Multi-host modelling of influenza A

Edward Hill, Michael Tildesley[a,b], Thomas House[a,b]

[a] *Centre for Complexity Science, University of Warwick, Coventry, UK*
[b] *Mathematics Institute, University of Warwick, Coventry, UK*

Abstract

Influenza A viruses are associated with most of the widespread influenza epidemics, and are the sole cause of the occasional global pandemics. They have many strains and the ability to inhabit many host species. We present a stochastic model, based on susceptible, infected and recovered dynamics, to explore the complex strain diversity and between-species epidemiology of influenza A. We apply the derived model in three distinct ways. First as a one host, multi-strain model, second as a multi-host, one strain model, and finally as a multi-host, multi-strain model. We address basic questions in strain and multiple host dynamics, focusing particularly on how the number of infected is affected by the type of model and the reasons behind this. Our results show how the ability of a strain to out compete others within a host population is linked to the length of the infectious period. The longer the infectious period, the more likely it is the strain will be able to persist and cause a greater number of infectious cases in the long-term. We show how the time taken for a strain to reach humanity from other animal species depends on the zoonotic pathway of the strain, with the pathway containing the lower number of species not always the quickest route.

1. Introduction

The virus influenza is a cause of considerable morbidity and mortality in humans. Influenza incorporates three virus types: types A, B and C. A key distinction in the molecular make up of these types is that type A and B influenza viruses contain eight RNA gene segments, while type C influenza viruses contain seven RNA gene segments. Consequently, influenza viruses exhibit different patterns of epidemiological and clinical behaviour [1]. Influenza A, in particular, has the ability to inhabit many host species and has many strains. As a result, type A viruses are associated with most of the widespread influenza epidemics, and are the sole cause of the occasional global pandemics. It remains of critical importance to understand how likely it is that more lethal influenza A strains will cause a pandemic in the human population.

Influenza A is notable for its annual epidemics and antigenic drift dynamics. The two most abundant surface proteins of type A influenza viruses, hemagglutinin (HA) and neuraminidase (NA), show the greatest extent of variability, undergoing subtle, minor changes in structure almost annually. Consequently, a large number of strains can be in circulation within a population at any given time. An antibody response raised against one strain may or may not be effective against a mutation of the original strain. Antigenic drift is the major reason for the success of these viruses in evading human immune mechanisms [1].

Humanity however does not act as the primary reservoir for type A influenza viruses. This role is fulfilled by wild avian species. Mutation of the viral genes not only creates different strains, but sometimes produces a strain that can be transmitted into a different host species. Of particular importance is the transmission between these wild avian species and domesticated poultry. In wild ducks, influenza A viruses cause no disease, although they can be excreted in high concentrations in the faeces. Ducks are essentially asymptomatic carriers of all influenza strains. Two types of avian influenza infection are transmitted from wild waterfowl to farmed ducks, chickens, turkeys and geese - referred to as LPAIV (low-pathogenic avian influenza virus) and HPAIV (high-pathogenic avian influenza virus). High-pathogenic avian influenza viruses are highly lethal to infected poultry, spreading very rapidly in the local region and leading, in outbreaks in recent times, to the loss of more than 100 million birds through disease and culling. However, on initial

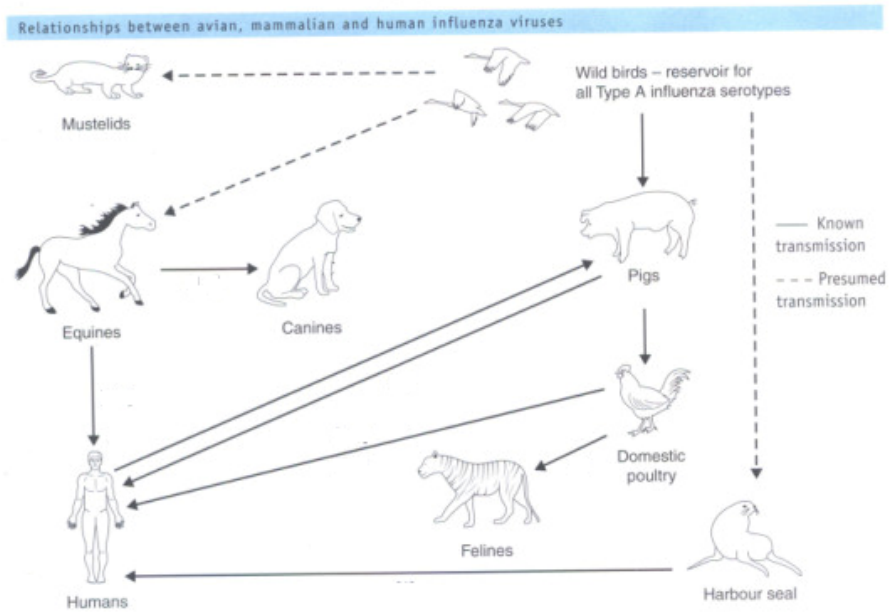

Fig. 1. Influenza type A transmitted between different animal species and humans. Adapted from [1].

infection of poultry these strains are avirulent (unable to produce disease). It is only following spread within the infected flock and mutation of the viral genes providing adaptation to their host that they acquire virulence and lethality. By a similar process, it is believed type A avian influenza viruses spread to, and caused severe outbreaks of illness and disease, in pigs, horses, harbour seals and whales. Influenza A now has a number of possible transmission routes between different animal species and humans (Fig. 1). The epidemiology of a single strain can be radically different depending on the host species [1].

Very occasionally, humans become infected with a virus bearing HA and/or NA antigens derived from non-human sources. These are essentially novel to humans. The HA and/or NA surface antigens of such viruses are not initially recognised by the specific human host defence mechanisms. Due to the viruses meeting with little or no established resistance, they can, following mutation and adaptation to their new host, spread relatively easily in the human species. This can give rise to a localised outbreak that may develop into a worldwide influenza pandemic [1]. An area of interest is the length of time it takes for a strain circulating within animal host species to be transmitted to humanity, referred to as the hit time.

The grave impact of an influenza pandemic was shown in recent years, caused by H1N1. In April 2009, H1N1 was first detected in the United States. This virus was a unique combination of influenza virus genes never previously identified in either animals or people. The virus genes were a combination of genes most closely related to North American swine-lineage H1N1 and Eurasian swine-lineage H1N1 influenza viruses. However, it became apparent that this new virus was circulating among humans and not among U.S. pig herds [2]. Shortly before the end of the pandemic (as of 1st August 2010), worldwide more than 214 countries and overseas territories or communities had reported laboratory confirmed cases of H1N1, including over 18,449 deaths (see Fig. 2) [3]. Work has since been carried out studying the 2009 H1N1 influenza pandemic, including the transmission characteristics and risk factors [4, 5]. A current concern is the H7N9 outbreak in China. Influenza A H7 viruses are a group of influenza viruses that normally circulate among birds. Although some H7 viruses (H7N2, H7N3 and H7N7) have occasionally been found to infect humans, no human infections with H7N9 viruses had been reported until recent reports from China. The disease is of concern because most patients have been severely ill, with many suffering from severe pneumonia. [6]. WHO was first notified of cases of H7N9 virus cases in China on 31st March 2013. As of 17th May 2013, a total of 131 laboratory-confirmed cases of human infection with H7N9 virus in China, including 36 deaths, had

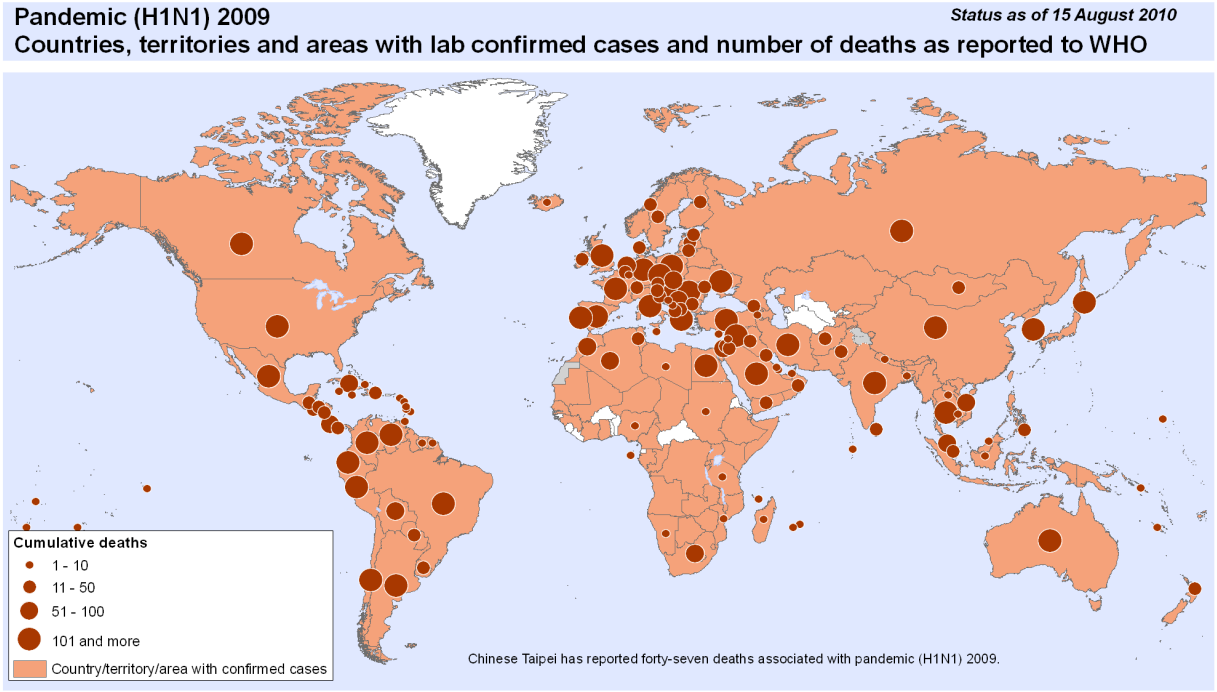

Fig. 2. Pandemic (H1N1) 2009 affected countries and deaths, status as of 15-August-2010. Adapted from [3].

been reported to WHO [7].

The complexity of the multi-strain, multi-host problem has been approached in a number of ways. The main obstacle to modelling systems with more than a few strains is the fact that the number of dynamic variables tends to grow with the number of strains. Gog and Grenfell [8] propose a simple model that is capable of capturing the dynamics of a large number of antigenic types that interact via host cross-immunity. This manages the complexity of many strains by using a status-based formulation, where the current immune state of the host is considered rather than the very complex immune history of exposure. Arinaminpathy and McLean [9] use a simple mathematical model to study potential epidemiological markers of adaptation, where pathogens are able to cross the species barrier and become established in a new host species, in this case humans. The model simulates a series of introductions of evolving pathogens via a multi-type branching process, leading to emergence. Modelling of specific classes of influenza A virus have been carried out. Smith et al. [10] quantified and visualized the antigenic evolution of influenza A (H3N2) virus in humans, where their approach offered a route to predicting the relative success of emerging strains.

In this study we illustrate a stochastic framework to model the complex strain diversity and between-species epidemiology of influenza A. We initially describe the dynamics in two simplified settings; a one host, multi-strain model, and a multi-host, one strain model. The one host, multi-strain model is used to explore how the transmission and recovery rates impact the ability of a strain to out compete all others, when every strain has the same reproductive ratio. The multi-host, one strain model is used to study the critical value of the cross-host transmission parameters that would prevent a cross-transmission event occurring, and how this is impacted by population size. We then consider a four-host multi-strain model, involving ducks, chickens, pigs, and humans, and consider two general questions: how does strain specific transmission, where some strains are more likely to cross into the human population than others (through different zoonotic pathways for example), impact the strain hit time? What happens to the strain hit time if varying strengths of cross-strain immunity, where infection by past strains may help protect individuals against infection from new strains, are introduced?

2. Methods

In standard models of disease transmission, individuals are classified with respect to their disease status. The chosen model was a stochastic susceptible-infected-recovered (SIR) model, which incorporated a demographic process. A given individual in a host population is classified in one of three classes for each strain. Susceptible individuals can catch the given strain if they are exposed to it. Infected individuals are infectious. They can spread the strain to any susceptible individuals they have contact with. Recovered individuals, alternatively referred to as removed individuals, are immune against infection from that strain. Either they have had the disease and recovered, or being infected by a different strain made them immune via cross-immunity.

The model was simulated by applying the Gillespie algorithm, a dynamic Monte Carlo method. Doing this allowed us to interpret parameters we could actually measure. In theory it would have been possible to derive analytic expressions describing the process, though these would have been too complex to solve (for example, see equation 1). A deterministic model was also considered, modelling the problem using ordinary differential equations. However, with each added strain and/or host the number of classes required grew, quickly becoming computationally intensive. Though this model gave the right temporal dynamics, a major issue using the deterministic model was we could get classes with occupancy values less than 1. Having such small values was not realistic. This would have prevented us from capturing features such as stochastic die out, which are of biological interest. A stochastic model analysed via simulations was deemed the most appropriate approach.

In what follows we made a number of assumptions; all newborns were susceptible to all strains, every individual was equally susceptible to a given strain, there was homogeneous mixing within the population, super infection was not possible (i.e. an individual could only be infected by at most one strain at a particular time), frequency dependence was used for intra-host susceptible/infected interactions, while density dependence was used for inter-host susceptible/infected interactions.

We made an important distinction between intra-host and inter-host interactions, following the approach used by Keeling and Rohani [11] in their multi-host and multi-strain models. Frequency dependence is where the force of infection (involving the interaction between susceptibles and infectious) is divided by the total population size. This is more applicable when the density of individuals is independent of population size. Therefore, frequency dependence was used to model intra-host susceptible/infected interactions. To model inter-host susceptible/infected interactions, the transmission term was not divided by the total population size. As we were dealing with separate populations and the interaction was likely to depend on the density of the species involved, we assumed density dependence.

The transition rates used in the model were assumed to be Poisson. For a given strain k, they were defined as follows:

Host j susceptible individual infected by host j infectious individual:

$$T\left(S_k^j - 1, I_k^j + 1, R_k^j | S_k^j, I_k^j, R_k^j\right) = \left(\beta_k^{j,j} I_k^j \frac{S_k^j}{N^j}\right)$$

Spontaneous infection of a host j individual:

$$T\left(S_k^j - 1, I_k^j + 1, R_k^j | S_k^j, I_k^j, R_k^j\right) = \alpha_k^j$$

Host i susceptible individual infected by host j infectious individual:

$$T\left(S_k^i - 1, I_k^i + 1, R_k^i | S_k^i, I_k^i, R_k^i\right) = \beta_k^{i,j} S_k^i I_k^j$$

Recovery of infected host j individual from strain k:

$$T\left(S_k^j, I_k^j - 1, R_k^j + 1 | S_k^j, I_k^j, R_k^j\right) = g_k^j I_k^j$$

where N^j is the host j population size, S_k^j, I_k^j and R_k^j are the total number of susceptible, infected and recovered host j individuals with respect to strain k, $\beta_k^{i,j}$ is the transmission rate of strain k to host species i from host species j, g_k^j is the recovery rate of a host j individual from strain k, and α_k^j is the spontaneous infection rate of a host j individual with strain k. Spontaneous infection was included here for completeness, but was not included in any of the simulations that produced our results.

Note, as we assumed an individual could not be infected by more than one strain at any given time, while they were infected by strain k they were no longer susceptible to any other strain. Thus, the number of individuals in class S_l^j, where $l \neq k$, also decreased by one. Once an individual had recovered from strain k, they became susceptible to all other strains once more. Thus, the number of individuals in class S_l^j, where $l \neq k$, increased by one. Cross-immunity was added in by assigning a probability to an individual gaining immunity to strain l when infected by strain k, where $l \neq$ k, assuming they were not already immune to strain l. If the cross-immunity event was a success, the individual was re-classified as being recovered with respect to strain l.

A demographic process was also included, with both a birth/death of a host j individual given by the rate $\delta^j N^j$, where δ^j is the individual death rate/birth rate for host species j.

One host, multi-strain model

Using a simplified version of the full model described above, we first looked at a one host, multi-strain model. In this section, we assumed there was no cross-strain immunity. The dynamics for a three strain, four strain, and five strain model were obtained. All results obtained for this model were averaged over 100 simulation runs. Each strain had the same reproductive ratio of 2, but no two strains had identical transmission and recovery rates. This allowed us to analyse what factors resulted in a particular strain emerging as the dominant strain (i.e. cause the greatest number of infected cases) out of a group, and whether this was impacted by the initial population size. In all simulations the initial conditions were fixed, with 5 distinct individuals infected by each strain, and all remaining individuals infected by all strains. Rather than starting with 1 infected individual per strain, 5 was chosen to reduce the chance of immediate die out. As the process is Markovian, and so memoryless, any future dynamics are not impacted by the starting initial condition. However, we still chose a low level of initial infected to be able to capture events such as die out, and to avoid biasing the probabilities in such a way that meant we influenced the dynamics.

Multi-host, one strain model

The second simplified version of the full model to be studied was a multi-host, one strain model. In this section we also assumed there was no cross-strain immunity.

Our main focus was on a two host, one strain model. The two hosts represented chickens and ducks. A single influenza A strain has two different behaviours for these species. In chickens, influenza A is highly transmissible, but the infectious period is very short. In contrast, due to influenza A being asymptomatic in ducks, less is known about the transmission rate from this species due to the difficulty in spotting when a duck is infected. However, the infectious period of influenza A is longer in ducks compared to chickens. The strain was set to have a reproductive ratio of 2 in both ducks and chickens. We investigated the dynamics obtained for the various susceptible and infected classes. Following this, we looked at the impact of adding more hosts.

One part of our analysis for the two host, one strain model was the critical value of the cross-host transmission parameters that would prevent a cross-transmission event occurring, and how this was impacted by population size. To find the critical value, the number of cross-species transmission events from one host to another was recorded up until the simulation run time reached 50 years. This was averaged over a number of simulations. As stochastic fluctuations were possible, the approximate critical value was taken to be when the average number of cross-transmission events was below 1 in the specified time period.

Further analysis was performed on the time series corresponding to the number of susceptible and infected individuals for a given host, by calculating cross-correlations and power spectra. The cross-correlation gives

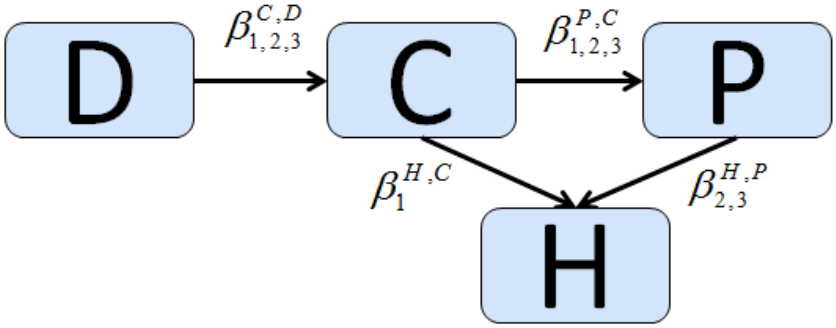

Fig. 3. Diagram of the possible routes of cross-transmission between different host species for the four host model. The numbers in the subscripts indicate the strains that are able to cross between those two particular hosts.

a measure of the similarity of two waveforms as a function of a time-lag applied to one of them. Defined mathematically:

$$(f \star h)[n] = \sum_{m=-\infty}^{\infty} f^*[m]h[n+m]$$

where f and h are continuous functions, and f^* denotes the complex conjugate of f. The power spectrum quantified the periodicity of the oscillations. This showed how the variance of the fluctuations was distributed over different spectral frequencies. A sharply peaked spectrum indicated structured oscillations with a dominant frequency where the spectrum was peaked. Due to the time series data points being unevenly spaced, the data was first interpolated, creating a time series with evenly spaced data points, on which a power spectrum calculation could be performed.

Full Model - four hosts, two/three strains

An application of the full model was used with four hosts and both two and three strains. The four hosts represented ducks, chickens, pigs and humans. The possible routes for strain cross-transmission are shown by the directed arrows in Fig. 3. Each simulation would begin with only the duck population containing infected individuals. Cross-transmission between hosts had to occur for any strain to infect a different host population from ducks. The model was simplified by each strain only having one possible transmission route from ducks to humans.

The magnitudes of the cross-transmission parameters chosen satisfied the following criteria:

$$\beta^{C,D} > \beta^{P,C} \geq \beta^{H,P} > \beta^{H,C}$$

The cross-transmission between ducks and chickens is more regular due to their genetic similarity, while the geographical closeness of chicken and pig populations results in cross-transmission events being a common occurrence. Compared to chickens, humans have a greater genetic similarity to pigs, due to them both being mammals. This increases the likelihood that the source species of a strain cross-transmitted into humans being pigs.

We were interested in how the pathway of infection and varying cross-immunity parameters would influence the hit time. Three distinct sets of cross-immunity parameters were defined; no cross-immunity, intermediate cross-immunity, full cross-immunity. In the no cross-immunity scenario, all cross-immunity parameters were zero for every host. For intermediate cross-immunity and full cross-immunity, chicken, pig and human individuals had a probability of 0.5 and 1 respectively of gaining cross-immunity to a strain they were still susceptible to, when infected by any strain. All strains the individual was still susceptible to were considered, each one separately. Ducks were assumed to gain no cross-immunity in both the intermediate and full cross-immunity scenarios, all related parameters were zero. This was to fit with the assumption that they are a reservoir for all influenza A strains.

A final factor that was considered was the impact cross-immunity had on the hit time of a strain that was introduced into the system at a later time. We investigated this using two strains. At the system start time, a strain was introduced in the duck population. A second strain was introduced into the duck population after a specified delay time. This was set to be equivalent to a week, and allowed time for the first strain to spread. Each of the three cross-immunity parameter sets was tested.

We used a survival function to capture the probability that a given strain will not have reached humanity by a specified time. The survival function is a property of any random variable that maps a set of events onto time. With T being a continuous random variable with cumulative distribution function $F(t)$, it is formally defined as follows:

$$P(T > t) = \int_t^\infty f(u)du = 1 - F(t)$$

It was possible to derive an expression for the probability that a given strain will have reached humanity by a specified time. The survival function could then be obtained. With the cross-transmission rate assumed to be Poisson, the waiting time for a cross-transmission event follows an exponential distribution. At time t, the probability a cross-transmission event occurs from an infected host j individual to a susceptible host i individual is:

$$F(t) = \left(1 - e^{-\int_0^t \beta^{i,j} I^j(s) S^i(s) ds}\right)$$

Taking into account all possible routes a strain can take to reach the target host, and the individual cross-transmission steps within each route, the solution for the probability of the strain reaching humanity by time t is given by:

$$F(t) = \sum_{\text{routes}} \left(\prod_{\text{Steps in route}} \left(1 - e^{-\int_0^t \beta^{i,j} I^j(s) S^i(s) ds}\right) \right) \tag{1}$$

We are unable to solve equation (1) analytically, due to the number of infected and susceptible individuals of each host changing with time. Therefore, simulations were used to obtain an approximate value for the probability of a strain reaching humanity by a given time t. Corresponding survival curves were produced using this approximation, calculated by $1 - F(t)$.

3. Results

One host, multi-strain model

For a three strain, four strain, and five strain model, with one host that starts with an initial population of 10,000, there was a consistent pattern in the number of individuals infected (Fig. 4). Under the assumption of all strains having the same reproductive ratio, the strain that is most transmissible is the first to have a large outbreak in the host population. However, these strains have a short infectious period (a week or less, corresponding to $g \geq 52$), so die out quickly. The strains that are less transmissible have an initial outbreak that occurs later, though the length of this is longer compared to the more transmissible strains. This is shown by the wider spikes in number of infected individuals in Fig. 4. The strains with longer infectious periods are able to persist, oscillating around a stable number. Comparing all the strains, the dominant strain is the one that has the longest infectious period, rather than the greatest transmission rate. Note, as the model is stochastic, the number infected by each strain goes to zero eventually when there is no spontaneous infection of individuals included in the system.

The major features of the dynamics appeared unaffected when the initial host population sizes were varied (Fig. 5). One noticeable difference was an increase in initial population size resulted in a slower decline in the number of infected to 0 for strains that are highly transmissible, but with a short infectious period. In our case, this can be seen with the dynamics for strain 1 and 2 with the varied initial population sizes. This corresponded to the average extinction time of these strains being increased.

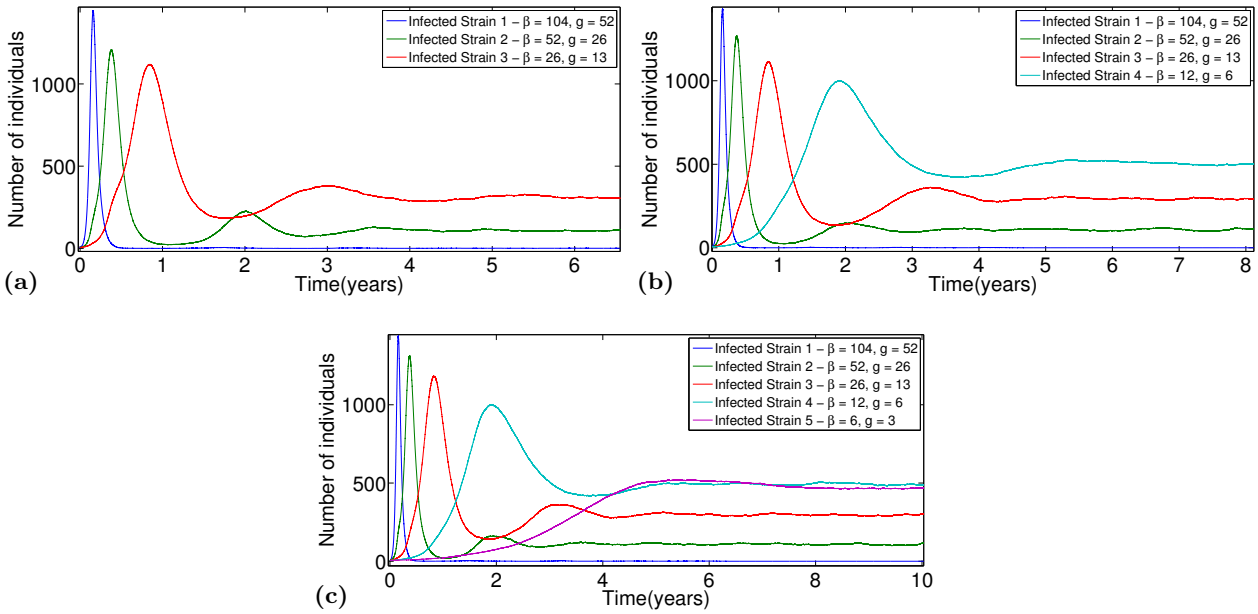

Fig. 4. Number of individuals infected by each strain for a (a) three strain, (b) four strain, and (c) five strain model, averaged over 100 simulations. All strains had a reproductive ratio of 2. For each simulation, the initial starting populations were 10,000 individuals, with 5 distinct individuals infected by each strain. Individual birth and death rates were set to 1.

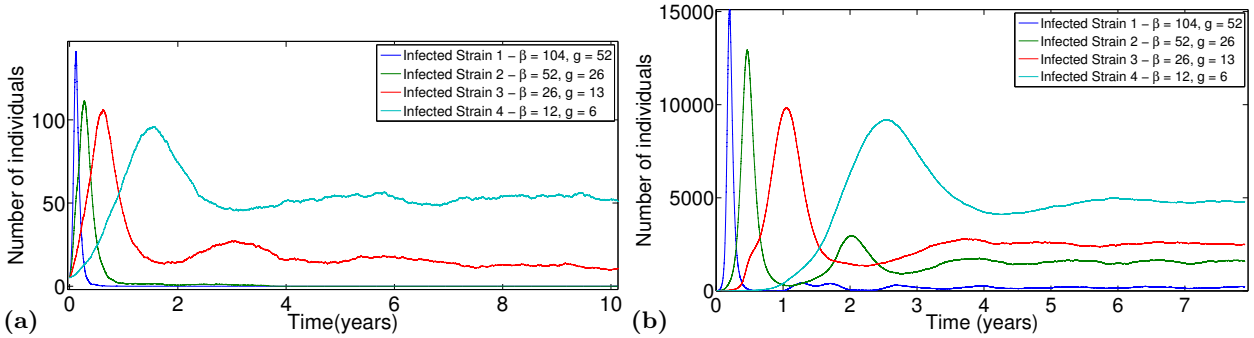

Fig. 5. Number of individuals infected by each strain in a four strain model. The initial starting populations were (a) 1,000 and (b) 100,000 individuals, with 5 distinct individuals infected by each strain. All strains had a reproductive ratio of 2. The model with an initial host population of 1,000 individuals was averaged over 100 simulations. The model with an initial host population of 100,000 individuals was averaged over 10 simulations. Individual birth and death rates were set to 1.

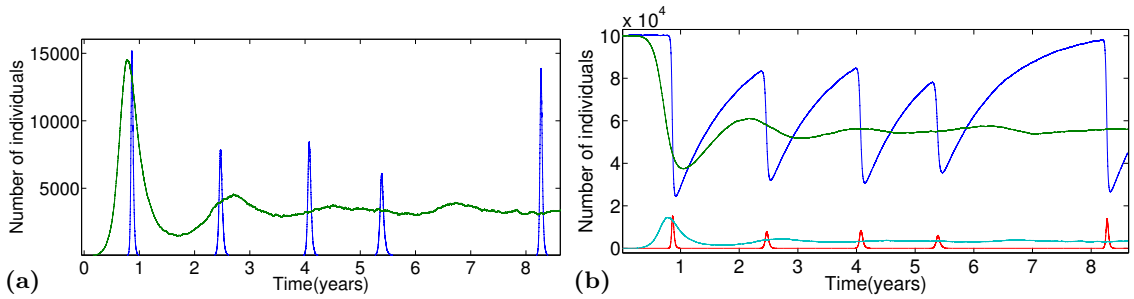

Fig. 6. Two host, one strain plots from one simulation. The strain had an R_0 value of 2 in both host species. For both types of host, the initial population size was 100,000. Individual birth and death rates were set to 1. Graph (a) shows the number of individuals per host infected by the strain with respect to time. Dark blue line - Host 1 infected, Green line - Host 2 infected. Graph (b) shows the number of individuals per host infected by, and susceptible to the strain with respect to time. Dark blue line - Host 1 susceptible, Green line - Host 2 susceptible, Red line - Host 1 infected, Light blue line - Host 2 infected. On all graphs, for host 1 (chickens), $\beta_1 = 208, g_1 = 104$, and for host 2 (ducks), $\beta_2 = 26, g_2 = 13$. Cross-transmission was only possible from host 2 to host 1, with $\beta^{1,2} = 1 \times 10^{-8}$. At the start of the simulation, there were 5 host 2 individuals infected and 0 host 1 individuals infected.

Multi-host, one strain model

A representative example of the infection dynamics for our two host, one strain model, where the two hosts represented chickens and ducks is given in Fig. 6a. It shows the number of individuals per host infected by the strain with respect to time. We restricted our attention to a strain that could be transmitted from ducks to chickens, but not from chickens to ducks. Following an initial epidemic, the number of host 2 infected individuals settles and oscillates about a stable value. In the host 1 population, the strain caused a series of intermittent short epidemics. After each outbreak, the strain was unable to persist in the host 1 population. The dynamics presented here replicate what was observed in real life during influenza pandemics, for example HPAIV H5N1 in South Asia. The spatial and temporal dynamics, along with risk factors, of the H5N1 infection in humans and poultry have been well studied in this region, in particular within China [12, 13], Thailand [14], Vietnam [15] and Bangladesh [16].

The number of individuals per host infected by, and susceptible to the strain with respect to time is shown in Fig. 6b. When a large outbreak occurred in either host population, the respective susceptible population suffered a sharp decline. A larger pool of host 1 individuals susceptible to the disease when the strain was cross-transmitted into the population resulted in greater numbers of infected individuals for that particular outbreak. Following each outbreak the number of host 1 susceptibles would recover, due to being replenished by births. The number of host 2 susceptible individuals eventually settled to a relatively stable value, corresponding to when the number of host 2 infected individuals also became more settled. However, both still display minor fluctuations.

Adding in more hosts, for which the strain has properties similar to when it infects chickens, led to the dynamics replicating the two host case (Fig. 7). Assuming the strain only initially infects individuals within one host population, the frequency of outbreaks for all remaining hosts was dependent on the cross-transmission parameter between itself and the host population where the strain initially started. It was found that raising the order of magnitude of the total number of individuals across all hosts results in a drop in the critical value of the cross-transmission parameter by the same order of magnitude (Table 1).

Using a representative example of two host, one strain dynamics (Fig. 8a), Figs. 8b-8g give the cross-correlations for all 6 pairs of combinations for the host 1 susceptible and infected, and host 2 susceptible and infected time series. The sequences were normalised so the autocorrelations at zero lag were identically 1. Both the host 1 susceptible and infected, and host 2 susceptible and infected cross-correlations (Figs. 8b-8c) had a negative peak when the time lag was 0, with a positive peak with a time lag of approximately -1 year. The remaining 4 combinations gave outputs that were viewed as noise, with no similarity between the waveforms. The power series for the host 2 infected time series, and both susceptible time series show

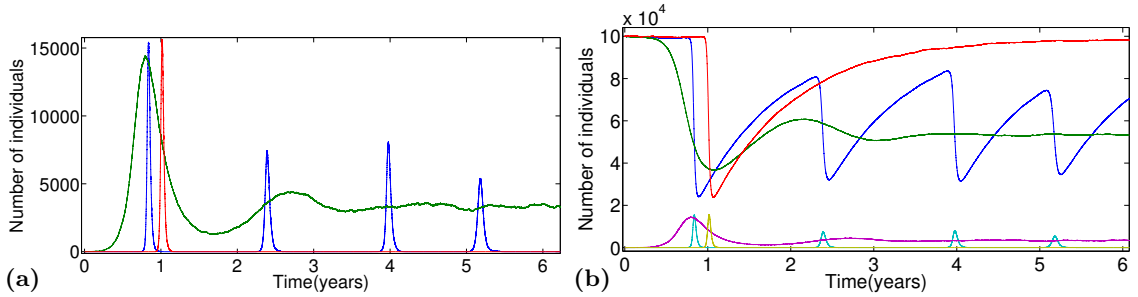

Fig. 7. Three host, one strain plots from one simulation. The strain had an R_0 value of 2 in all host species. For all three hosts, the initial population size was 100,000. Individual birth and death rates were set to 1. Graph (a) shows the number of individuals per host infected by the strain with respect to time. Dark blue line - Host 1 infected, Green line - Host 2 infected, Red line - Host 3 infected. Graph (b) shows the number of individuals per host infected by, and susceptible to the strain with respect to time. Dark blue line - Host 1 susceptible, Green line - Host 2 susceptible, Red line - Host 3 susceptible, Light blue line - Host 1 infected, Purple line, Host 2 infected, Yellow line - Host 3 infected. On all graphs, for host 1 and host 3 , $\beta_{1,3} = 208, g_{1,3} = 104$, and for host 2, $\beta_2 = 26, g_2 = 13$. Cross-transmission was only possible from host 2 to host 1, with $\beta^{1,2} = 1 \times 10^{-8}$, and host 2 to host 3, with $\beta^{3,2} = 1 \times 10^{-9}$. At the start of the simulation, there were 5 host 2 individuals infected, 0 host 1 individuals infected, and 0 host 3 individuals infected.

notable peaks for frequencies under 1Hz (Fig. 9).

Full Model - four hosts, two/three strains

All simulations using this model started with initial host population sizes of 100,000 for ducks, chickens and pigs, and 10,000 for humans. Individual birth and death rates were set to 1. Within each host, all strains had the same transmission and reproductive rates. The rates used were: Ducks, $\beta = 26$, $g = 13$; Chickens, $\beta = 208$, $g = 104$; Pigs and Humans, $\beta = 52$, $g = 26$. For each strain, the possible routes of cross-transmission between host species were different (Fig. 3), as were the values of the cross-transmission parameters.

First, we considered only two strains that were both introduced into the duck population at the same time. For all three cross-immunity conditions, on average strain 2 reached humanity in less time compared to strain 1. While the strain 2 hit times obtained similar mean and standard deviation values for each cross-immunity case, the strain 1 hit time was more varied and longer when full cross-immunity was used (Fig. 10a). The strain 2 survival curves are very similar for each cross-immunity condition, further showing the little impact cross-immunity condition had on the strain 2 hit time. The strain 1 survival curves are also similar up until a probability of approximately 0.15, where the curve corresponding to full cross-immunity begins to decrease less severely (Fig. 11a). This is what results in the increased variability and higher average hit time for strain 1 in the full cross-immunity case.

For the three strain case, the relative hit times values when comparing strains 1 and 2 are similar to the two strain case, though both have slightly longer and more variable hit times. Strain 3 is the quickest strain

Table 1. Approximate critical rates for cross-species transmission for varying starting population sizes. Both host 1 and host 2 start with the same initial population size. The strain transmission and recovery rates used were for host 1 (chickens), $\beta_1 = 208, g_1 = 104$, and for host 2 (ducks), $\beta_2 = 26, g_2 = 13$.

Initial population	Approximate critical cross-transmission rate)
$1,000$	$\mathcal{O}(10^{-7})$
$10,000$	$\mathcal{O}(10^{-9})$
$100,000$	$\mathcal{O}(10^{-11})$

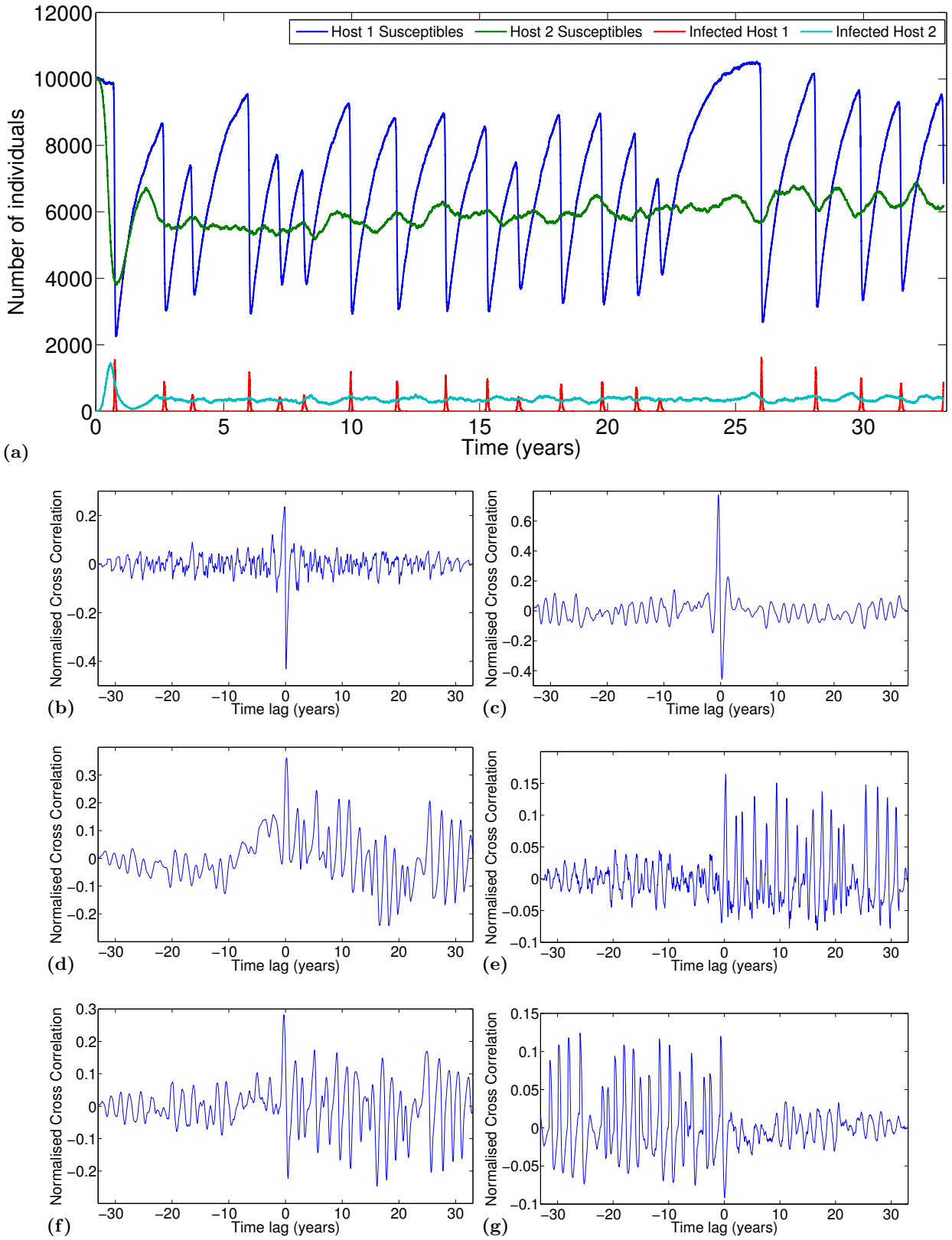

Fig. 8. Cross-correlations from a representative example of two host, one strain dynamics, shown in (a). The sequences were normalised so the autocorrelations at zero lag were identically 1.0. (b)-(g) show the cross-correlations for: (b) host 1 susceptibles and infected, (c) host 2 susceptibles and infected, (d) host 1 and host 2 susceptibles, (e) host 1 and host 2 infected, (f) host 1 susceptibles and host 2 infected, (g) host 2 susceptibles and host 1 infected.

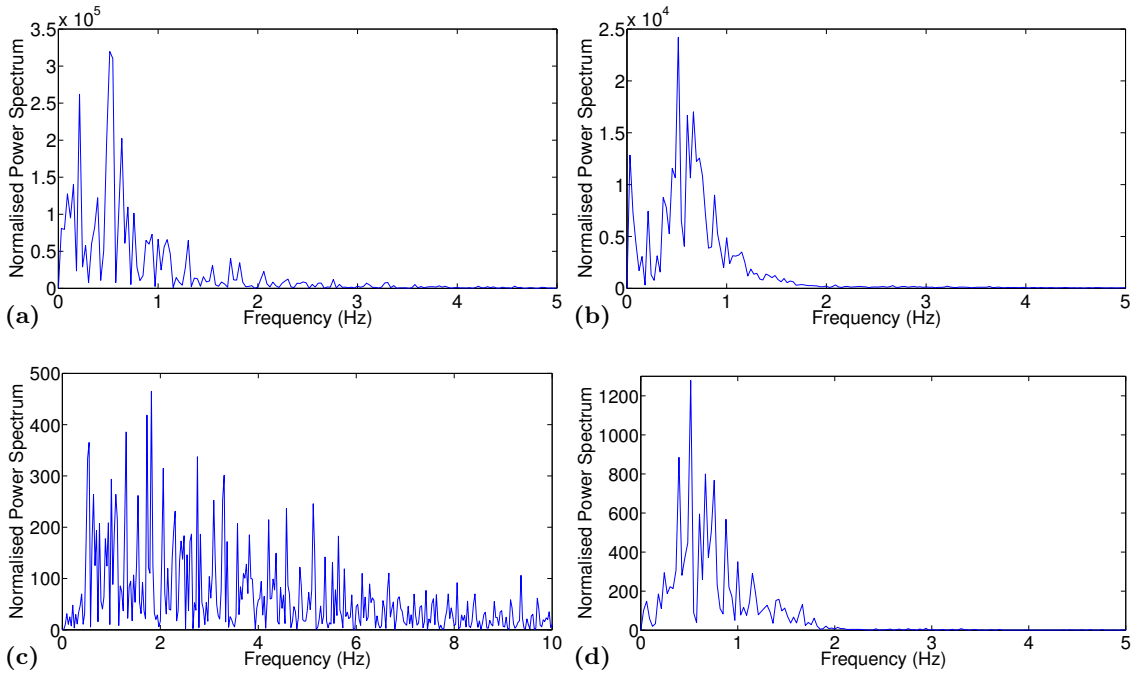

Fig. 9. Power spectra for the (a) host 1 susceptible time series, (b) host 2 susceptible times series, (c) host 1 infected time series, (d) host 2 infected time series from a representative example of two host, one strain dynamics (Fig. 8a).

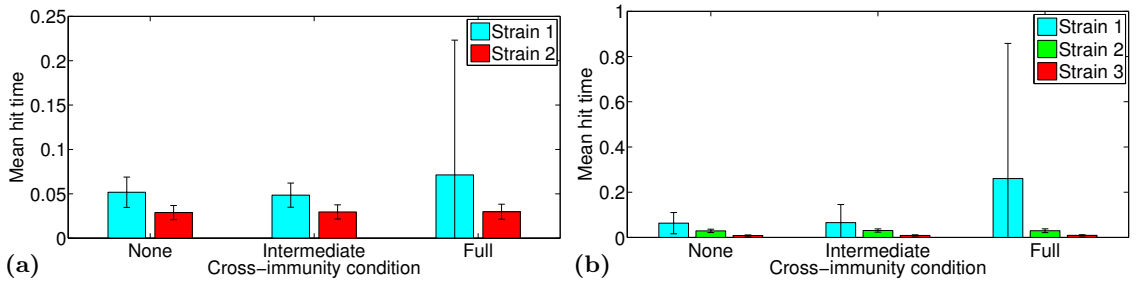

Fig. 10. Mean values for the humanity hit times from the (a) four host, two strain model, (b) four host, three strain model. The error bars give the range of values that were both within 1 standard deviation of the mean and greater than 0.

to reach humanity on average in all three cross-immunity scenarios (Fig. 10b). We note that just like strain 2, the mean and standard deviation values show little variation between the different cross-immunity cases. As in the two strain case, the strains that can only reach humanity via transmission from chickens to pigs and then pigs to humans (i.e. strains 2 and 3) have very similar survival functions, regardless of the type of cross-immunity condition used. The strains that could be transmitted directly from chickens to humans (i.e. strain 1) did see variation in the survival function based on the cross-immunity condition imposed (Fig. 11b). The strain 1 survival curves are similar up until a probability of approximately 0.2, where the curve corresponding to full cross-immunity begins to flatten and decrease with a shallower gradient.

Using a modified four host, two strain model, that had the two strains being introduced at separate times, Figs. 12a-12b present the mean humanity hit times for the strains. Both cases of either strain 1 or strain 2 being introduced first, followed by the remaining strain after the simulation time reached a week were considered. All other aspects of the four host, two strain model remained the same as before. The following comparisons are with respect to the case where both strains were introduced into the duck

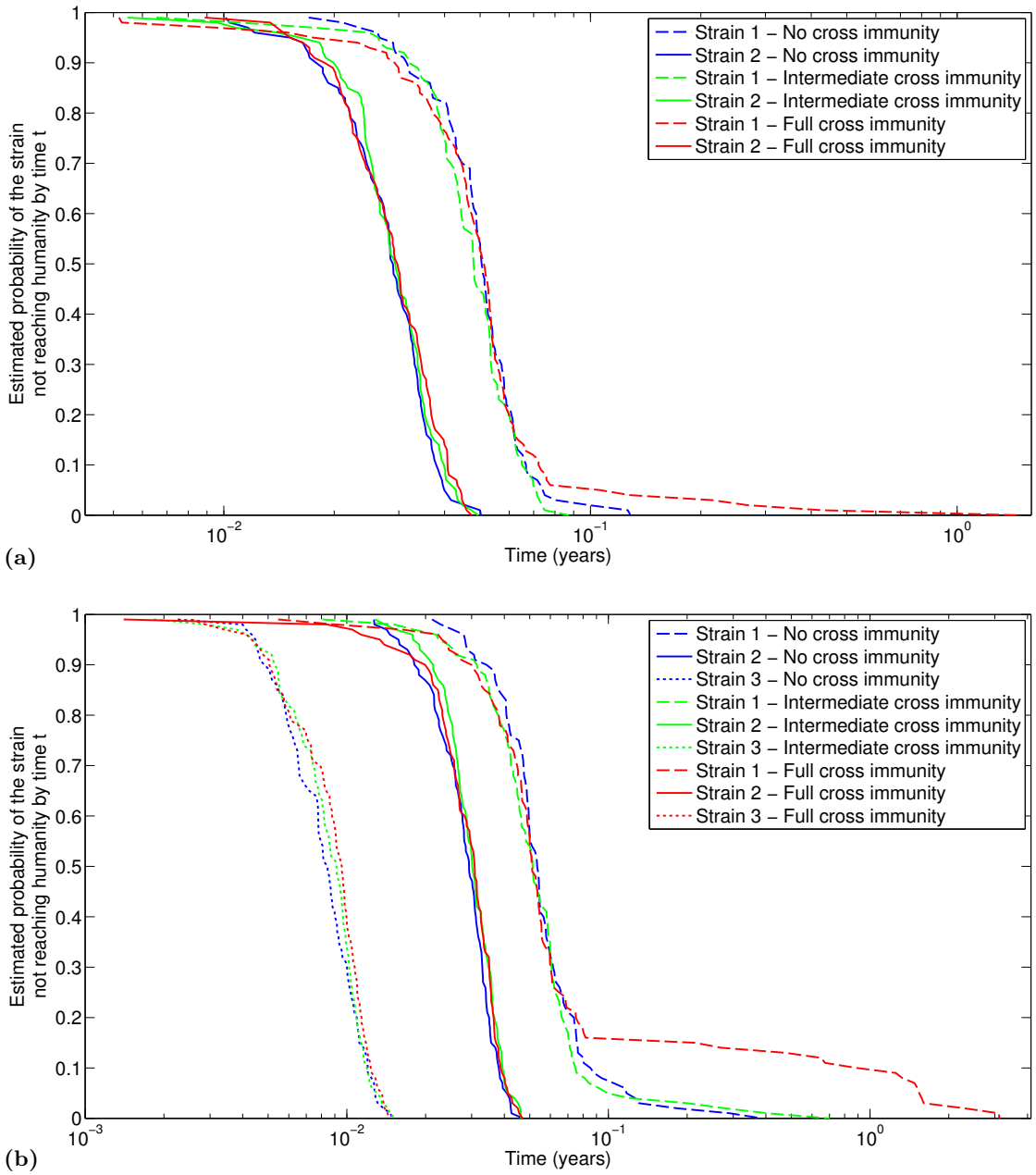

Fig. 11. Survival function of the probability of each strain not reaching humanity by time t for the (a) four host, two strain model and (b) four host, three strain model, with varying cross-immunity parameters. Initial host population sizes were as follows: Ducks, chickens and pigs - 100,000, humans - 10,000. Within each host, the two or three strains had the same transmission and reproductive rates. These were: Ducks, $\beta = 26$, $g = 13$; Chickens, $\beta = 208$, $g = 104$; Pigs and Humans, $\beta = 52$, $g = 26$. In (a), there were initially 5 ducks infected by strain 1, and 5 infected by strain 2. Cross-transmission rates were all zero except: $\beta_{1,2}^{D,C} = 0.01$, $\beta_1^{C,H} = 0.000001$, $\beta_{1,2}^{C,P} = 0.0001$, $\beta_2^{P,H} = 0.001$. In (b), each of the 3 strains initially had 5 distinct ducks infected. Cross-transmission rates were all zero except: $\beta_{1,2,3}^{D,C} = 0.01$, $\beta_1^{C,H} = 0.000001$, $\beta_{1,2}^{C,P} = 0.0001$, $\beta_3^{C,P} = 0.001$, $\beta_2^{P,H} = 0.001$, $\beta_3^{C,P} = 0.001$. The survival function were constructed using humanity hit times obtained from 100 simulation runs.

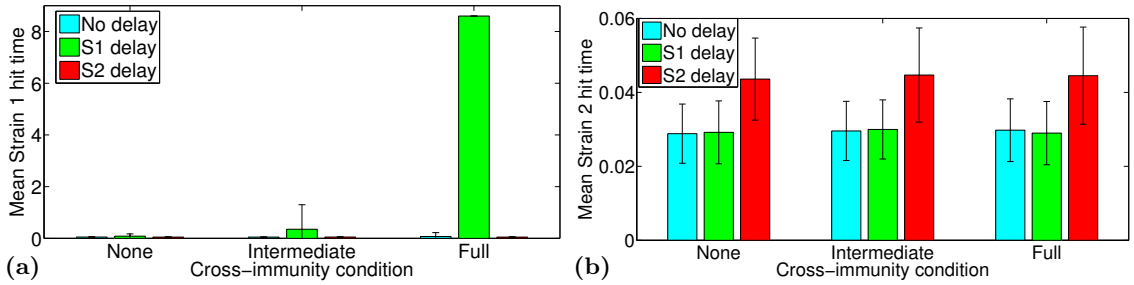

Fig. 12. Mean values for the humanity hit times from the four host, two strain model for (a) strain 1, and (b) strain 2, with a time delay included before the introduction of the second strain. No delay corresponds to strains 1 and 2 having their initial duck infected cases introduced at the same time. S1 delay corresponds to strain 1 having their initial duck infected cases introduced a week after the first strain 2 infected duck cases. S2 delay corresponds to strain 2 having their initial duck infected cases introduced a week after the first strain 1 infected duck cases. The error bars give the range of values that were both within 1 standard deviation of the mean and greater than 0. In the delay model where strain 2 was introduced first with the full cross-immunity condition, strain 1 never reached humanity within the permitted number of events (10 million). The simulation time finished greater than 8.6 years in every run, so this was taken as the value in the plot, though the average had to be greater. As a result, there was no standard deviation value available for that specific example, and no error bar is plotted.

population together at the same time. For the case where strain 1 was introduced before strain 2, strain 1 had a slight reduce in variability in hit time, though only in the full cross-immunity case did the average hit time drop significantly. The average strain 1 hit time seemed unaffected by the cross-immunity condition imposed if strain 1 was introduced before strain 2. Strain 2 had an approximate 50% increase in average hit time in all cross-immunity cases, with a slight increase in variability. Despite this, and being introduced into the system after strain 1, strain 2 still had a lower average hit time in all cross-immunity cases. Strain 2 also once again had the characteristic of the hit time not being affected by the cross-immunity criteria.

This was also seen when strain 2 was introduced before strain 1. The mean and standard deviation values for strain 2 saw little change from the no delay case. In contrast, strain 1 was heavily impacted by being introduced later. With stronger cross-immunity effects strain 1 saw a rapid rise in hit time, and the variability in hit time grew. The effect was so great that in the full cross-immunity scenario strain 1 never reached humanity before the maximum number of events allowed by the simulation, 10 million, was reached. The simulation time finished greater than 8.6 years on each run, so the average hit time had to exceed this value. The mean hit time values and standard deviations used in Figs. 10 and 12 are given in the Appendix.

Fig. 13 gives the survival functions of the probability of each strain not reaching humanity a time t after being introduced, with one strain introduced after a specified delay time, compared to the survival curves obtained for the case with no time delay. The graphs in the left-hand column (Figs. 13a, 13c and 13e) were for the case when strain 1 had a delayed introduction, and the graphs in the right-hand column (Figs. 13b, 13d and 13f) were for the case when strain 2 had a delayed introduction.

First, looking at the strain 1 delay case, the strain 2 survival functions are very similar for each cross-immunity scenario. They also match the survival function obtained for strain 2 in the case when both strains 1 and 2 are introduced at the same time. The strain 1 survival curves had vast differences to the no delay strain 1 survival curves. Each one had a longer tail and a less severe negative gradient, with a higher estimated probability of the strain not having reached humanity for any given time after strain 1 was introduced.

Second, looking at the strain 2 delay case, the strain 1 survival functions are this time similar to each other across each cross-immunity scenario and compared to the no delay case. One discrepancy is in the full cross-immunity case (Fig. 13f), where the strain 1 survival function has a much longer tail. The strain 2 survival function appears shifted compared to the respective no delay survival function. This was caused by the increase in average hit time, but little change in the variability in the hit time from the no delay case to the strain 2 delayed case.

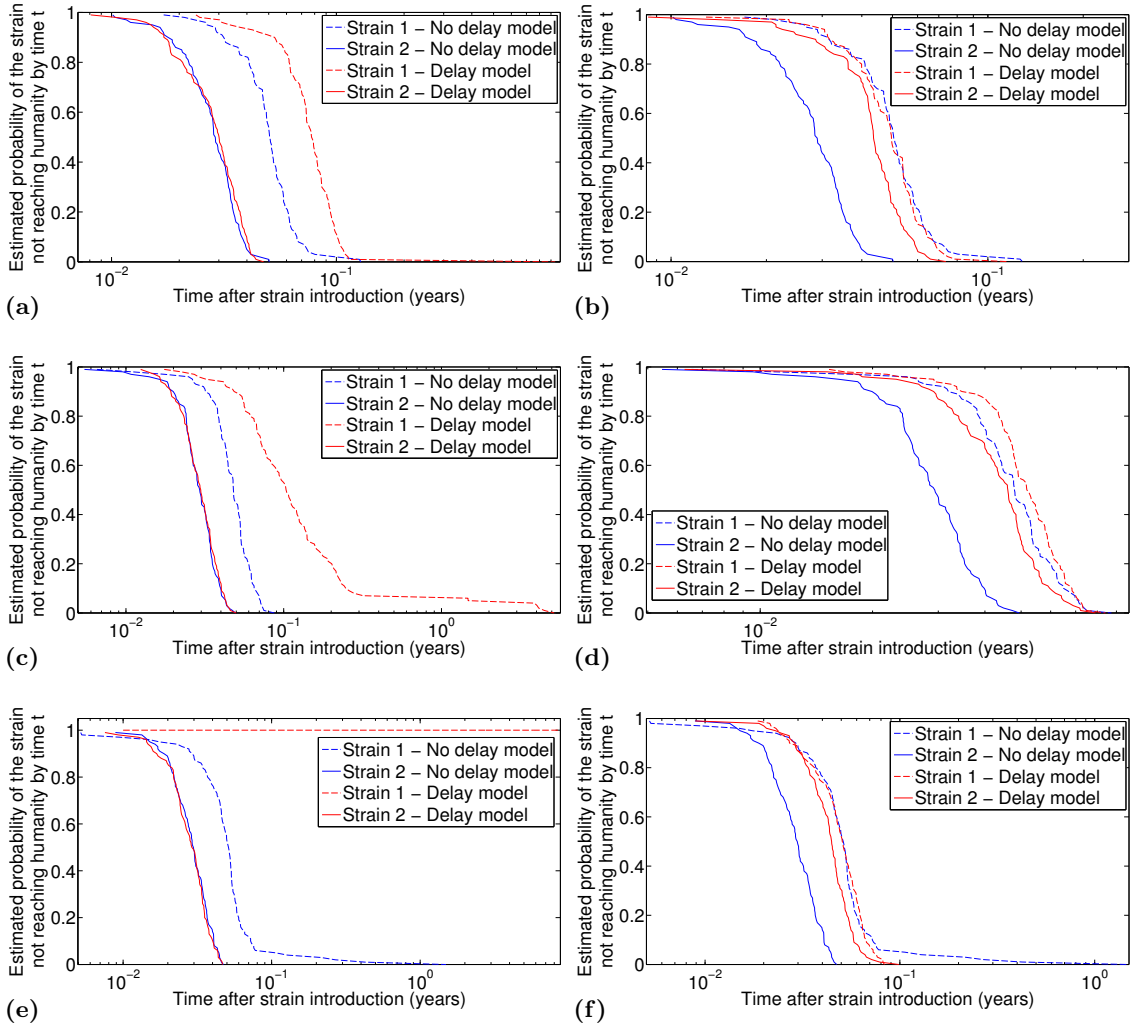

Fig. 13. Survival functions of the probability of each strain not reaching humanity a time t after being introduced for the four host, two strain model, with the following conditions: (a) no cross-immunity, strain 2 introduced before strain 1, (b) no cross-immunity, strain 1 introduced before strain 2, (c) intermediate cross-immunity, strain 2 introduced before strain 1, (d) intermediate cross-immunity, strain 1 introduced before strain 2, (e) full cross-immunity, strain 2 introduced before strain 1, (f) full cross-immunity, strain 1 introduced before strain 2. The survival functions were compared to those obtained when no time delay was included. Initial host population sizes were as follows: Ducks, chickens and pigs - $100,000$, humans - $10,000$. Within each host, the two strains had the same transmission and reproductive rates. These were: Ducks - $\beta = 26$; $g = 13$, Chickens, $\beta = 208$, $g = 104$; Pigs and Humans, $\beta = 52$, $g = 26$. For the graphs in the left-hand column, there were initially 0 ducks infected by strain 1, and 5 infected by strain 2. For the graphs in the right-hand column, there were initially 5 ducks infected by strain 1, and 0 infected by strain 2. After the time since the first strain was introduced exceeded $1/52$, corresponding to a week, 5 ducks were infected by the second strain. Cross-transmission rates were all zero except: $\beta_{1,2}^{D,C} = 0.01$, $\beta_{1}^{C,H} = 0.000001$, $\beta_{1,2}^{C,P} = 0.0001$, $\beta_{2}^{P,H} = 0.001$. The survival functions were constructed using humanity hit times obtained from 100 simulation runs.

4. Discussion

This study presents a model that captures two main features of influenza A; the large number of different strains it consists of, and the ability to infect multiple host species.

One host, multi-strain model

For a number of strains with the same reproductive ratio, R_0, a longer infectious period results in that particular strain dominating. The strain with the longest infectious period has the largest susceptible pool of individuals to infect. This results in an effective reproductive ratio $R_{\text{eff}} = R_0\bar{S}$ that is greater than other strains, where $\bar{S}$ is the number of individuals susceptible to the given strain.

We saw that as the initial host population size increases, the extinction time for each strain grew. It is known that the probability of extinction, for a population of size n, is given by [11]:

$$P_{\text{ext}}(n) = \frac{1}{R_k^n}$$

where $R_k = \frac{R_0 S_k}{N}$ is the effective reproductive ratio for strain k, and S_k is the number of individuals susceptible to strain k. Our findings correspond with this result. With a larger initial host population size, on average that particular strain will persist and spread infection for a longer period of time.

Multi-host, one strain model

The infection number dynamics observed for the two host, one strain model (see Fig. 6) and the three host, one strain model (see Fig. 7) can be explained with similar reasoning used for the one host, multi-strain model. Duck-like influenza dynamics, with lower transmission but a longer infectious period, sees the number of infected settle and have small fluctuations, with the disease able to persist in the population. The low transmissibility means only a small portion of the initial susceptible pool of individuals is infected in the first epidemic. Together with the long recovery period, this allows natural births to replenish the susceptible pool at a rate that keeps up the number becoming no longer susceptible, either through infection or natural death. The result is a high effective reproductive ratio (above 1) with a relatively stable number of infected, and a strain that is able to persist in the host population. Chicken-like influenza dynamics, with higher transmission but a very short infectious period, sees frequent outbreaks in the population occur, with none able to persist. High transmissibility leads to the majority of the susceptible pool of individuals becoming infected at the start of each outbreak. The result is a low effective reproductive ratio (below 1), causing the outbreak to die out. The short infectious period is what causes the overall length of each outbreak to be short-lived.

With regards to the critical value of the cross-transmission parameter in the two host model (Table 1), our result suggests the critical value of the cross-transmission parameter $\beta_{1,2}$ is reduced by the same order the total population size was increased by. An increase in either population size by a certain order of magnitude will result in the number of susceptible chickens and/or infected ducks increasing by a similar order of magnitude (dependent on the host type whose population size was increased). Consequently, the number of cross-transmission events will increase if the cross-transmission parameter $\beta_{1,2}$ remains unchanged. Reducing the value of $\beta_{1,2}$ by a similar order of magnitude to the increase in total population size counteracts this and would make the occurrence of a cross-transmission event relatively rare.

The cross-correlations obtained for the pairs of combinations of host 1 susceptible and infected, and host 2 susceptible and infected time series from Fig. 8a are an expected result (Fig. 8). When an individual in one of the host populations becomes infected, the number of infected increases by one, and the number of susceptibles decreases by one at the same time. The result is the peaks in the number of infected occurring at a similar time to the troughs in the number of susceptibles, though they do not occur at the exact same time due to the existence of a recovered class and the presence of a demographic process. This gives a negative correlation when the time lag is 0 (Figs. 8b-8c). The remaining 4 pairs of combinations for the host 1 susceptible and infected, and host 2 susceptible and infected time series do not directly influence one another, resulting in waveforms that are dissimilar. Therefore, there was no significant cross-correlations

observed for these cases (Figs. 8d-8g). In the two host model, the susceptible and infected number of individuals per host may oscillate with some underlying frequencies (Fig. 9). The exception to this was the host 1 (corresponding to chickens) infected curve. An outbreak could only occur in the host 1 population to begin with if there was a cross-transmission event from host 2 (corresponding to ducks), and there was no immediate infection die out. Due to the process being stochastic, this would lead to outbreaks occurring at irregular intervals, resulting in an uninformative power spectrum as shown in Fig. 9c. Further study is required however to obtain more conclusive results.

Full Model - four hosts, two/three strains

We initially considered the scenario of a group of strains emerging within the duck population at the same time, effectively competing against each other to reach humanity first. Comparing the average hit times (Fig. 10) and the survival functions (Fig. 11), the suggestion is the zoonotic pathway with two cross-transmission steps required, rather than one step that has a lower rate of transmission, appears to be a more efficient way for a strain to reach humanity in the shortest time. This is not impacted by the three cross-immunity conditions that we imposed. Adding in additional strains that reached humanity via the longer zoonotic pathway, via chickens to pigs, and then pigs to humans, appears to further lengthen the average hit time of strains that can be transmitted directly from chickens to humans, in our case strain 1.

The alterations seen in the survival functions when one strain is introduced later than the other, and no cross-immunity occurs, compared to the no delay model are purely created by the assumption that super infection is not possible. The strain that was introduced first had time to establish an infected population in the various host species. Therefore, when the second strain was introduced later on the susceptible pool of individuals that could be infected within each host species was reduced by the number currently infected by the other strain. The average time taken for the strain to cross from one species to the next increased as a result, due to the probability of a cross-transmission event reducing relative to other event rates. This is what caused the average hit time to increase in the strain that was introduced last in comparison to when both strains cause their first infected cases at the same time (see Figs. 13a-13b).

In the case of strain 1 being introduced last, the effect caused by not allowing super infection was further compounded when intermediate cross-immunity or full cross-immunity was assumed. This is captured by the strain 1 delay model survival functions in Figs. 13c and 13e. Now individuals infected by strain 2 would potentially or definitely always gain immunity to strain 1. This would result in a permanently reduced strain 1 susceptible pool. The rate of cross-transmission of strain 1 from ducks to chickens, and then chickens to humans would reduce relative to the other event rates. As a consequence the average hit time will grow, and the variability in hit time will also increase.

When strain 2 was introduced last, the type of cross-immunity did not seem to influence the average hit time. The survival functions in all three cases were also similar (see Figs. 13b, 13d and 13f). One suggestion as to why this occurred is the time delay corresponding to one week was not enough of a head start for strain 1 to establish a large enough infected population within each host species, and not reducing the susceptible pool enough for strain 2 as a result. The cross-transmission parameters for strain 2 were high enough for the rate of a cross-transmission event occurring in each step of the strain 2 pathway to humanity to not be significantly impacted by the slightly reduced susceptible pool of individuals.

There are currently several strains circulating in the human population, such as H5N1, H7N9 and H1N1. If a new strain of influenza were to emerge in humanity, we would want to know the characteristics that would make it capable of causing an epidemic. Our delay model simulations provide us an insight into those characteristics. Those strains would need to have significant antigenic difference to the strains already present, to lower the potential cross-immunity to the strain, and be able to reach humanity via the quickest zoonotic pathway. Further work is needed to establish what is required for the strains to be able to persist in the human population, following the initial infection event.

Further Work

A number of extensions to this work can be pursued. The assumption of no super infection can be relaxed, allowing an individual to be infected by more than one strain at once. How this changes the infection dynamics observed can be explored. In reality strains are continually mutating, creating new

strains that no individual is immune to. We currently do not capture this behaviour. We would require the use of real life data to get accurate parameter values. The four host model can be extended to cope with populations that are not randomly mixing and that have some form of spatial structure. Metapopulation techniques can be used to model communities or patches, all containing a population of each host species. The effect of cross-community transmission rates on the total number infected by each strain could then be explored. For this extended community model, we would expect the initial condition for infected individuals to be more influential on the dynamics observed. This provides another potential direction for further study.

References

[1] R. Jennings and R. C. Read. *Influenza:Human and Avian*. Royal Society of Medicine Press, 2006.

[2] Centers for Disease Control and Prevention. *The 2009 H1N1 Pandemic: Summary Highlights, April 2009-April 2010* . Available at: `http://www.cdc.gov/h1n1flu/cdcresponse.html`, 2010. [Accessed: 24 April 2013].

[3] World Health Organization. *Pandemic (H1N1) 2009 - update 112* . Available at: `http://www.who.int/csr/don/2010_08_06/en/index.html`, 2010. [Accessed: 24 April 2013].

[4] A. Mastin, P. Alarcon, D. Pfeiffer, J. Wood, S. Williamson, I. Brown, and B. Wieland. Prevalence and risk factors for swine influenza virus infection in the English pig population. *PLoS Currents Influenza*, **3** :RRN1209, 2011.

[5] L. Opatowski, C. Fraser, J. Griffin, E. de Silva, M. D. V. Kerkhove, E. J. Lyons, S. Cauchemez, and N. M. Ferguson. Transmission Characteristics of the 2009 H1N1 Influenza Pandemic: Comparison of 8 Southern Hemisphere Countries. *PLoS Pathogens*, **7** (9):e1002225, 2011.

[6] World Health Organization. *Frequently Asked Questions on human infection with influenza A(H7N9) virus, China* . Available at: `http://www.who.int/influenza/human_animal_interface/faq_H7N9/en/index.html`, 2013. [Accessed: 24 April 2013].

[7] World Health Organization. *Human infection with avian influenza A(H7N9) virus in China - update* . Available at: `http://www.who.int/csr/don/2013_05_17/en/index.html`, 2013. [Accessed: 28 May 2013].

[8] J. R. Gog and B. T. Grenfell. Dynamics and selection of many-strain pathogens. *PNAS*, **99** (26):17209–17214, 2002.

[9] N. Arinaminpathy and A. R. McLean. Evolution and emergence of novel human infections. *Proc. R. Soc. B*, **276** : 3937–3943, 2009.

[10] D. J. Smith, A. S. Lapedes, J. C. de Jong, T. M. Bestebroer, G. F. Rimmelzwaan, A. D. M. E. Osterhaus, and R. A. M. Fouchier. Mapping the Antigenic and Genetic Evolution of Influenza Virus. *Science*, **305** :371–376, 2004.

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

[12] R. J. S. Magalhães, X. Zhou, B. Jia, F. Guo, D. U. Pfeiffer, and V. Martin. Live Poultry Trade in Southern China Provinces and HPAIV H5N1 Infection in Humans and Poultry: The Role of Chinese New Year Festivities. *PLoS ONE*, **7** (11):e49712, 2012.

[13] V. Martin, D. U. Pfeiffer, X. Zhoi, X. Xiao, D. J. Prosser, F. Guo, and M. Gilbert. Spaital Distribution and Risk factors of Highly Pathogenic Avian Influenza (HPAI) H5N1 in China. *PLoS Pathogens*, **7** (3):e1001308, 2011.

[14] M. Gilbert, X. Xiao, P. Chaitaweesup, W. Kalpravidh, S. Premashthira, S. Boles, and J. Slingenbergh. Avian influenza, domestic ducks and rice agriculture in Thailand. *Agriculter, Ecosystems and Environment*, **119** :409–415, 2007.

[15] P. G. T. Walker, S. Cauchemez, R. Métras, D. H. Dung, D. Pfeiffer, and A. C. Gnahi. A Bayesian Approach to Quantifying the Effects of Mass Poultry Vaccination upon the Spatial and Temporal Dynamics of H5N1 in Northern Vietnam. *PLoS Computational Biology*, **6** (2):e1000683, 2010.

[16] L. Loth, M. Gilbert, M. G. Osmani, A. M. Kalam, and X. Xiao. Risk factors and clusters of Highly Pathogenic Avian Influenza H5N1 outbreaks in Bangladesh. *Preventive Veterinary Medicine*, **96** :104–113, 2010.

Appendices

The following tables give the mean and standard deviation values for strain hit times in the four host, two strain and four host, three strain models.

Table .2. Mean and standard deviation values for the humanity hit times from the four host, two strain model. All values are given to 3 significant figures.

	Mean humanity hit time (yrs)		Hit time standard deviation (yrs)	
Cross-immunity condition	Strain 1	Strain 2	Strain 1	Strain 2
None	0.0518	0.0288	0.0171	0.00800
Intermediate	0.0485	0.0296	0.0136	0.00800
Full	0.0713	0.0298	0.152	0.00848

Table .3. Mean and standard deviation values for the humanity hit times from the four host, three strain model. All values are given to 3 significant figures.

	Mean humanity hit time (yrs)			Hit time standard deviation (yrs)		
Cross-immunity condition	Strain 1	Strain 2	Strain 3	Strain 1	Strain 2	Strain 3
None	0.0633	0.0290	0.00846	0.0470	0.00735	0.00269
Intermediate	0.0655	0.0306	0.00880	0.0800	0.00723	0.00266
Full	0.260	0.0297	0.00910	0.598	0.00834	0.00277

Table .4. Mean values for the humanity hit times from the four host, two strain model, with a time delay included before the introduction of the second strain. All values are given to 3 significant figures. No delay corresponds to strains 1 and 2 having their initial duck infected cases introduced at the same time. S1 delay corresponds to strain 1 having their initial duck infected cases introduced a week after the first strain 2 infected duck cases. S2 delay corresponds to strain 2 having their initial duck infected cases introduced a week after the first strain 1 infected duck cases.

	Strain 1 mean hit time (yrs)			Strain 2 mean hit time (yrs)		
Cross-immunity condition	No delay	S1 delay	S2 delay	No delay	S1 delay	S2 delay
None	0.0518	0.0852	0.0498	0.0288	0.0292	0.0436
Intermediate	0.0485	0.353	0.0522	0.0296	0.0300	0.0447
Full	0.0713	> 8.6	0.0501	0.0298	0.0300	0.0445

Table .5. Standard deviation values for the humanity hit times from the four host, two strain model, with a time delay included before the introduction of the second strain. All values are given to 3 significant figures. In the delay model where strain 2 was introduced first, strain 1 never reached humanity within the permitted number of events (10 million). As a result, there was no standard deviation value available for that specific example. No delay corresponds to strains 1 and 2 having their initial duck infected cases introduced at the same time. S1 delay corresponds to strain 1 having their initial duck infected cases introduced a week after the first strain 2 infected duck cases. S2 delay corresponds to strain 2 having their initial duck infected cases introduced a week after the first strain 1 infected duck cases.

	Strain 1 hit time standard dev.(yrs)			Strain 2 hit time standard dev.(yrs)		
Cross-immunity condition	No delay	S1 delay	S2 delay	No delay	S1 delay	S2 delay
None	0.0171	0.0884	0.0138	0.00800	0.00850	0.0111
Intermediate	0.0136	0.947	0.0127	0.00800	0.00800	0.0127
Full	0.152	-	0.0149	0.00848	0.00855	0.0131