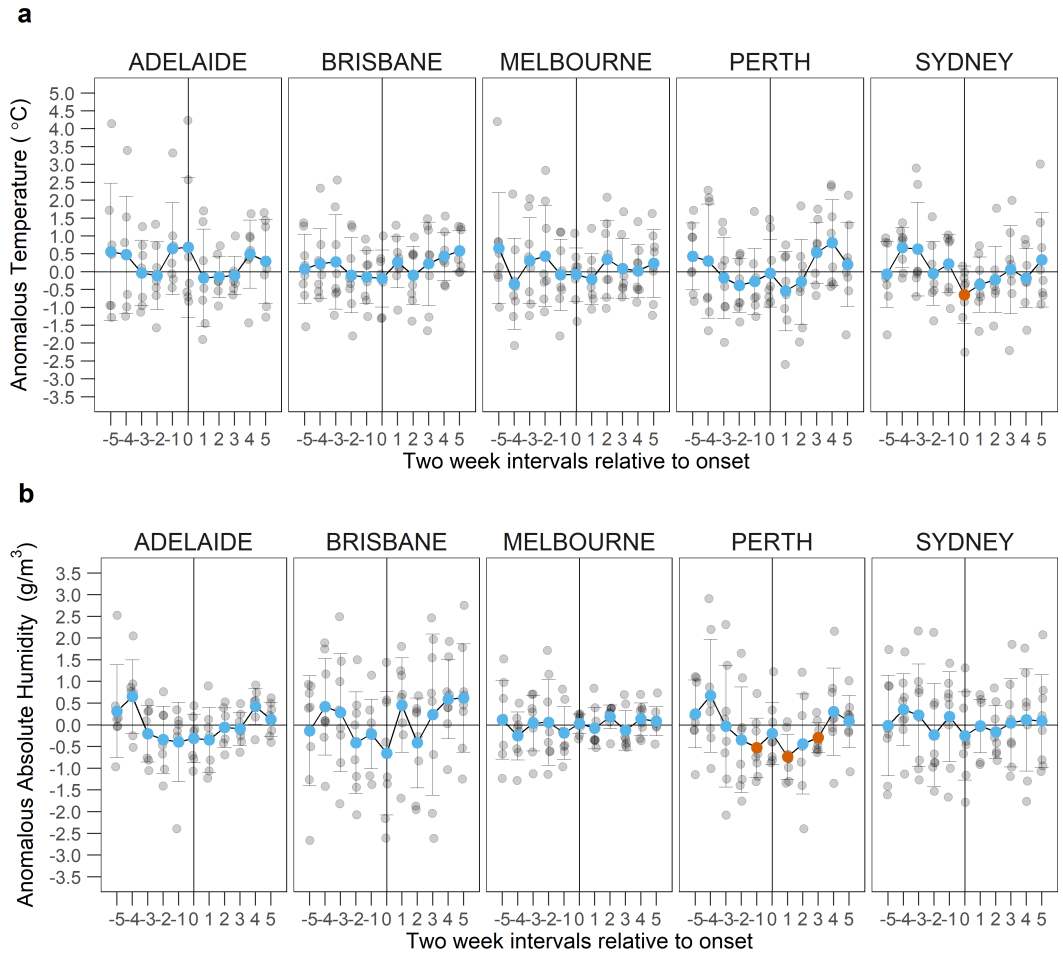


Fig. B.5 Robustness of climatic factor analyses. Utilising only the timing estimates by Geoghegan et al.[92], I assessed if more generally, the onset of influenza A epidemic activity in the seasons from 2007 to 2015 was preceded by periods of anomalous climatic conditions; see Appendix B.1. (a) Anomalous temperature T' and (b) absolute humidity AH' prior to and after epidemic onset for each of all five cities. Epidemic onset is marked by the vertical line at 0. For the earliest onset epidemic in each season and city (15 epidemics per city), T' and AH' for each time point are represented by grey points: a point below the horizontal line denotes that the value is lower than the 31 year city-specific mean. Blue points show the mean T' and AH' for that two week period for all epidemics within the study period in a particular city. Time periods with statistically significantly (p<0.05) reductions in mean T' or AH' from the 31-year average are shown in orange. Time periods with statistically significantly ($p < 0.05$) reductions in mean T' or AH' from the 31-year average are shown in orange. In the two-week period immediately prior to epidemic onset, there is a statistically reduction in AH' of $0.522gm^{-3}$ in Perth ($p = 0.039$, Wilcoxon one-sample test), which is roughly equivalent to a 3.13% reduction in relative humidity. This result was not statistically significant after correcting for multiple testing (Holm correction).

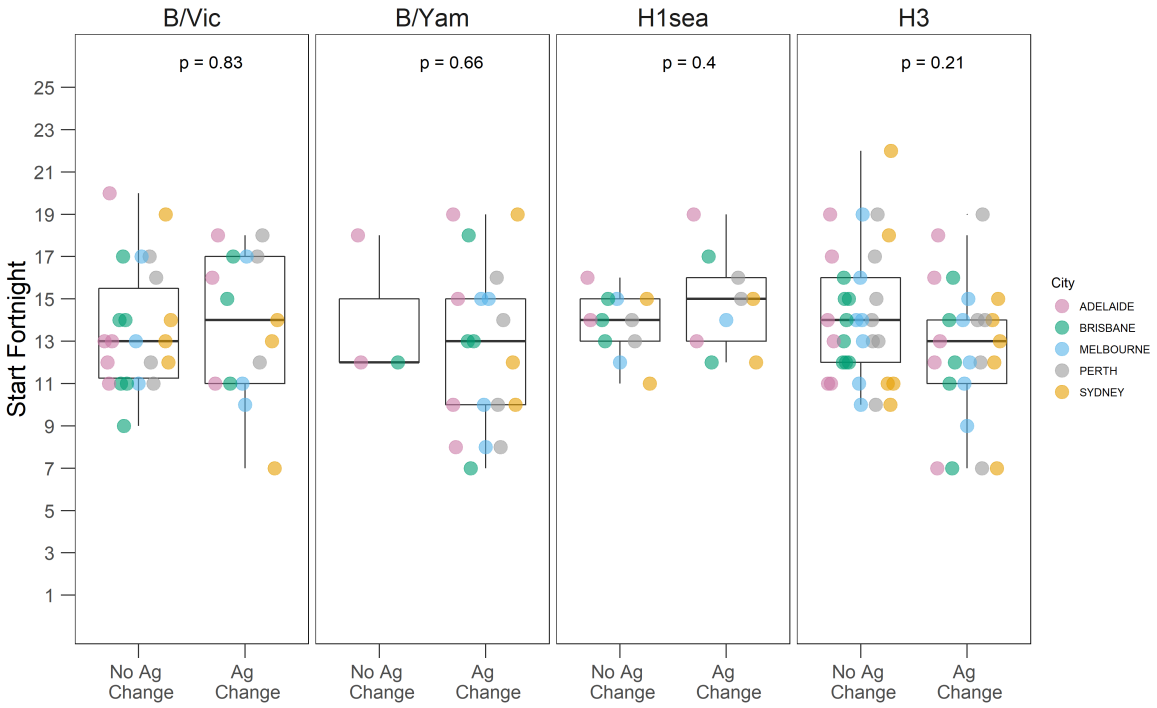


Fig. B.6 Effect of antigenic change on epidemic onset timing. Epidemic onset timings were compared between seasons associated with and without the epidemic level circulation of a new major antigenic variant. *p* values are from Wilcoxon two sample tests (n = 37, 26, 22 and 63 for B/Vic, B/Yam, A/H1sea and A/H3 respectively). Box plots show the median, first and third quartile values, as well as overall range. Each point corresponds to one epidemic in a city and the box plots show the median, first and third quartile values, and range.

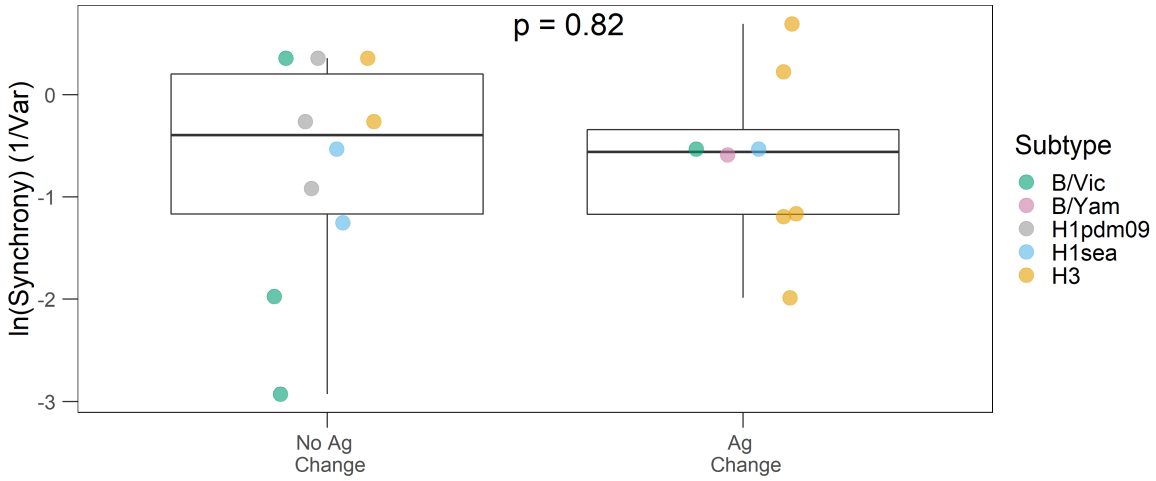


Fig. B.7 Effect of antigenic change on the spatio-temporal synchrony of epidemics. Epidemic synchrony was compared between seasons associated with and without the epidemic level circulation of a new major antigenic variant. Synchrony is quantified as the reciprocal of the variance in onset timings for seasons, where epidemic activity for an antigenic variant was present in all five cities. *p* values are from Wilcoxon two sample tests (n = 18). Box plots show the median, first and third quartile values, as well as overall range.

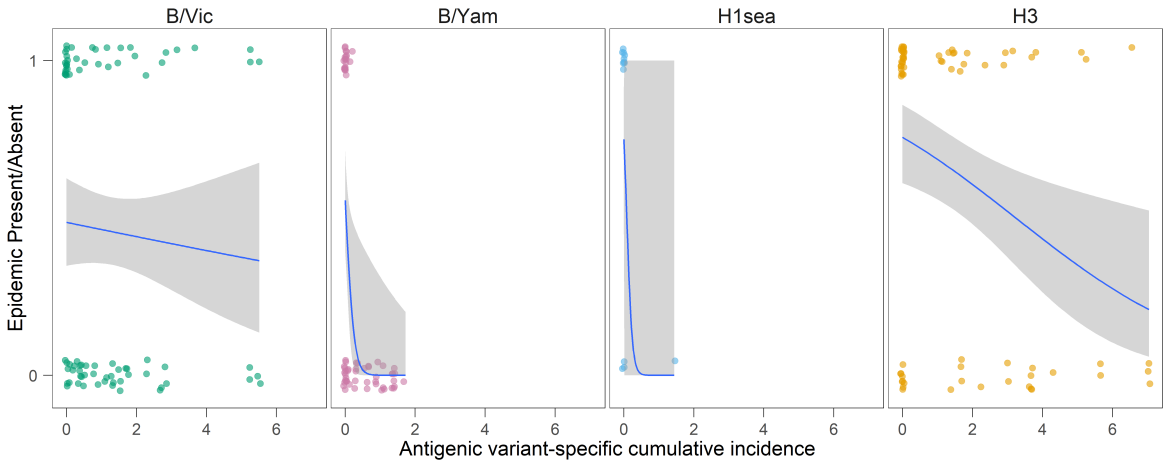


Fig. B.8 Effect of prior immunity on the probability of successful epidemic initiation. For each antigenic variant, I examined whether epidemic levels of activity were present or absent in each of the seasons from its initial detection to its replacement by the next variant. . Antigenic variant-specific cumulative incidence was measured relative to the city-specific mean epidemic size, where 1 is equivalent to the mean epidemic incidence. Binary logistic regression models were fitted for each subtype (n = 81, 65, 13 and 72 for B/Vic, B/Yam, A/H1sea and A/H3 respectively). The 95% confidence interval is denoted by the grey shaded area. See Table B.5 for *OR* from the binary logistic regressions.

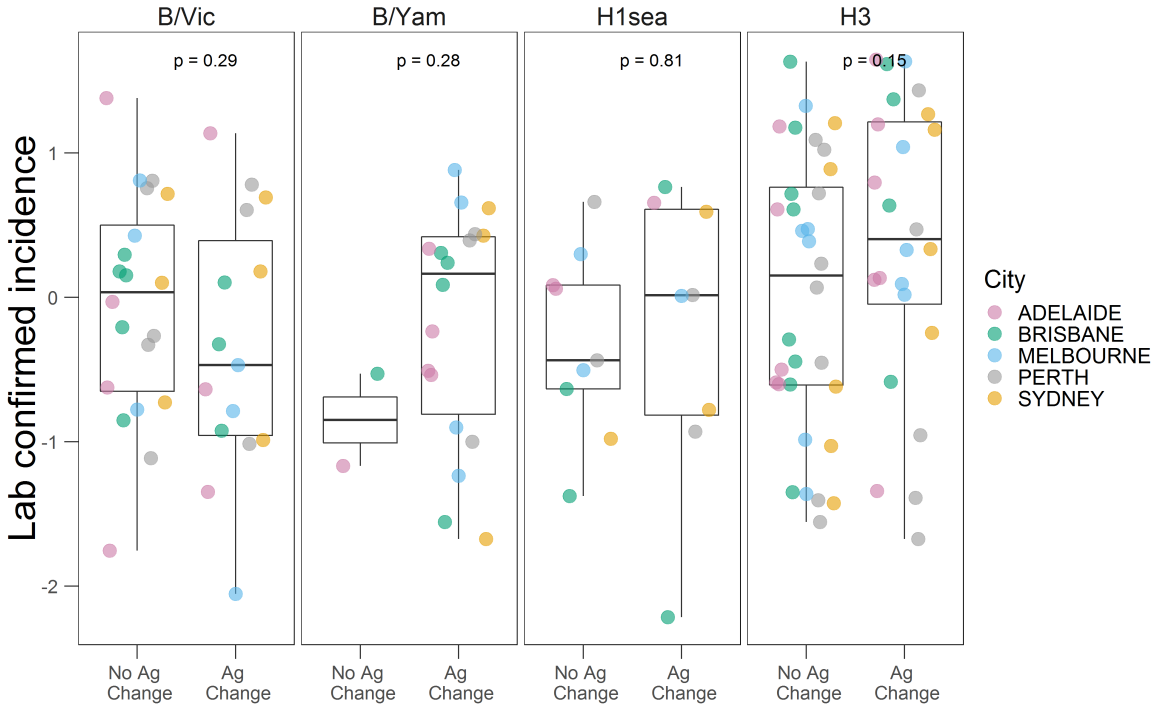


Fig. B.9 Robustness of analyses of the effect of antigenic change on epidemic incidence, towards potential antigenic characterisation errors. Analyses as part of main text (Figure 3) attribute all A/H3 cases in 2004 to California/7/2004 and in 2005 to A/Wisconsin/67/2005 antigenic variants, due to delays in updating vaccine strain nomenclature. Here, I make no such assumptions. Epidemic incidence were compared between seasons associated with and without the epidemic level circulation of a new major antigenic variant. Within each subtype, incidence for individual epidemics were log transformed and subtracted by the city-specific mean of log incidence, to allow for comparison between cities. p values are from Wilcoxon two sample tests (n = 37, 26, 22 and 63 for B/Vic, B/Yam, A/H1sea and A/H3 respectively). Each point corresponds to one epidemic in a city and the box plots show the median, first and third quartile of the transformed values, and range.

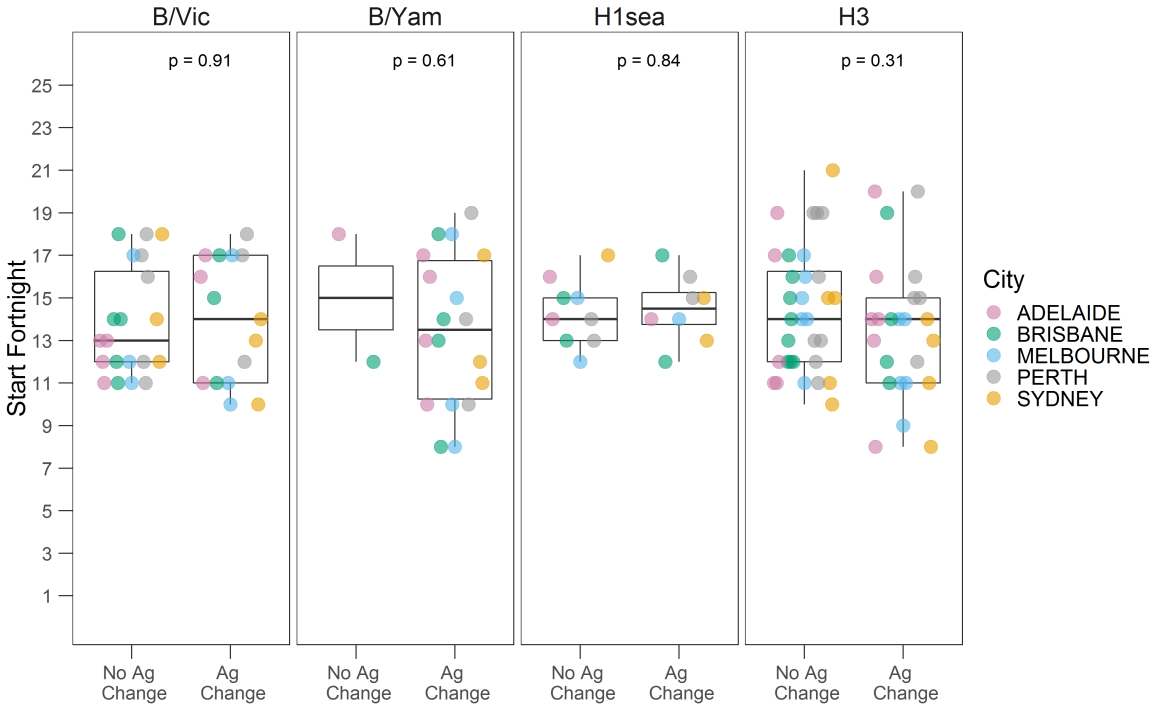


Fig. B.10 Robustness of analyses of the effect of antigenic change on epidemic onset timing, towards potential antigenic characterisation errors. Analyses in Fig. B.6 attribute all A/H3 cases in 2004 to California/7/2004 and in 2005 to A/Wisconsin/67/2005 antigenic variants, due to delays in updating vaccine strain nomenclature. Here, I make no such assumptions. Epidemic onset timing was compared between seasons associated with and without the epidemic level circulation of a new major antigenic variant. *p* values are from Wilcoxon two sample tests (n = 37, 26, 22 and 63 for B/Vic, B/Yam, A/H1sea and A/H3 respectively). Box plots show the median, first and third quartile values, as well as overall range. Each point corresponds to one epidemic in a city and the box plots show the median, first and third quartile values, and range.

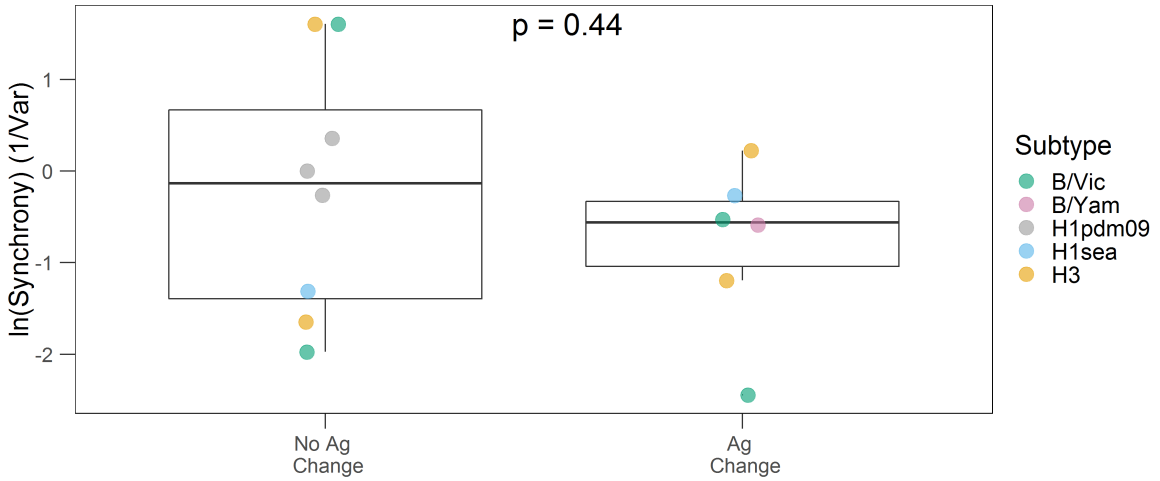


Fig. B.11 Robustness of analyses of the effect of antigenic change on the spatio-temporal synchrony of epidemics, towards potential antigenic characterisation errors. Analyses as part of main text (Fig. B.7) attribute all A/H3 cases in 2004 to California/7/2004 and in 2005 to A/Wisconsin/67/2005 antigenic variants, due to delays in updating vaccine strain nomenclature. Here, I make no such assumptions. The spatio-temporal synchrony of epidemics is compared between seasons associated with and without the epidemic level circulation of a new major antigenic variant. Synchrony is quantified as the reciprocal of the variance in onset timings for seasons, where epidemic activity for an antigenic variant was present in all five cities. *p* values are from Wilcoxon two sample tests (n = 14). Box plots show the median, first and third quartile values, as well as overall range.

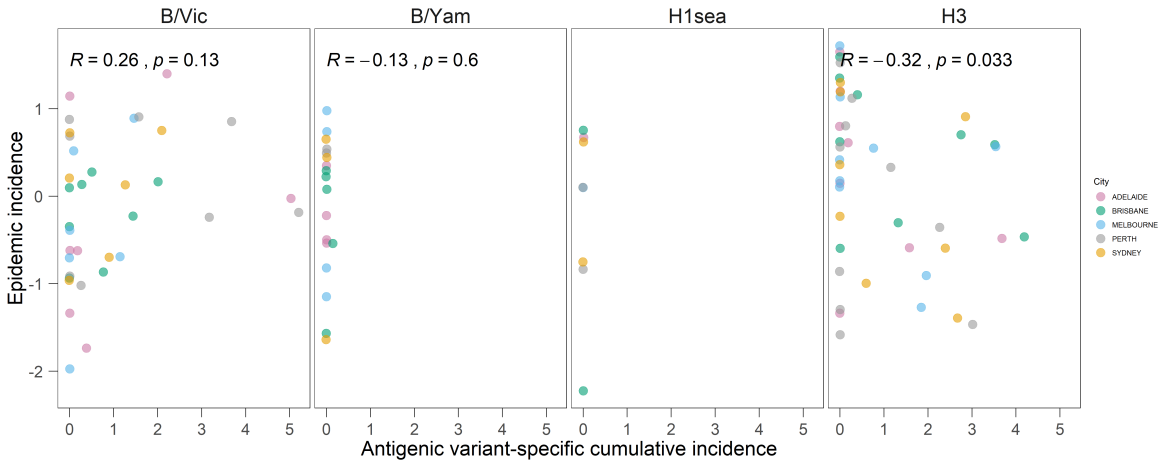


Fig. B.12 Robustness of analyses of the effect of antigenic variant-specific cumulative incidence on subsequent epidemic incidence, towards potential antigenic characterisation errors. Analyses as part of main text (Figure 4) attribute all A/H3 cases in 2004 to California/7/2004 and in 2005 to A/Wisconsin/67/2005 antigenic variants, due to delays in updating vaccine strain nomenclature. Here, I make no such assumptions. Within each subtype, incidence for individual epidemics were log transformed and subtracted by the city-specific mean of log incidence, to allow for comparison between cities. Cumulative incidence was measured relative to the city-specific mean epidemic size, where 1 is equivalent to the mean epidemic incidence. r and p values are from Pearson's correlation tests (n = 37, 20, 9 and 45 for B/Vic, B/Yam, A/H1sea and A/H3 respectively). Note that antigenic variants of B/Yam and H1sea rarely initiated multiple epidemics during the study period.

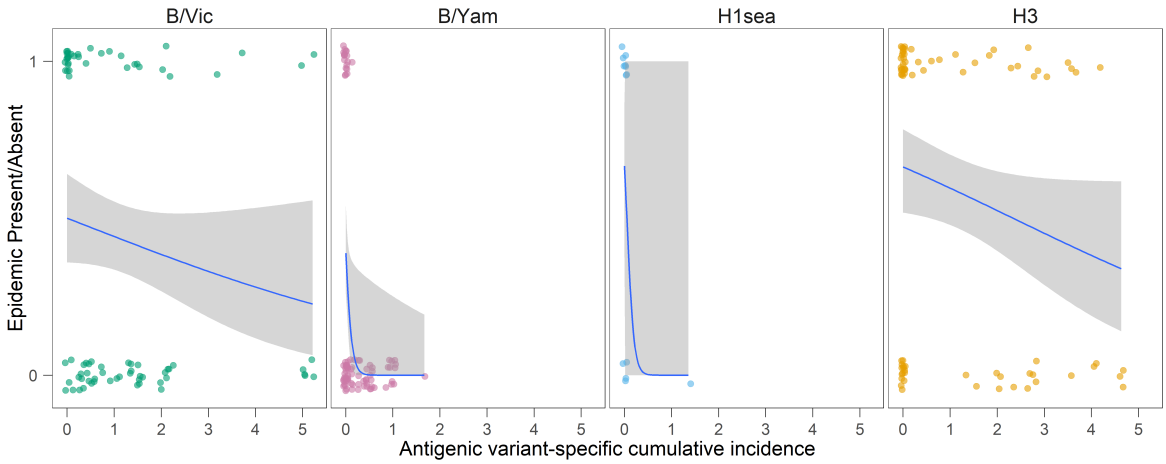


Fig. B.13 Robustness of analyses of the effect of antigenic variant-specific cumulative incidence on probability of successful epidemic initiation, towards potential antigenic characterisation errors. Analyses as part of main text (Fig. B.8) attribute all A/H3 cases in 2004 to California/7/2004 and in 2005 to A/Wisconsin/67/2005 antigenic variants, due to delays in updating vaccine strain nomenclature. Here, I make no such assumptions. For each antigenic variant, I examined whether epidemic levels of activity were present or absent in each of the seasons from its initial detection to its replacement by the next variant. . Antigenic variant-specific cumulative incidence was measured relative to the city-specific mean epidemic size, where 1 is equivalent to the mean epidemic incidence. Binary logistic regression models were fitted for each subtype (n = 81, 65, 13 and 72 for B/Vic, B/Yam, A/H1sea and A/H3 respectively). The 95% confidence interval is denoted by the grey shaded area. See Table B.6 for *OR* from the binary logistic regressions.

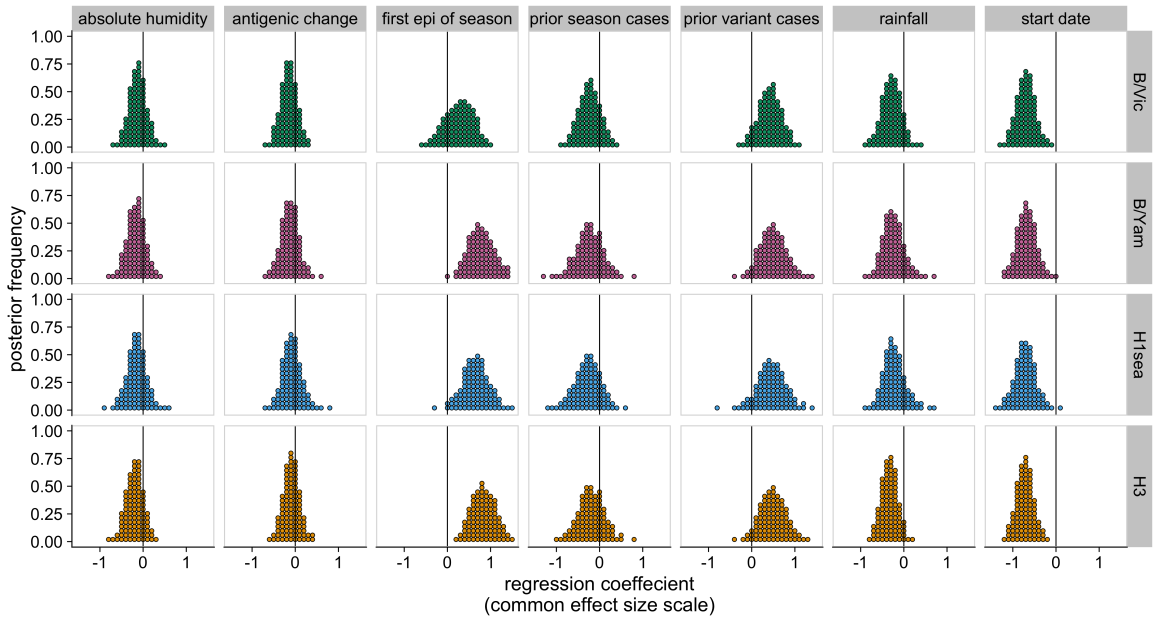


Fig. B.14 Joint contributions of climatic and virological factors on epidemic incidence for individual subtypes. Using the Bayesian multilevel model, posterior distributions for the effects of climate, timing, and antigenic variables on epidemic size were estimated for individual subtypes. Both predictors and outcome variables are standardized, so effects are shown on a common scale.

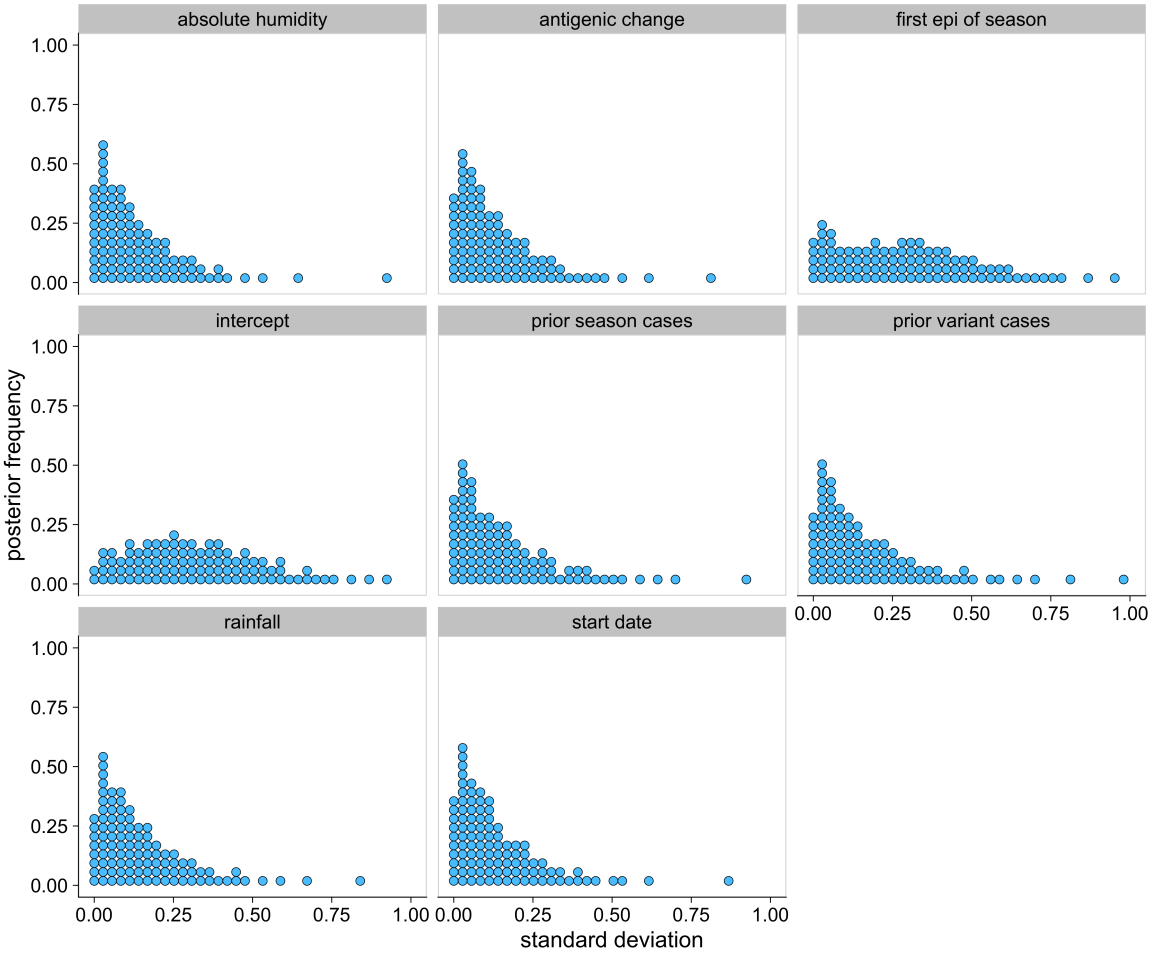


Fig. B.15 Posterior distributions of standard deviations for subtype-specific effect sizes about the mean across all subtypes. Model estimates place standard deviations close to zero, suggesting that there is limited evidence from this dataset for variation among subtypes in the effects of climate, timing, or antigenicity on epidemic size.

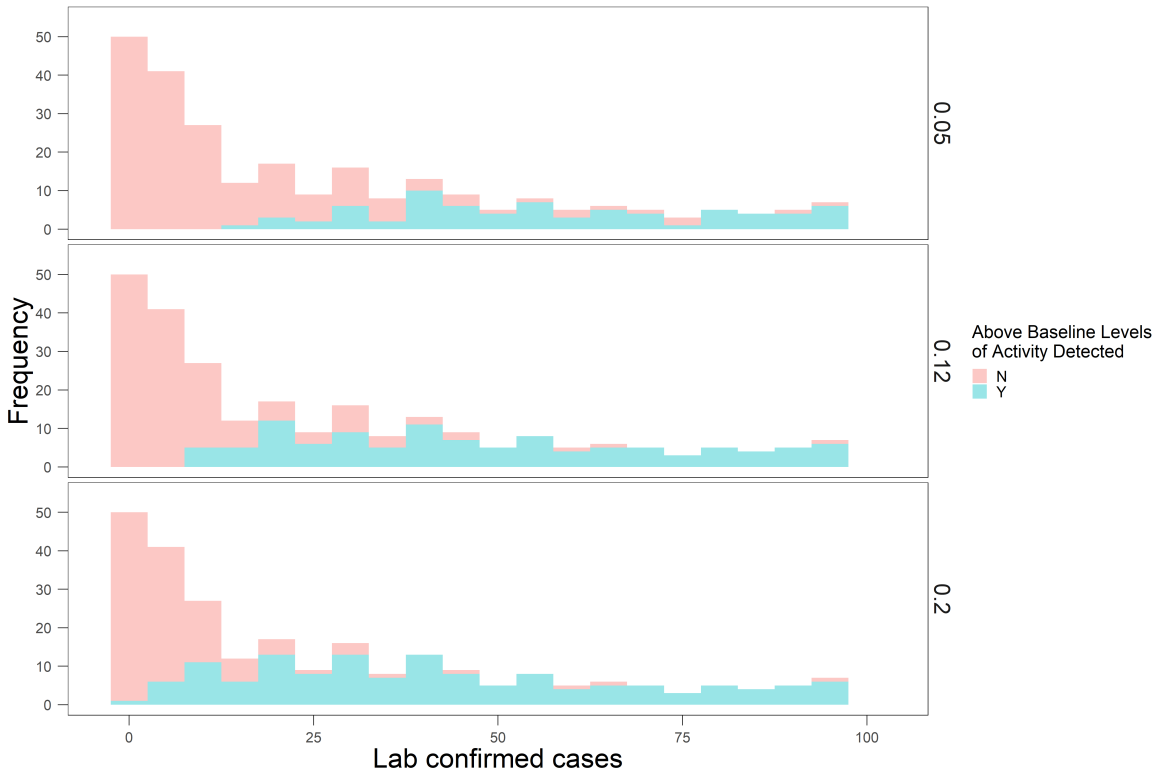


Fig. B.16 Histogram of the number of lab confirmed cases for seasons with and without above baseline levels of activity. In order to focus on the sensitivity and specificity of the algorithm in identifying epidemics, for case-scarce seasons, only seasons with less than 100 cases were plotted.

The threshold value y_α, which if exceeded marks the onset of an epidemic, and thus the sensitivity of the detection algorithm are determined by the quantile parameter α. In the main text, $\alpha = 0.12$ was used, since it identified epidemic onset and end timings that corresponded well with visual inspection of the raw time series. I repeated the estimation of onset and end timings using $\alpha = 0.05\,\&\,0.2$ (plotted in separate facets), which increased and reduced threshold values respectively. When the threshold values are lowered ($\alpha = 0.2$), there is a shift in epidemic size distribution to the left: non-epidemic activity early in the calendar year were conflated as periods of above-baseline levels of epidemic activity.

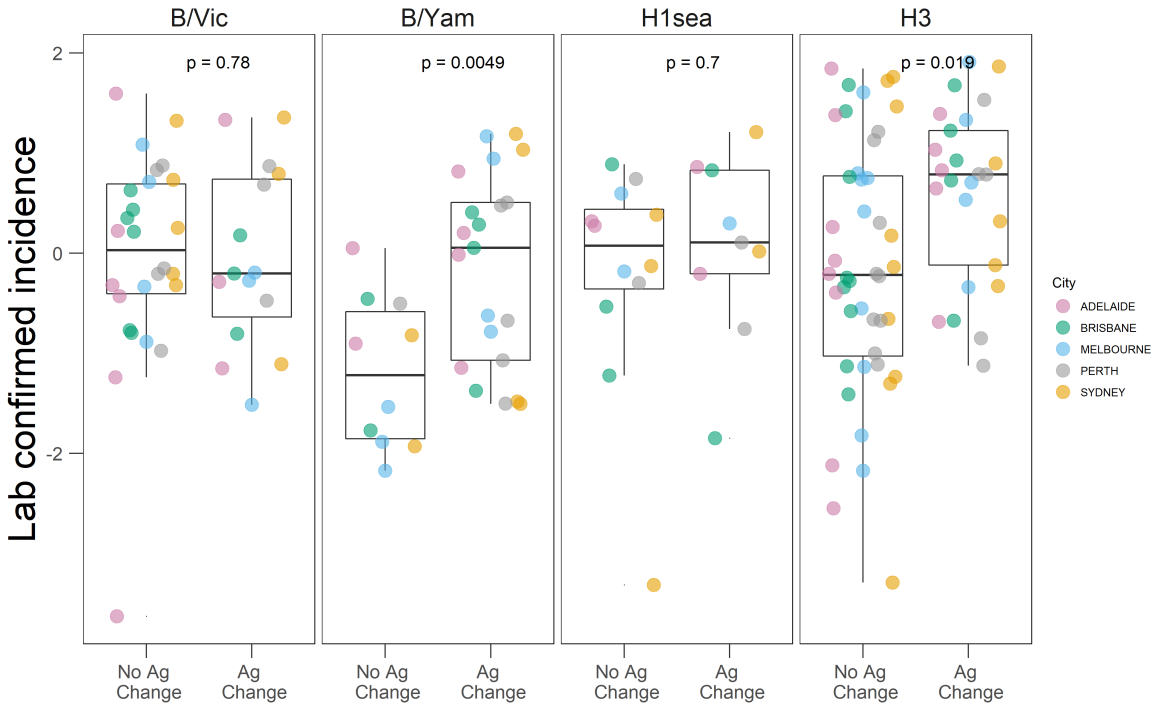


Fig. B.17 Robustness of analyses of the effect of antigenic change on epidemic incidence towards lowering of epidemic detection threshold values ($\alpha = 0.2$). Reducing the threshold resulted in spurious activity being designated as epidemics, which in turn inflated the number of small epidemics observed and the apparent association between antigenic change and larger epidemics.

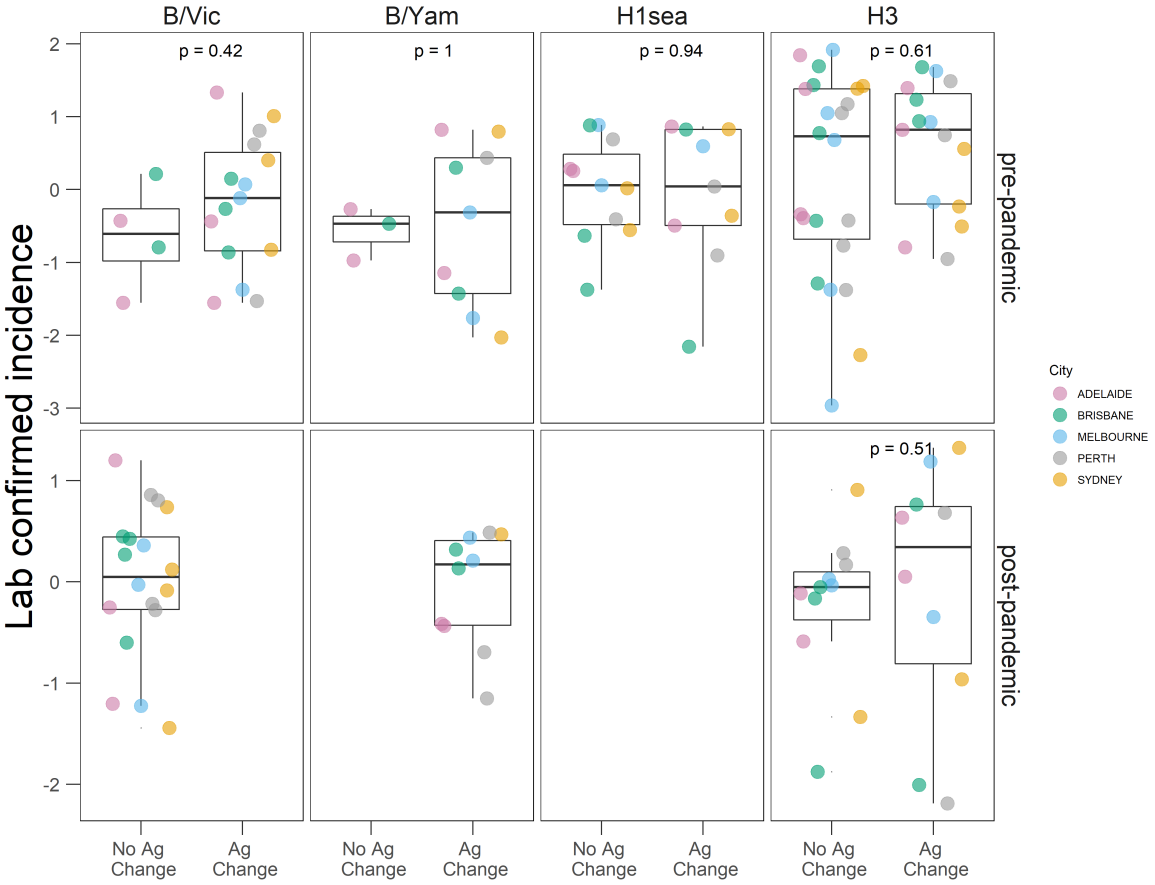


Fig. B.18 Robustness of analyses of the effect of antigenic change on epidemic incidence towards potential differences in surveillance intensity between the pre- and post-pandemic eras. Within each subtype, incidence for individual epidemics were log transformed and subtracted by the city- and era-specific mean of log incidence, to allow for comparison between cities. *p* values are from Wilcoxon two sample tests. Each point corresponds to one epidemic in a city and the box plots show the median, first and third quartile of the transformed values, and range.

B.4 Appendix B Tables

n-week block	Observed Mean T' (°C)	Observed Mean AH' (gm^{-3})
1	0.0294 (0.556)	-0.113 (0.304)
2	0.0158 (0.544)	-0.089 (0.304)
3	-0.0183 (0.471)	-0.0538 (0.356)

Table B.1 Observed mean climatic values over n-week continuous blocks prior to the first epidemic onset of each season. Mean T' and AH' fluctuations were aggregated across all five cities and compared against the bootstrapped distribution of random samples of observed mean T' and AH' values to determine statistical significance. The associated p values (the observed value's quantile within the bootstrap distribution) are shown in parentheses.

City	Observed Mean T' (°C)	Observed Mean AH' (gm^{-3})	n-week block
Adelaide	-0.134 (0.345)	**-0.366 (0.021)**	2
Brisbane	0.272 (0.875)	0.0967 (0.627)	2
Melbourne	-0.302 (0.154)	-0.179 (0.123)	2
Perth	0.211 (0.765)	-0.359 (0.0708)	2
Sydney	0.101 (0.675)	0.241 (0.818)	2
Adelaide	0.00971 (0.542)	-0.215 (0.0695)	4
Brisbane	0.294 (0.932)	0.18 (0.766)	4
Melbourne	-0.0572 (0.407)	-0.0117 (0.44)	4
Perth	-0.0538 (0.425)	-0.249 (0.0945)	4
Sydney	-0.113 (0.29)	-0.149 (0.236)	4
Adelaide	0.0322 (0.591)	-0.0913 (0.267)	6
Brisbane	0.0791 (0.68)	-0.0324 (0.444)	6
Melbourne	0.00498 (0.514)	-0.0201 (0.405)	6
Perth	-0.0745 (0.374)	-0.087 (0.273)	6
Sydney	-0.133 (0.217)	-0.0382 (0.406)	6

Table B.2 Observed mean climatic values over n-week continuous blocks prior to the first epidemic onset of each season. For each city, mean T' and AH' fluctuations compared against the bootstrapped distribution of random samples of observed mean T' and AH' values in order to determine statistical significance. The associated p values (the observed value's quantile within the bootstrap distribution) are shown in parentheses. Note, the result for Adelaide (bolded) were not significant after correcting for multiple testing (Holm correction).

n-week block	Observed Mean T' (°C)	Observed Mean AH' (gm^{-3})
2	0.0588 (0.574)	-0.219 (0.237)
4	0.0098 (0.524)	-0.235 (0.169)
6	0.0759 (0.629)	-0.133 (0.26)

Table B.3 Utilising only the timing estimates by Geoghegan et al.[92], I assessed if more generally, the onset of influenza A epidemic activity in the seasons from 2007 to 2015 was preceded by periods of anomalous climatic conditions; see Appendix B.1. Observed mean climatic values over n-week continuous blocks prior to the first epidemic onset of each season. Mean T' and AH' fluctuations were aggregated across all five cities and compared against the bootstrapped distribution of random samples of observed mean T' and AH' values, in order to calculate the statistical significance non-parametrically: the associated p values (the observed value's quantile within the bootstrap distribution) are shown in parentheses.

City	Observed Mean T' (°C)	Observed Mean AH' (gm^{-3})	n-week block
Adelaide	0.651 (0.948)	-0.394 (0.054)	2
Brisbane	-0.146 (0.329)	-0.208 (0.323)	2
Melbourne	-0.085 (0.425)	-0.183 (0.195)	2
Perth	-0.272 (0.267)	-0.522 (0.063)	2
Sydney	0.219 (0.754)	0.19 (0.697)	2
Adelaide	0.27 (0.826)	**-0.367 (0.029)**	4
Brisbane	-0.122 (0.328)	-0.31 (0.191)	4
Melbourne	0.176 (0.722)	-0.062 (0.352)	4
Perth	-0.328 (0.167)	-0.434 (0.052)	4
Sydney	0.085 (0.643)	-0.019 (0.468)	4
Adelaide	0.168 (0.756)	**-0.311 (0.041)**	6
Brisbane	0.01 (0.522)	-0.112 (0.353)	6
Melbourne	0.22 (0.8)	-0.027 (0.416)	6
Perth	-0.276 (0.176)	-0.299 (0.098)	6
Sydney	0.268 (0.892)	0.062 (0.601)	6

Table B.4 Utilising only the timing estimates by Geoghegan et al.[92], I assessed if more generally, the onset of influenza A epidemic activity in the seasons from 2007 to 2015 was preceded by periods of anomalous climatic conditions; see Appendix B.1. For each city, mean T' and AH' fluctuations compared against the bootstrapped distribution of random samples of observed mean T' and AH' values, in order to calculate the statistical significance non-parametrically: the associated p values (the observed value's quantile within the bootstrap distribution) are shown in parentheses. In Adelaide, the AH' of -0.367gm^{-3} and -0.311gm^{-3} in the 4- and 6- week blocks (in bold) immediately preceding epidemics are roughly equivalent to decreases in relative humidity of 2.88% and 2.40% respectively. Both results were not statistically significant after correcting for multiple testing (Holm correction).

Subtype	Term	*OR*	*OR* adjusted *SE*	*p* value
B/Vic	(intercept)	0.946	0.276	0.848
B/Vic	Cumulative incidence	0.913	0.134	0.534
B/Yam	(intercept)	1.25	0.433	0.514
B/Yam	Cumulative incidence	0.000213	0.00083	0.0303
H1sea	(intercept)	3	2	0.0994
H1sea	Cumulative incidence	2.41E-06	0.00661	0.996
H3	(intercept)	3.1	1.08	0.00111
H3	Cumulative incidence	0.705	0.0913	0.00692

Table B.5 Binary logistic regression assessing the effect of antigenic variant-specific cumulative incidence on the probability of successful epidemic initiation for each subtype. Note: $OR < 1$ implies that increased cumulative incidence results in a reduction in the probability of successful initiation.

Subtype	Term	*OR*	*OR* adjusted *SE*	*p* value
B/Vic	(intercept)	1	0.296	0.996
B/Vic	Cumulative incidence	0.79	0.136	0.171
B/Yam	(intercept)	0.636	0.196	0.142
B/Yam	Cumulative incidence	2.35E-06	1.49E-05	0.0415
H1sea	(intercept)	2	1.22	0.258
H1sea	Cumulative incidence	1.49E-06	0.00433	0.996
H3	(intercept)	1.97	0.611	0.0284
H3	Cumulative incidence	0.748	0.118	0.0663

Table B.6 Binary logistic regression assessing the effect of antigenic variant-specific cumulative incidence on the probability of successful epidemic initiation for each subtype. Analyses as part of main text (Fig. 4.6) attribute all A/H3 cases in 2004 to California/7/2004 and in 2005 to A/Wisconsin/67/2005 antigenic variants, due to delays in updating vaccine strain nomenclature. Here, I make no such assumptions. Note: $OR < 1$ implies that increased cumulative incidence results in a reduction in the probability of successful initiation.

Appendix C

Appendix to Chapter 5

C.1 Appendix C Figures

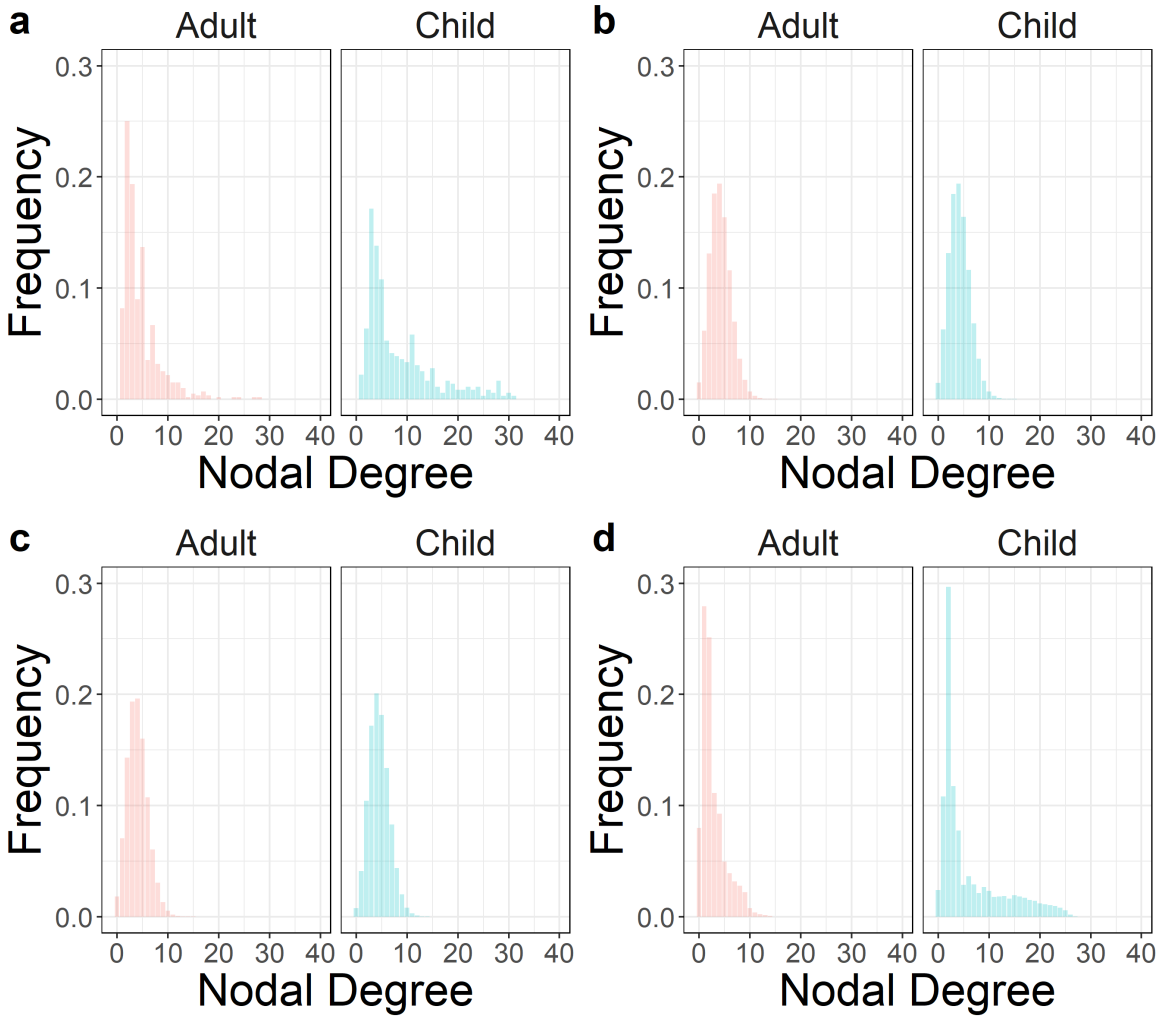


Fig. C.1 Degree distributions for adults and children. (a) POLYMOD[186]; (b) "*No structure*"; (c) "*Household*"; (d) "*Social structure*"

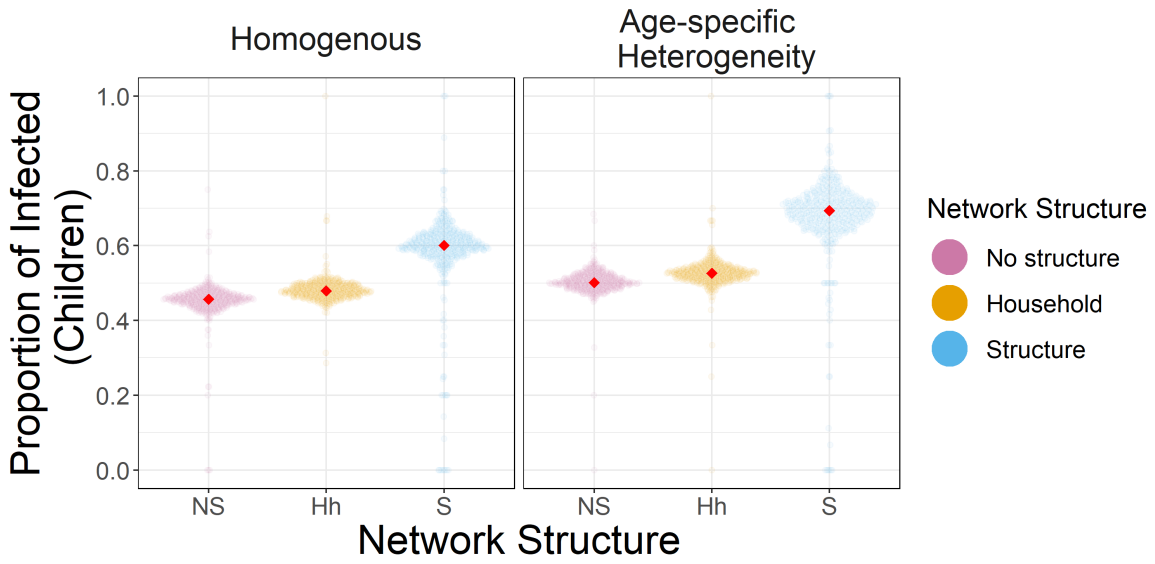


Fig. C.2 Age composition of season 1 epidemics. Individual points represent the proportion of total infected that are children from one stochastic simulation; the red point represents the median value. "*Social structure*" and age-specific distribution of immune classes contribute to the preferential infection of children.

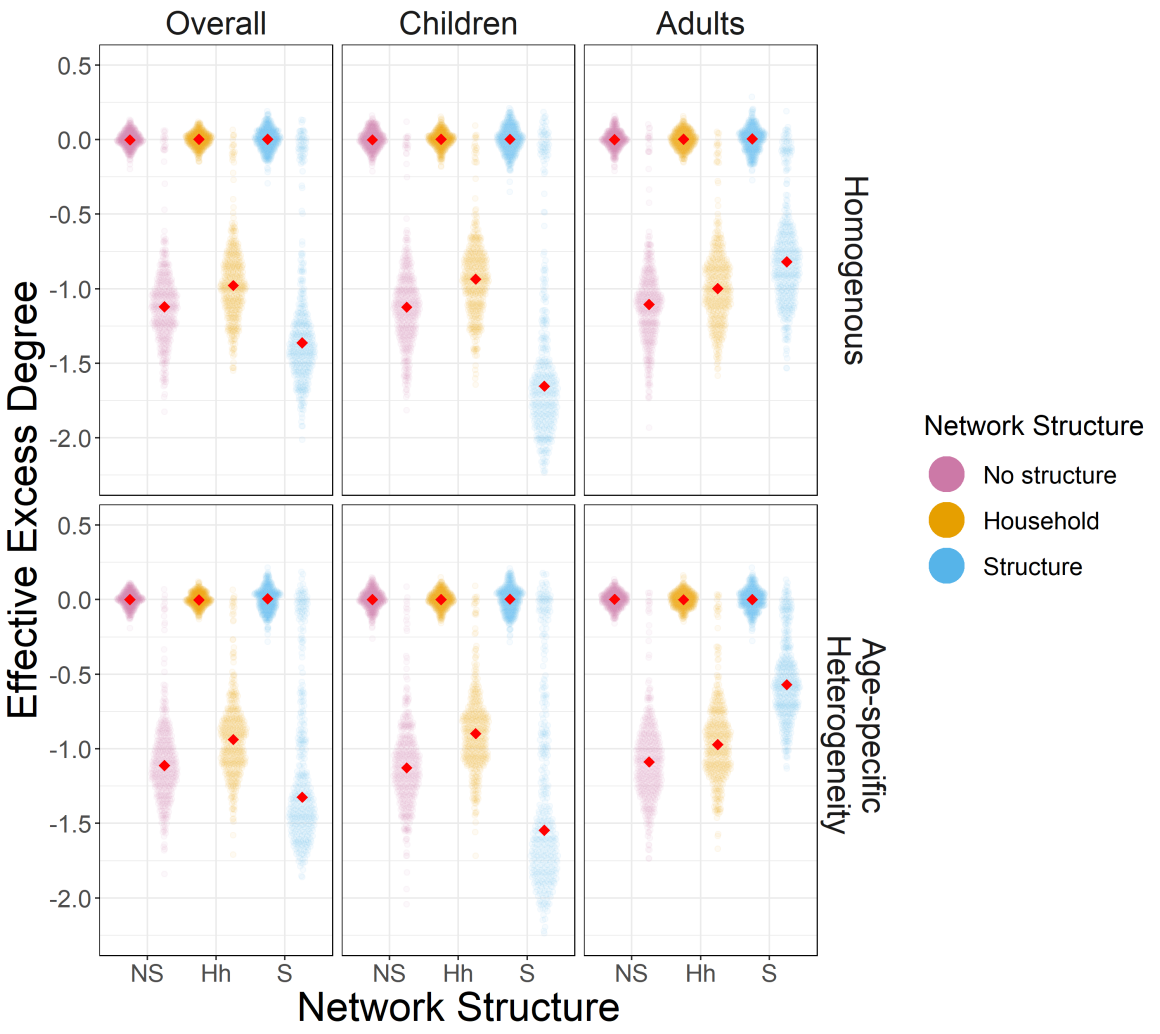


Fig. C.3 The impact of the Season 1 epidemic on network connectivity. For each network structure and immunity combination, the left and right distributions show the average effective excess degree distributions for the original *pre-S1* and residual *post-S1* networks respectively; the red point represents the median value. To make the effective excess degree distributions comparable, for each combination and age group, *pre-S1* and *post-S1* values were normalised relative to the *pre-S1* mean. The reduction on effective excess degree and the frailty of residual *post-S1* networks are maximised through the combined effects of "social structure" and immune class distributions that differ between age groups. This effect is concentrated within the children, who begin as the better-connected age group but consequently experience the greatest reduction in their connectivity.

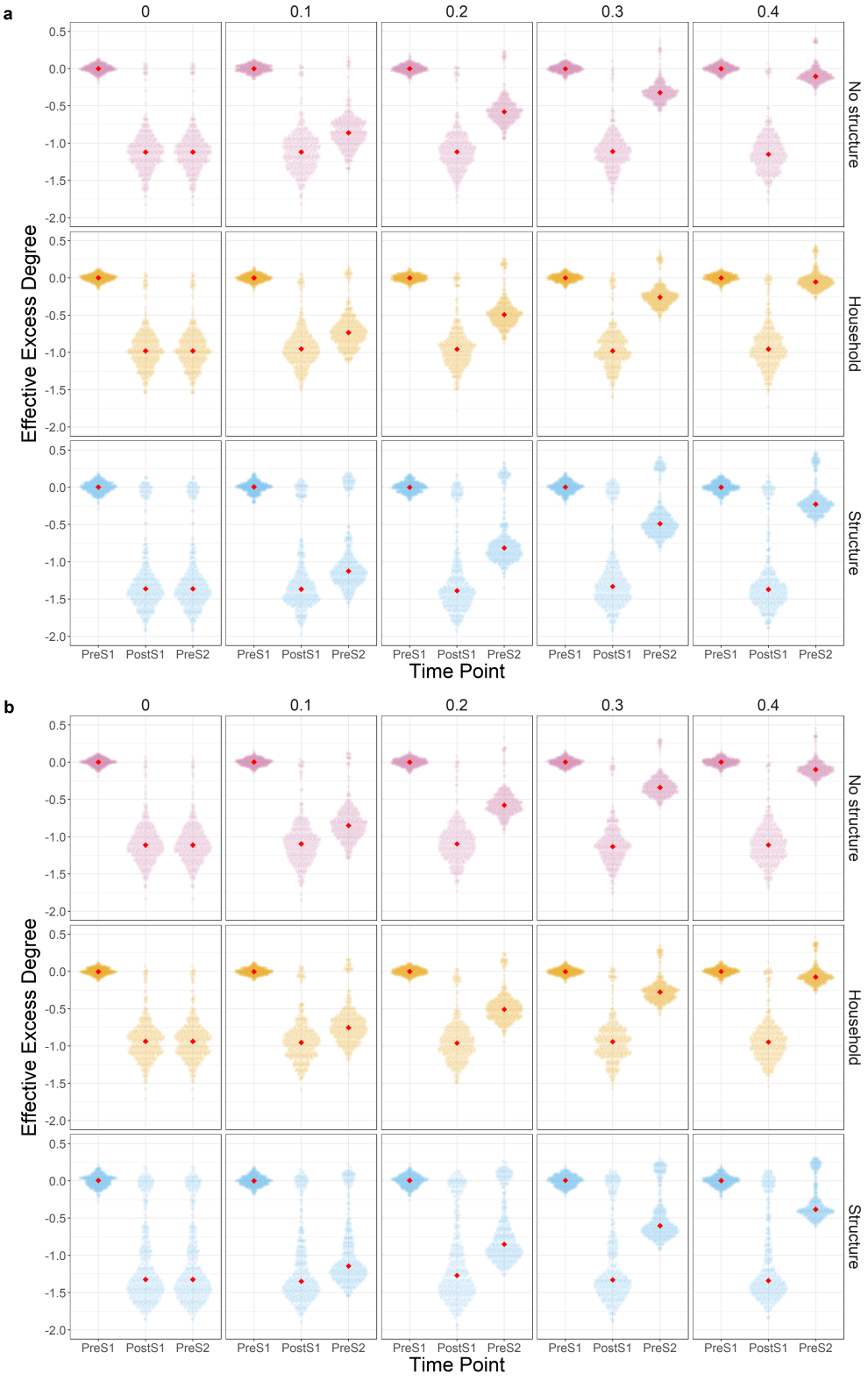


Fig. C.4 The impact of interseasonal immune waning on network connectivity. Immune class distributions are (a) homogenous and (b) differ between age groups. Caption continues on next page.

Fig. C.4 (Previous page.) The proportion of S_0 individuals resusceptibilised (δ) is shown by the top facet label, steadily increasing from left to right. For each network structure and immunity combination, the left, middle and right distributions show the average effective excess degree distributions for the original *pre-S1*, residual *post-S1* and resusceptibilised *pre-S2* networks respectively; the red point represents the median value. To make the effective excess degree distributions comparable, for each combination and age group, *pre-S1*, *post-S1* and *pre-S2* values were normalised relative to the *pre-S1* mean. For both (a) and (b), the *pre-S2* effective degree increases with δ: by $\delta = 0.4$, connectivity is restored to levels comparable to *pre-S1* values, except for the case of "*social structure*". The rate at which effective excess degree is restored is lowest in the presence of both "*social structure*" and age-specific heterogeneities in immune class distributions (bottom row of (b)).

C.2 Appendix C Tables

Network structure	ERGM term	Estimate	Std. Error	z value	Pr(>|z|)
No structure	Edges	-5.637	0.020	-281.108	0
Household	Edges	-6.022	0.024	-248.481	0
	Household	6.000	0.054	110.360	0
Structure	Edges	-9.935	0.236	-42.079	0
	Household	10.915	0.233	46.816	0
	Daycare	9.400	0.244	38.581	0
	School	9.150	0.236	38.690	0
	Work	7.916	0.234	33.791	2.65E-250
	Child-Adult	-0.697	0.092	-7.566	3.86E-14
	Adult (Degree >7)	1.906	0.236	8.089	6.01E-16
	Child (Degree >15)	1.355	0.244	5.557	2.74E-08

Table C.1 Coefficients for ERGM terms. "*No structure*" has no specified terms. "*Household*" structure incorporates assortative mixing between individuals of same household index. "*Social structure*" is parameterised based on the contact distribution from the POLYMOD study [186] with assortative mixing within households, daycare-centres, schools and workplaces. Additional degree-based terms were added to force a positive skew.

Network Structure	Immune Class Distribution	Term	*OR*	*OR* adjusted *SE*	*p* value
No structure	Homogenous	Intercept	0.0333	0.00374	1.38E-201
No structure	Homogenous	Proportion resusceptibilised	798000	351000	5.95E-210
No structure	Age-specific Heterogeneity	Intercept	0.0287	0.00334	3.24E-204
No structure	Age-specific Heterogeneity	Proportion resusceptibilised	1430000	650000	3.11E-213
Household	Homogenous	Intercept	0.0467	0.00483	3.84E-193
Household	Homogenous	Proportion resusceptibilised	414000	173000	6.80E-210
Household	Age-specific Heterogeneity	Intercept	0.0546	0.00543	6.51E-188
Household	Age-specific Heterogeneity	Proportion resusceptibilised	141000	56000	4.30E-197
Structure	Homogenous	Intercept	0.0836	0.0076	7.23E-164
Structure	Homogenous	Proportion resusceptibilised	3200	1090	1.66E-123
Structure	Age-specific Heterogeneity	Intercept	0.128	0.0106	3.38E-136
Structure	Age-specific Heterogeneity	Proportion resusceptibilised	399	126	1.19E-80

Table C.2 Binary logistic regression assessing the effect of immune waning on the probability of successful S2 epidemics. Note that $OR > 1$ implies that greater interseasonal waning results in a greater probability of successful epidemic initiation.

Appendix D

Appendix to Chapter 6

D.1 Appendix D Figures

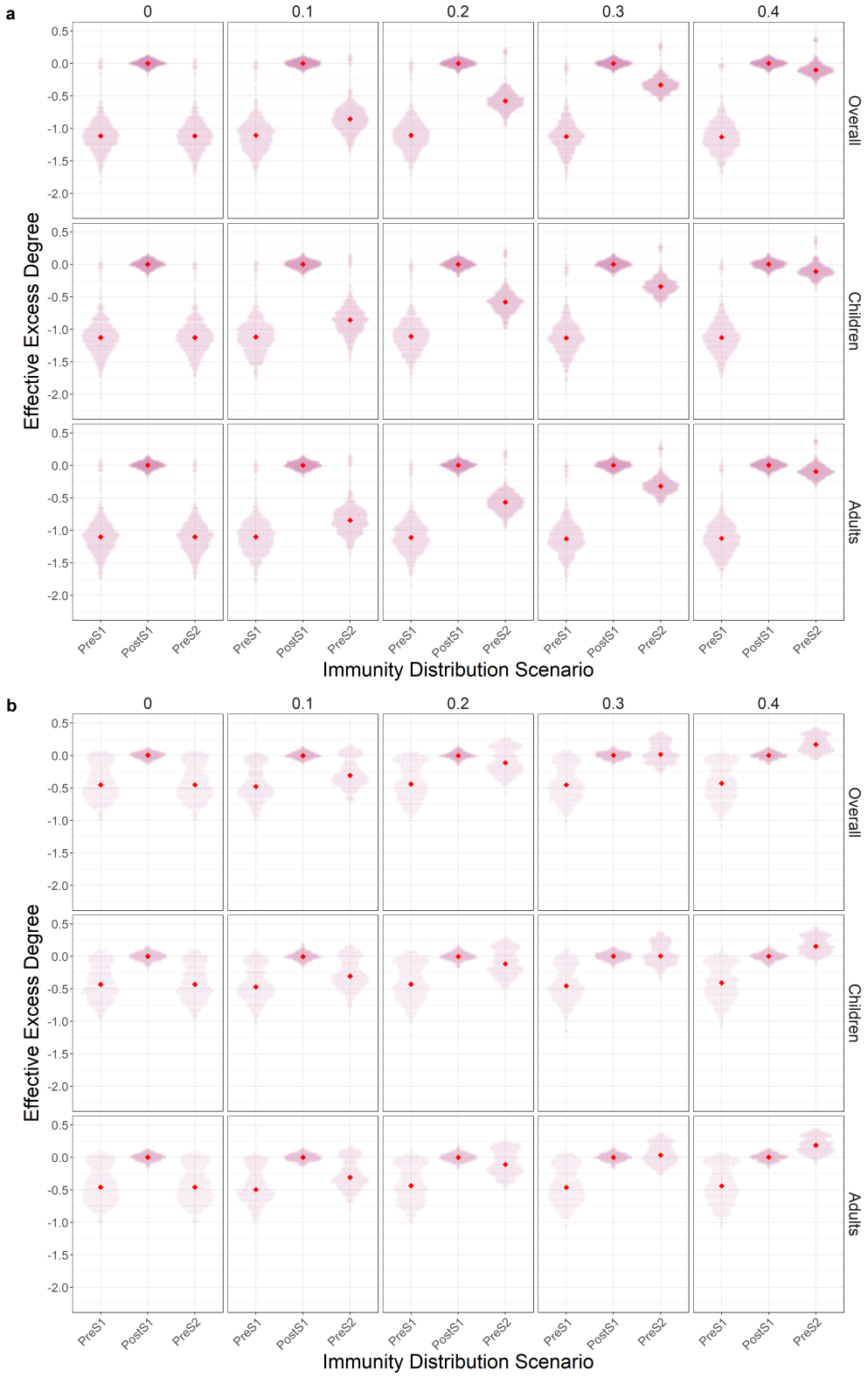


Fig. D.1 The impact impact of the S_{broad} immune class on the ability of interseasonal immune waning in restoring the connectivity of networks with "no structure". The S_{broad} immune class is absent in (a) and present in (b). The initial *pre-S1* susceptibility profile of both (a) and (b) are otherwise the same.

Fig. D.1 (Previous page.) The proportion of S_0 individuals resusceptibilised (δ) is shown by the top facet label, steadily increasing from left to right. For each facet plot, the left, middle and right distributions show the average effective excess degree distributions for the original *pre-S1*, residual *post-S1* and resusceptibilised *pre-S2* networks respectively; the red point represents the median value. To make the effective excess degree distributions comparable, for each combination and age group, *pre-S1*, *post-S1* and *pre-S2* values were normalised relative to the *pre-S1* mean. Due to the lack of any assortatitve mixing within age groups, S_{broad} are interspersed within the network, thus raising the *post-S1* and *pre-S2* levels of effective excess degree for both age groups.

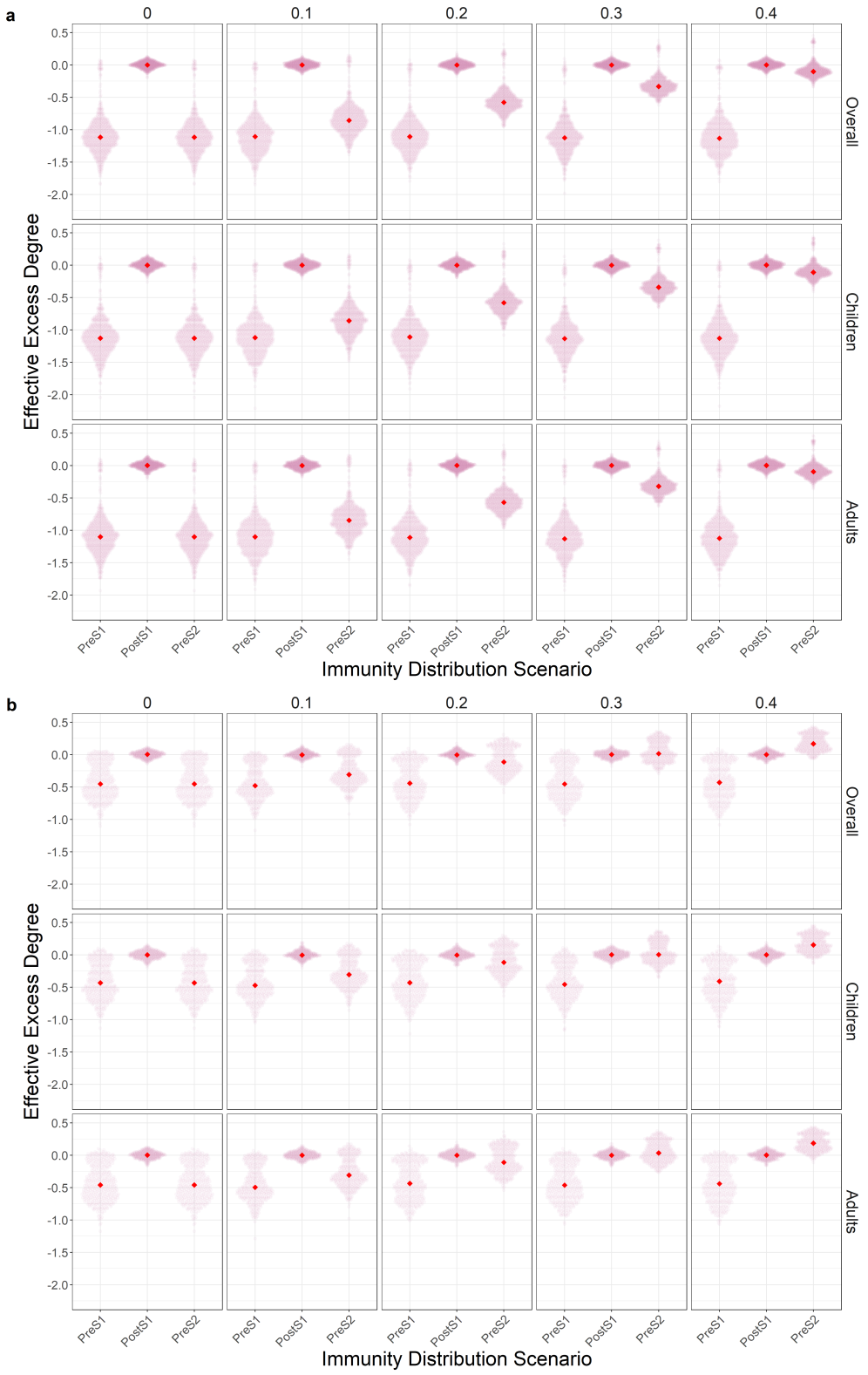


Fig. D.2 The impact impact of the S_{broad} immune class on the ability of interseasonal immune waning in restoring the connectivity of networks with "no structure". The S_{broad} immune class is absent in (a) and present in (b). The initial *pre-S1* susceptibility profile of both (a) and (b) are otherwise the same.

Fig. D.2 (Previous page.) The proportion of S_0 individuals resusceptibilised (δ) is shown by the top facet label, steadily increasing from left to right. For each facet plot, the left, middle and right distributions show the average effective excess degree distributions for the original *pre-S1*, residual *post-S1* and resusceptibilised *pre-S2* networks respectively; the red point represents the median value. To make the effective excess degree distributions comparable, for each combination and age group, *pre-S1*, *post-S1* and *pre-S2* values were normalised relative to the *pre-S1* mean. The presence of S_{broad} adults has its greatest, albeit still limited, effects on the *post-S1* and *pre-S2* connectedness of adults, with higher values of effective excess degree, the distribution of which exhibits lower variance. In contrast, the *post-S1* and *pre-S2* effective excess degree distributions for children, as well as the overall network, remain unchanged.

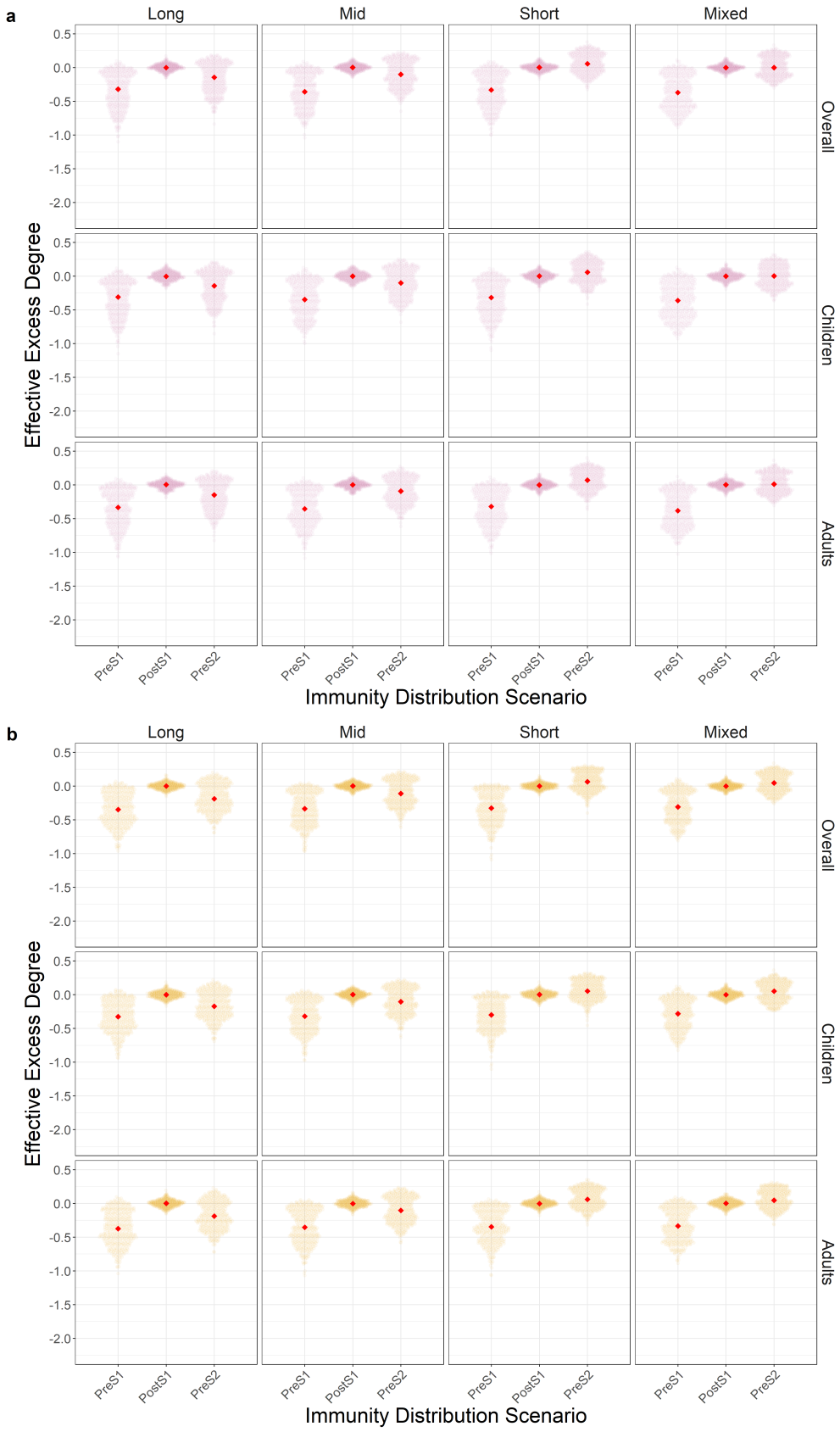


Fig. D.3 The impact of variable interseasonal immune waning on restoring connectivity in networks with (a) "*no structure*" and (b) "emphhousehold" structure. Immune waning scenarios and the associated age group-specific rates are listed in Table 6.1.

Fig. D.3 (Previous page.) The left, central and right distributions show the original *pre-S1*, residual *post-S1* and resusceptibilised *pre-S2* distributions for average effective excess degree respectively; the red point denotes the median value. To make the effective excess degree distributions comparable, *pre-S1*, *post-S1* and *pre-S2* values were log-transformed and normalised relative to the *pre-S1* mean for that particular combination. Whilst the distribution of immune classes differ between age groups, the lack of strong network structure or segregation means that children are only marginally more affected than adults by epidemics (Figs. C.3 and C.4); otherwise, children and adults are indistinguishable so little benefit is gained from differing immune waning rates between the age groups.

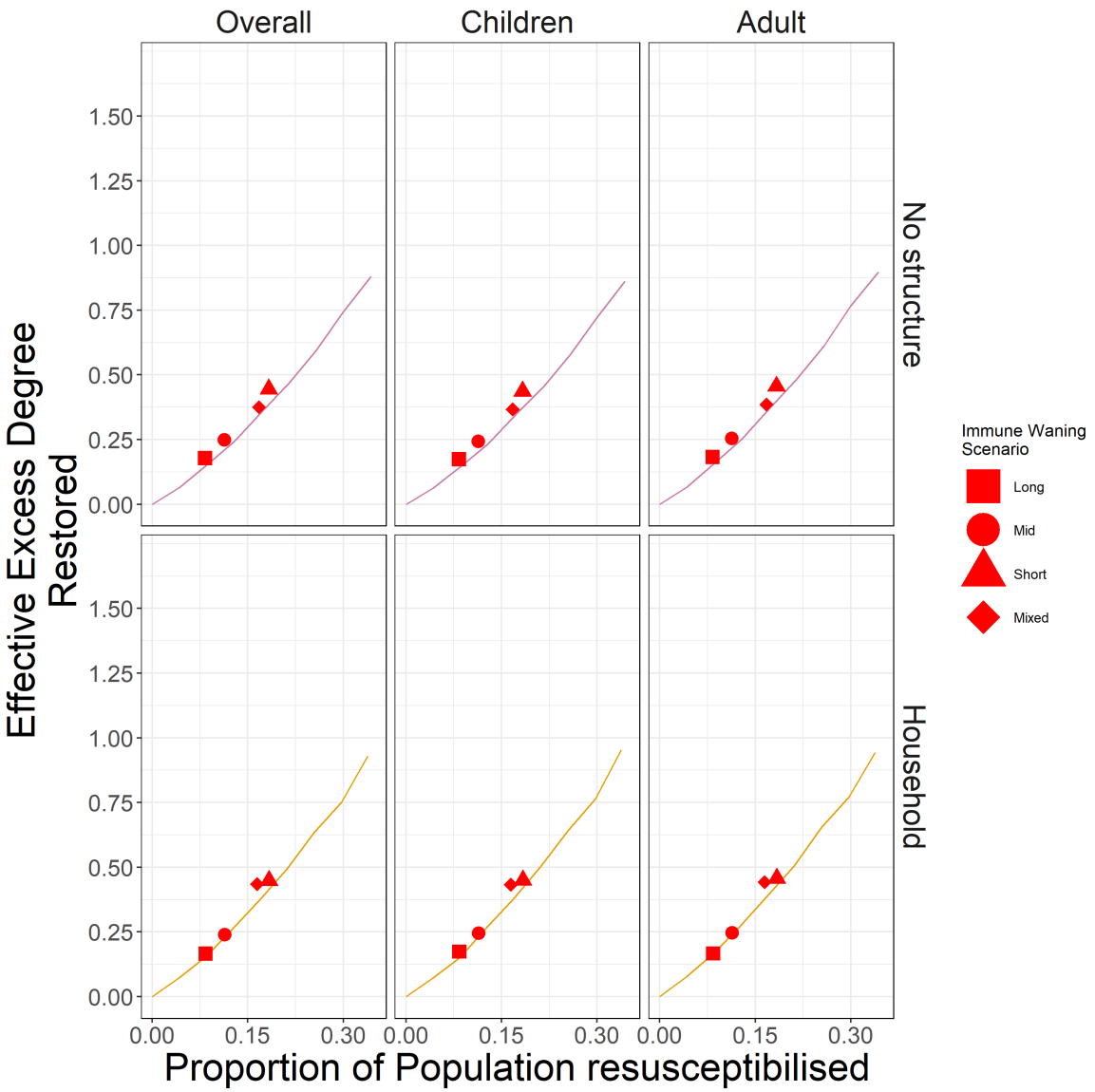


Fig. D.4 The effect of variable interseasonal immune waning in restoring the effective excess degree of the *post-S1* network. The various immune waning scenarios (see Table 6.1 for corresponding waning rates) are compared against the null homogenous waning model, which is denoted by the blue line that shows the mean amount of effective excess degree restored (Eq. (5.7)) for a given number of S_0 individuals resusceptibilised. With only weak correlations between nodal degree and susceptibility, varying the immune waning rates between age groups does little to restore network connectivity above and beyond the homogenous waning null models (denoted by lines).

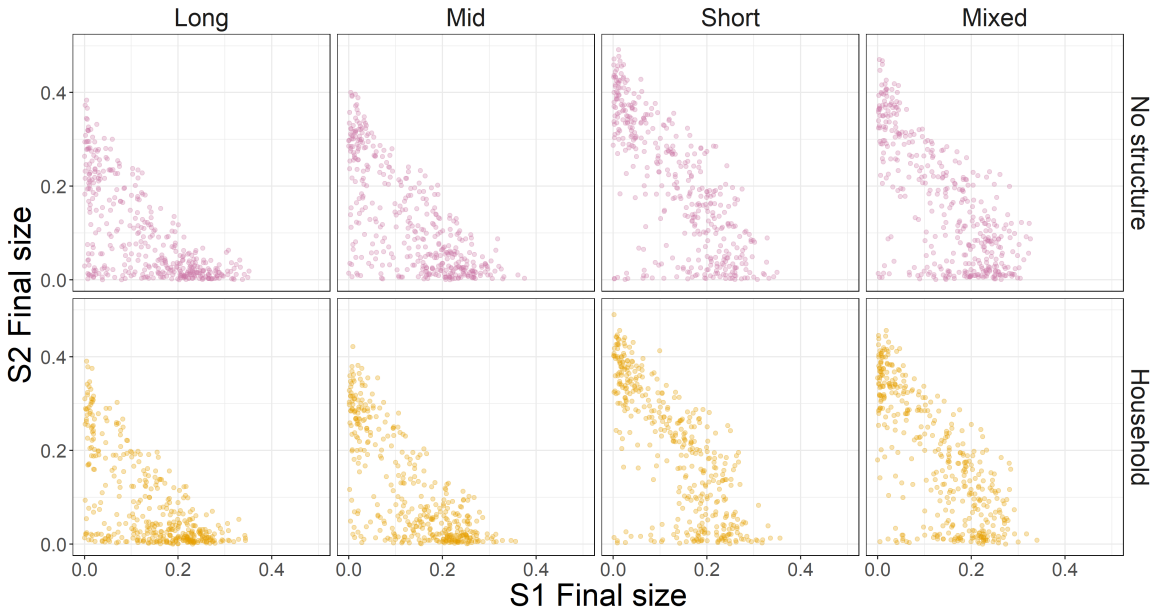


Fig. D.5 The effect of variable interseasonal immune waning on the final size of consecutive epidemics. The corresponding immune waning rates for each scenario are listed in Table 6.1. Individual points denote the final size for the *S1* and *S2* epidemics for a single model realisation. As in Chapter 5, the lack of significant network structure results In networks that are less affected by frailty, thus requiring less loss of immunity to enable *S2* epidemics.

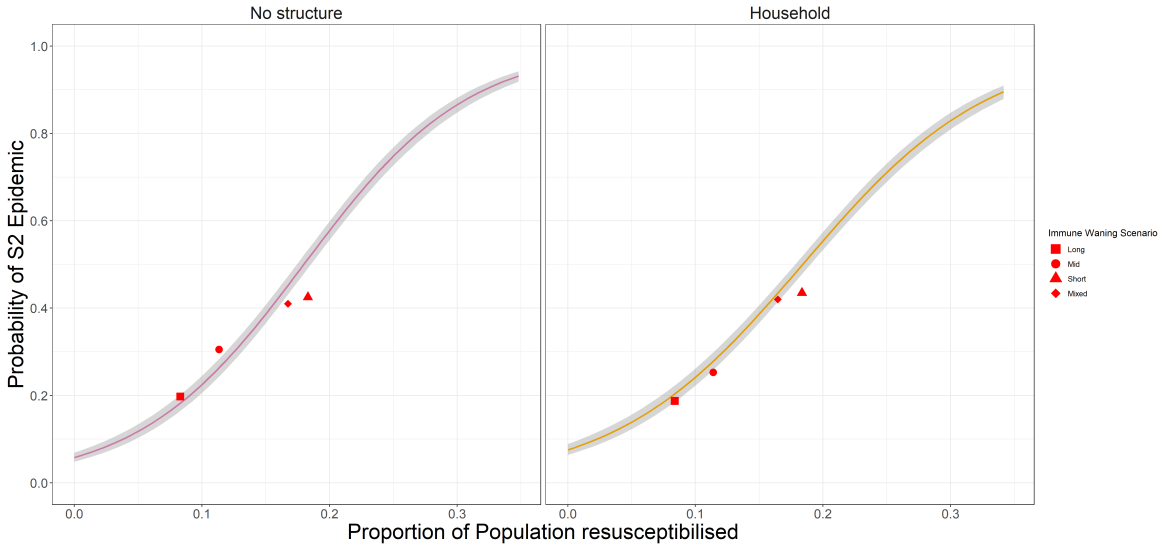


Fig. D.6 The effect of variable interseasonal immune waning on the probability of a major *S2* epidemic. The various immune waning scenarios (red points) were contrasted against the null homogenous waning model: a binary logistic regression model (blue line) was fitted to the latter and shows the probability of successful *S2* epidemics for a given number of S_0 individuals resusceptibilised. The probability of a successful *S2* epidemic is given by the proportion of realisations, with total incidence above 0.05. Again, varying the immune waning rates between age groups has little additional impact, given the fact that children and adults are basically indistinguishable, in the absence of strong correlations between nodal degree and susceptibility.

D.2 Appendix D Tables

Network Structure	Broadly Immune Class	Term	*OR*	*OR* adjusted SE	*p* value
No structure	Absent	Intercept	0.0287	0.00334	3.24E-204
		Proportion resusceptibilised	1430000	650000	3.11E-213
	Present	Intercept	0.319	0.0282	3.89E-38
		Proportion resusceptibilised	3350	1380	4.58E-86
Household	Absent	Intercept	0.0546	0.00543	6.51E-188
		Proportion resusceptibilised	141000	56000	4.30E-197
	Present	Intercept	0.345	0.0301	3.31E-34
		Proportion resusceptibilised	2450	997	4.73E-82
Structure	Absent	Intercept	0.128	0.0106	3.38E-136
		Proportion resusceptibilised	399	126	1.19E-80
	Present	Intercept	0.147	0.0152	1.90E-76
		Proportion resusceptibilised	206	78.7	4.23E-44

Table D.1 Binary logistic regression assessing the impact of broadly immune S_{broad} class on the probability of successful *S2* epidemics. Note that $OR > 1$ implies that greater interseasonal waning results in a greater probability of successful epidemic initiation.