Journal Article

Dynamics of the 2004 avian influenza H5N1 outbreak in Thailand: The role of duck farming, sequential model fitting and control

Author(s):
Retkute, Renata; Jewell, Chris P.; Van Boeckel, Thomas P.; Zhang, Geli; Xiao, Xiangming; Thanapongtharm, Weerapong; Keeling, Matt; Gilbert, Marius; Tildesley, Michael J.

Publication Date:
2018-11-01

Permanent Link:
https://doi.org/10.3929/ethz-b-000295721

Originally published in:

Rights / License:
Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International

This page was generated automatically upon download from the ETH Zurich Research Collection. For more information please consult the Terms of use.
Dynamics of the 2004 avian influenza H5N1 outbreak in Thailand: The role of duck farming, sequential model fitting and control

Renata Retkute⁎, Chris P. Jewell, Thomas P. Van Boeckel, Geli Zhang, Xiangming Xiao, Weerapong Thanapongtharm, Matt Keeling, Marius Gilbert, Michael J. Tildesley

⁎ Corresponding author.
E-mail address: r.retkute@warwick.ac.uk (R. Retkute).

ARTICLE INFO

Keywords:
Avian influenza
Disease control
Spatial heterogeneity
Vaccination

ABSTRACT

The Highly Pathogenic Avian Influenza (HPAI) subtype H5N1 virus persists in many countries and has been circulating in poultry, wild birds. In addition, the virus has emerged in other species and frequent zoonotic spillover events indicate that there remains a significant risk to human health. It is crucial to understand the dynamics of the disease in the poultry industry to develop a more comprehensive knowledge of the risks of transmission and to establish a better distribution of resources when implementing control. In this paper, we develop a set of mathematical models that simulate the spread of HPAI H5N1 in the poultry industry in Thailand, utilising data from the 2004 epidemic. The model that incorporates the intensity of duck farming when assessing transmission risk provides the best fit to the spatiotemporal characteristics of the observed outbreak, implying that intensive duck farming drives transmission of HPAI in Thailand. We also extend our models using a sequential model fitting approach to explore the ability of the models to be used in “real time” during novel disease outbreaks. We conclude that, whilst predictions of epidemic size are estimated poorly in the early stages of disease outbreaks, the model can infer the preferred control policy that should be deployed to minimise the impact of the disease.

1. Introduction

Since the emergence of highly pathogenic avian influenza (HPAI) H5N1 in the late 1990s, the virus has had significant impact on poultry industries around the world and has posed a serious threat to public health. The majority of outbreaks occur in East and South Asia, a region that contains half of the population of the world. Whilst the probability of a human avian influenza (AI) pandemic may appear low based on the poor ability of the virus to adapt to the upper human airway (Peiris et al., 2007), the death toll of such an event may be catastrophic. As a result of the HPAI H5N1 panzootic, there have been 860 confirmed human cases of avian influenza A (H5N1) in 16 countries and 454 deaths (a human case fatality of 53%) as of 28th May, 2018, based on World Health Organization (WHO) statistics. The majority of humans infected with the virus work in professions that involve close contact with potentially infected poultry (de Bruin et al., 2017) and therefore it is crucial to understand the dynamics of the disease in the poultry industry to develop a more comprehensive knowledge of the risks of transmission to humans.

In Thailand, H5N1 was first detected in the poultry industry in January 2004 (Tiensin et al., 2005). The first wave of infection took place from January to May and resulted in 193 reported AI outbreaks. A second infection wave started in July 2004 and culminated in March 2005, with 1,492 outbreaks notified during this period. Approximately 62 million birds were killed either through infection or through targeted culling as a control measure. In addition, during the 2004 H5N1 epidemic there were 17 human cases of infection and 12 deaths (Tiensin et al., 2005; Auewarakul, 2008). In an attempt to improve case detection, the Thai government implemented an X-ray survey (Gilbert et al., 2006) in September 2004. Control policies such as localised movement restrictions and 1km ring culling were introduced to reduce the risk of further spread of the disease; vaccination was not used.
The spatial distribution of HPAI H5N1 has been studied in several countries and regions using spatial statistical models (Gilbert et al., 2006, 2007, 2008, 2010, 2014; Tiensin et al., 2007a,b; 2009; Paul et al., 2010, 2014, 2016; Stegeman et al., 2011; Sematimba et al., 2012; Artois et al., 2017) and mathematical models (Truscott et al., 2007; Tiensin et al., 2007a,b; Boender et al., 2007; Sharkey et al., 2008; Jewell et al., 2009a,b,c; Minh et al., 2011; Walker et al., 2012; Hill et al., 2017, 2018). Evidence from previous work suggests that different landscapes, production systems and water-related variables are risk factors that will promote disease transmission. In addition, mathematical models such as the one that we present here can improve understanding about the transmission dynamics of infectious diseases and help with an assessment of the effectiveness of control strategies applied during outbreaks (Stegeman et al., 2011). A key challenge when implementing control is the lack of clinical signs in many duck species - whilst chickens have a very high mortality rate for H5N1 (Yu et al., 2007; Jeong et al., 2009), ducks are largely asymptomatic but can readily transmit the disease. In addition, in Thailand, as in many other countries in South East Asia, the presence of free grazing ducks, that feed year round on rice paddies, have been shown to be a strong predictor for the presence of the disease in the landscape (Gilbert et al., 2007, 2008). It is therefore crucial to develop a more detailed understanding of the factors leading to persistence of AI in poultry, the conditions that are most suitable to disease transmission, and intervention strategies that will minimize the future impact of the disease.

In this paper, we develop a set of mathematical models to simulate the spread of HPAI H5N1 in the poultry industry in Thailand. We use detailed demographic and epidemiological data from the 2004 epidemic and fit our models to epidemiological surveillance (X-Ray) data. Our framework will allow us to establish the risk factors that result in transmission of the disease that will in turn enable us to effectively target intervention strategies. We will also consider the ability of the models to be used in “real time”, whereby model parameters are obtained by only using data that are observable at a given stage of the epidemic. This enables us to determine the potential for HPAI models to be used during outbreak both in terms of establishing transmission risk and in determining the most appropriate intervention policy. It also allows us to quantify how progressive accumulation of information shapes the knowledge of outbreak dynamics and the robustness of control strategies.

2. Materials and methods

2.1. Data

In this paper, we use demographic data compiled by the Department of Livestock Development (DLD, Bangkok) that was constructed through the X-Ray surveys that were implemented in response to the H5N1 outbreak in 2004. Demographic data was recorded at the farm/owner level and consists of the number of chicken and ducks in each flock, and a unique identifier of the subdistrict where each farm is located.

Owing to imperfect reporting at the onset of the epidemic, data for the first wave of infection in Thailand are largely incomplete. Therefore, for parameter estimation and analysis, we use the outbreak data from the second wave of the epidemic that took place from July 2004. In this second wave, more than 1400 outbreaks were reported and over 62 million birds died or were culled as part of the control policy (Tiensin et al., 2005). The outbreak data set includes the unique identifier of the subdistrict as well as the total number of sick, dead and destroyed animals.

In order to implement our mathematical model, we use geographical data for each subdistrict including the locations of subdistrict boundaries and the area of each subdistrict. We infer indices of neighbouring subdistricts by analysing intersection sets of subdistrict boundaries. We then calculate the length of all shared boundaries by fitting a linear spline through shared points on boundaries of neighbour subdistricts. The proportion of area used by rice paddy fields within each subdistrict was derived from the moderate resolution imaging spectroradiometer (MODIS) sensor onboard the NASA Terra satellite (Xiao et al., 2006).

2.2. Mathematical models

In our model, the basic unit of infection is the flock, and we assume that all birds (chicken and ducks) within a flock become infected. The model is stochastic time-discrete (1 day time-step). At any time, t, flocks can be partitioned into four classes: susceptible, exposed, infectious, or removed. The total number of infectious flocks at time t are defined as \( n_i(t) \). The probability that a susceptible flocks \( k \) is infected at time \( t \) is given by

\[
p_k(t) = 1 - \exp(-\lambda_k(t)),
\]

where \( \lambda_k(t) \) is the force of infection on flock \( k \) at time \( t \).

We develop a set of mathematical models of varying complexity depending on environmental and epidemiological factors with the force of infection having the following form:

\[
\lambda_k(t) = \frac{(1 + e_0) (\delta (1 + n_i(t)) + S_b(t) T_k(t))}{\text{environmental factors} \times \text{epidemiological factors}},
\]

Here \( e_0(t) \) is an environmental factor, which accounts for the effect of the spatial location, \( \delta \) is the background term, \( S_b(t) \) is the number of susceptible flocks \( k \) and \( T_k(t) \) is the transmission rate owing to the presence of infectious flocks within the subdistrict or in the neighbouring subdistrict. We have included \( 1 + n_i(t) \) in the model in order to allow the probability of spontaneous infections even in the absence of infectious flocks. This situation can occur for example due to the presence of a wildlife reservoir or the import of infected poultry from other countries. The spatial heterogeneity is implied in the term \( e_0(t) \), which depends on the environmental attributes of the subdistrict that flock \( k \) belongs to. We use a sigmoidal function to account for environmental factors

\[
H(x, \epsilon, n) = \frac{1}{(x/\epsilon)^n + 1},
\]

where \( x \) describes the value of the environmental factor at the geographical unit to which the susceptible flocks belongs. We develop a meta-population framework for the term \( T_k(t) \) following the approach of Buhnerkempe et al. (2014).

In our modelling framework, we consider transmission via one of five mechanisms: (i) spatially independent transmission to any subdistrict in Thailand, (ii) transmission within a subdistrict, (iii) transmission across borders to neighbouring subdistricts, (iv) transmission driven by presence of rice paddies and (v) transmission driven by duck farm intensity. By varying the combination of these factors, we have formulated five mathematical models with increasing complexity: a random process model (Model A), a spatial model (Model B), a spatial model incorporating rice density (Model C), a spatial model incorporating duck intensity (Model D), and a spatial model incorporating both rice density and duck intensity (Model E). Details of the equations governing each model are given below and in the Supplementary Information. List of parameters is shown in Table 1.

2.2.1. Model A: Random process model

The simplest model for a HPAI outbreak is that of complete spatial randomness, whereby transmission events are distributed independently according to a uniform probability distribution over all susceptible flocks. Under this scenario, the infection pressure at time \( t \) in any susceptible flock is:

\[
\lambda_k(t) = \delta (1 + n_i(t)).
\]

The biological meaning of the model is that the outbreak dynamics...
does not depend on any environmental factors or the structure of the poultry sector in Thailand. This naive model, whilst unlikely to be representative of heterogenous transmission routes, provides a baseline for epidemic spread for comparison with the spatially explicit models.

2.2.2. Model B: Spatial model

The spatial model utilises a classical metapopulation approach, based upon the model first used by Buhnerkempe et al. (2014). This model is an extension of the random process model, with the addition of within and between subdistrict transmission. The model therefore incorporates local density-dependent spread and contains three additive terms representing the different transmission scenarios - the background term, $\delta$, within-subdistrict transmission with transmission rate $\beta w$, and local cross-border transmission with transmission rate $\beta b$ from the set of neighbour subdistricts $NB_k$. The framework assumes that transmission is dependent upon the poultry industry features within a given subdistrict and within neighbouring subdistricts and ignores the presence of other landscape features that may result in increased transmission risk. Under this scenario, the transmission within or between subdistricts at time $t$ is:

$$ T_k(t) = \beta w W_{li}(t) + \beta b \sum_{i \in NB_k} B_{ki} l_i(t), $$

where $l_i(t)$ and $L_k(t)$ represent the number of infectious flocks within the subdistrict $i$ and in the neighbour subdistrict $k$ respectively. Details of the analytical expression of $W_{li}$ and $B_{ki}$, where both terms account for a probability of contact between premises within and between subdistricts. Details of the equations governing each term are given in the Supplementary Information.

2.2.3. Model C: Spatial (rice) model

Given that previous work suggests that the presence of free grazing ducks can result in increased transmission risk in Thailand (Gilbert et al., 2007), here we use data on rice fields in the country (Xiao et al., 2006) and extend the meta-population model described in the spatial model above to consider the likelihood of increased transmission risk in subdistricts with a high density of rice paddies. In order to account for the role of rice paddies in the infection dynamics, we use a sigmoidal function that operates as a scaling factor for the infection pressure. We define $r_{ck}$ as the fraction of the area of the subdistrict occupied by rice paddies. Here a higher infection pressure is presumed to be present for subdistricts with a higher proportion of rice paddy fields. This dependence is described by the scaling, $a_R$, threshold, $c_R$, and power, $n_R$. Therefore, we have that $e_k = a_R H(r_{ck}, c_R, n_R)$.

2.2.4. Model D: Spatial (duck) model

This model builds upon the spatial model but additionally considers the intensity of duck farming within a subdistrict as an additional factor influencing transmission. The intensity of the duck industry, $d_{ck}$, is formalised as a $\log_{10}$ of the size of largest duck flock within a subdistrict plus one. This dependence is described by the scaling, $a_D$, threshold, $c_D$ and power, $n_D$ given that the environmental factor is $e_k = a_D H(d_{ck}, c_D, n_D)$.

2.2.5. Model E: Spatial (duck and rice) model

Our final model incorporates dependence on all factors, i.e. spatial structure, rice density and intensity of duck farming. Here, $e_k = a_D H(r_{ck}, c_R, n_R) + a_D H(d_{ck}, c_D, n_D)$.

2.3. Fitting the models to outbreak data

Having defined the full modelling framework, we will consider this suite of nested mathematical models (A–E) to determine the factors that contribute to transmission of HPAI H5N1 within Thailand. Each model is fitted to outbreak data using a Bayesian likelihood approach that determines the parameters that best captures the observations (for details see Supplementary Material section below).

To establish the ability of each model to capture the spatiotemporal characteristics of the observed outbreak, we have simulated our models using the posterior distribution for parameters. We used these simulations to determine the ability of each model at capturing both the temporal and the spatial epidemic profiles from the 2004 outbreak. This provides an indication of whether our modelling frameworks are able to mimic the fine-scale transmission dynamics between poultry farms in the country. To assess goodness-of-fit of model predictions, we inspected the temporal and spatial profiles to ensure our simulations produced similar patterns to the observed outbreak. We quantify spatial similarity between observed and predicted maps by calculating the Jaccard index given

$$ J(\text{obs}, \text{sim}) = \frac{\sum_k \min(\text{obs}_k, \text{sim}_k)}{\sum_k \max(\text{obs}_k, \text{sim}_k)}, $$

where $\text{obs}_k$ and $\text{sim}_k$ are the observed and simulated number of infected flocks in subdistrict $k$. Here, the higher the value of $J$, the better the fit to the spatial epidemic. This index gives us a measure of the spatial accuracy of our model predictions at the subdistrict level.

2.4. Sequential analysis

The model parameterisation that we have described in the section above determines parameter values by fitting the set of models to the entire observed spatiotemporal outbreak in Thailand. This gives us an understanding of the level of complexity required to capture HPAI epidemic dynamics. It is also important to explore the potential of the model, so that we can assess whether the frameworks can be used to provide policy advice for ongoing epidemics. For a sequential analysis, we re-estimated model parameters by incrementally increasing the length of the outbreak data, evaluating how increasing amounts of outbreak data altered and influenced the parameter estimation and model projections.

In order to investigate the utility of our model predictions and control options in real time during outbreaks, we fitted the model only using outbreak data that would be available at particular points during the epidemic. The model was therefore parameterised using reported data up to day 10 (183 outbreaks) and then sequentially on days 15 (269 outbreaks), 20 (439 outbreaks), 25 (593 outbreaks), 30 (697 outbreaks), 35 (815 outbreaks), 50 (1170 outbreaks), 75 (1295 outbreaks), 100 (1324 outbreaks), 125 (1388 outbreaks), 150 (1430 outbreaks), 175 (1433 outbreaks), and 200 (the entire outbreak, 1435 outbreaks).

<table>
<thead>
<tr>
<th>Table 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>List of parameters.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description (units)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\delta$</td>
<td>Background term</td>
</tr>
<tr>
<td>$n(t)$</td>
<td>Number of infectious flocks</td>
</tr>
<tr>
<td>$l_{i}(t)$</td>
<td>Number of susceptible flocks in a district $i$ at time $t$</td>
</tr>
<tr>
<td>$\beta w$</td>
<td>Transmission rate within the subdistrict ($(flock)^{-1} (day)^{-1}$)</td>
</tr>
<tr>
<td>$\beta b$</td>
<td>Transmission rate from the neighbour subdistrict ($(flock)^{-1} (day)^{-1}$)</td>
</tr>
<tr>
<td>$\rho c$</td>
<td>Power coefficient for chicken</td>
</tr>
<tr>
<td>$\rho d$</td>
<td>Power coefficient for ducks</td>
</tr>
<tr>
<td>$\zeta$</td>
<td>Multiplier for relative infectiousness and transmissibility of ducks to chicken</td>
</tr>
<tr>
<td>$r_{ck}$</td>
<td>Fraction of area occupied by rice paddies</td>
</tr>
<tr>
<td>$a_R$</td>
<td>Scaling coefficient for rice density</td>
</tr>
<tr>
<td>$c_R$</td>
<td>Threshold for rice density</td>
</tr>
<tr>
<td>$n_R$</td>
<td>Power coefficient for rice density</td>
</tr>
<tr>
<td>$d_{ck}$</td>
<td>Intensity of duck industry</td>
</tr>
<tr>
<td>$a_D$</td>
<td>Scaling coefficient for intensity of duck farming</td>
</tr>
<tr>
<td>$c_D$</td>
<td>Threshold for intensity of duck farming</td>
</tr>
<tr>
<td>$n_D$</td>
<td>Power coefficient for intensity of duck farming</td>
</tr>
</tbody>
</table>
2.5. Control options

Once fully parameterised, we will use the models to investigate the effectiveness of a range of interventions that could be implemented to control the outbreak. We will analyse outbreak projections under three competing strategies: (i) culling of identified infected premises (IP) only, (ii) culling of IPs and farms within a given radius of IPs (ring culling) and (iii) IP culling and reactive vaccination with a given capacity within infected subdistricts.

The spatially aggregate nature of the meta-population model that we have developed prevents us from applying a culling radius directly to each IP to determine which flocks will be culled. In order to mimic ring culling in our metapopulation framework we therefore reduce the number of susceptible flocks within an infected subdistrict as a fraction of the subdistrict area to the area of a circle with a particular culling radius. We analyse projections based on the following culling radius values: 0.25 km, 0.5 km, 1 km, 2 km and 5 km.

As an alternative to ring culling, we explore the impact of reactive vaccination within any subdistrict reporting infection. For this control policy, following the removal of infected flocks, we assume that a given percentage of the poultry within an infected subdistrict are vaccinated and become immune to infection. Challenge trials show that vaccines can induce immunity (Qiao et al., 2003), reduce H5N1 excretion (van der Goot et al., 2008) or reduced morbidity and mortality to zero (EFSA, 2007) as soon as 7 days after vaccination. Therefore, we make the assumption that any poultry that are targeted for vaccination will acquire immunity 7 days after the relevant infected premises is reported. We then explore the impact of vaccination upon the outbreak size and duration, given vaccine capacities of 50%, 70% and 90% of all poultry within each infected subdistrict. See Supplementary Information for more details. As optimal control actions for control of disease outbreaks may depend on objectives (Probert et al., 2016), we evaluated the effectiveness of local control by analysing the projected number of infected flocks, the number of culled flocks and the duration of the outbreak. Real-time decision-making capability was tested by evaluating interventions based on the models fitted using outbreak data throughout the course of an epidemic (Probert et al., 2018).

2.6. Assessment of the economic impact of control strategies

Economic losses incurred by epidemics can be broadly divided into two categories: direct losses and consequential losses (Meuwissen et al., 1999). Direct losses refer to the costs of the execution of the control strategy, for example, compensation paid for destroyed animals or the cost of vaccines. Consequential losses are the long-term consequences, for example, due to supply and delivery problems or zoonotic transmission. Any intervention will have a combination of both direct and consequential losses.

The goal is to find a control which minimises the following cost function:

\[ C(c_1, c_2, c_3) = c_1 p_{\text{cull}} + c_2 p_{\text{vac}} + c_3 p_{\text{inf}}, \]

where \( p_{\text{cull}} \) is the mean expected number of culled flocks, \( c_1 \) is the cost associated with culling of a single flock, \( p_{\text{vac}} \) is the mean expected number of vaccinated poultry, \( c_2 \) is the cost associated with vaccination of one hundred birds, \( p_{\text{inf}} \) is the mean expected number of infected flocks, and \( c_3 \) is the cost associated with the monetary impact of a single infected flock (e.g. the risk of zoonotic transmission to professions that involve close contact with potentially infected poultry or wild-bird transmission). In the results section we investigate the optimal policy that should be implemented as the relative costs of \( c_1, c_2 \) and \( c_3 \) vary.

3. Results

3.1. Model parameterisation

We used a Bayesian framework to parameterise each model to the 2004 epidemics of H5N1 in Thailand. In order to compare the five modelling frameworks and establish a goodness of fit to the observational data, we calculate the deviance information criterion (DIC). Models with a lower DIC value implies a closer fit to the observational data compared to the alternative models and therefore this model is preferred. For DIC calculations we use the parameter values generated from our MCMC runs for parameter estimation (Spiegelhalter et al., 2002). We calculated the following values for the four models: \( DIC_A = 37478.2, DIC_B = 33763.5, DIC_C = 35396.7, DIC_D = 33164.3 \) and \( DIC_E = 33208.3 \). The main decrease in DIC comes from taking into account the spatial transmission component of the model (i.e. model B vs A). In this analysis, we find that the Jaccard index takes the following values for each model: \( J_A = 0.033, J_B = 0.043, J_C = 0.088, J_D = 0.112, \) and \( J_E = 0.108. \) The DIC and Jaccard index values indicate that Model D, incorporating the intensity of the duck industry, is preferred over all other modelling frameworks, indicating that it is the intensive duck farming industry that is predominantly driving transmission in Thailand.

For all models, the fitted posterior mean and 95% credible intervals for estimated parameters are summarised in Table 2. As the complexity of the modelling framework increases, the value of the background term \( \delta \) decreases, as local scale characteristics drive transmission for the more complex models. When considering rice paddy density as a risk factor (Model C) similar parameter estimates are found for the background term and within-subdistrict transmission rate when compared with the purely spatial Model B. Interestingly Model D, that includes intensive duck farms as a risk factor, produces lower parameter value estimates for \( \delta, \rho_w \) and \( \rho_b \), suggesting that the presence of intensive duck farms may have a strong influence on transmission.

To compare each model’s capacity to reproduce observed temporal and spatial summary statistics the 2004 avian influenza H5N1 outbreak in Thailand, we have simulated 1000 outbreaks for each model. Model D appears to most accurately capture the spatiotemporal dynamics of

---

Table 2

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Model (A)</th>
<th>Model (B)</th>
<th>Model (C)</th>
<th>Model (D)</th>
<th>Model (E)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \delta \times 10^{-8} )</td>
<td>2.55 (2.43, 2.68)</td>
<td>1.23 (1.14, 1.30)</td>
<td>1.23 (1.16, 1.30)</td>
<td>0.77 (0.69, 0.84)</td>
<td>0.64 (0.52, 0.76)</td>
</tr>
<tr>
<td>( \beta_w \times 10^{-6} )</td>
<td>3.60 (2.85, 4.52)</td>
<td>3.45 (3.22, 3.65)</td>
<td>0.99 (0.76, 1.12)</td>
<td>0.33 (0.24, 0.41)</td>
<td>0.33 (0.24, 0.41)</td>
</tr>
<tr>
<td>( \beta_b \times 10^{-5} )</td>
<td>1.71 (1.38, 2.22)</td>
<td>1.08 (1.03, 1.13)</td>
<td>0.41 (0.34, 0.52)</td>
<td>0.14 (0.10, 0.18)</td>
<td>0.14 (0.10, 0.18)</td>
</tr>
<tr>
<td>( p_C )</td>
<td>0.38 (0.29, 0.47)</td>
<td>0.34 (0.30, 0.38)</td>
<td>0.41 (0.32, 0.48)</td>
<td>0.28 (0.16, 0.37)</td>
<td>0.28 (0.16, 0.37)</td>
</tr>
<tr>
<td>( p_v )</td>
<td>0.26 (0.22, 0.31)</td>
<td>0.21 (0.18, 0.24)</td>
<td>0.03 (0.01, 0.08)</td>
<td>0.04 (0.02, 0.09)</td>
<td>0.04 (0.02, 0.09)</td>
</tr>
<tr>
<td>( \xi )</td>
<td>3.2 (2.3, 3.9)</td>
<td>3.6 (3.4, 3.9)</td>
<td>2.8 (2.3, 3.2)</td>
<td>2.9 (2.4, 3.7)</td>
<td>2.9 (2.4, 3.7)</td>
</tr>
<tr>
<td>( \alpha_s )</td>
<td>0.71 (0.65, 0.74)</td>
<td>0.61 (0.58, 0.63)</td>
<td>3.19 (1.98, 5.51)</td>
<td>0.64 (0.48, 0.88)</td>
<td>0.64 (0.48, 0.88)</td>
</tr>
<tr>
<td>( \alpha_v )</td>
<td>3.2 (3.1, 3.3)</td>
<td>13.5 (11.2, 16.7)</td>
<td>3.4 (3.2, 3.7)</td>
<td>3.9 (3.3, 4.7)</td>
<td>5.9 (5.0, 6.8)</td>
</tr>
<tr>
<td>( n_0 )</td>
<td>16.6 (12.3, 19.4)</td>
<td>16.6 (12.3, 19.4)</td>
<td>3.9 (3.3, 4.7)</td>
<td>5.2 (4.2, 6.4)</td>
<td>5.2 (4.2, 6.4)</td>
</tr>
</tbody>
</table>
the HPAI outbreak in Thailand (Fig. 1, second column). For this model, spatial predictions of spread are much more representative of the true 2004 outbreak, with infected subdistricts predominantly in the centre and west of Thailand, in the region north of Bangkok. In addition, the temporal profile captures both the epidemic peak and the epidemic tail, providing supporting evidence of the significant role of the intensive duck farming industry in transmission of HPAI H5N1 in Thailand.

The spatial model, Model B, does not accurately capture the geographical spread of the outbreak, with significant overestimates of infected farms in the east of the country and underestimates in the main

Fig. 1. Predicted and observed outbreak distributions. Maps show the mean number of infected flocks in each subdistrict.
epidemic hotspots. The model is also unable to accurately capture the temporal profile of the epidemic tail (Fig. 1, third column). We see remarkably similar predictions for Model C (Fig. 1, fourth column), though in this case overpredictions of spread in the east of Thailand are even greater, owing to the high densities of rice paddies in that part of the country. In addition, southern subdistricts of Thailand tend to have a higher average flock density, which will result in Models B and C overpredicting spread in these regions.

The random process model is unable to capture the spatial profile of the true outbreak, given that the model assumes random transmission characteristics. The model performs marginally better at predicting temporal behaviour, though slightly overestimates the epidemic peak and underestimates the number of farms infected in the epidemic tail (Fig. 1, last column).

3.2. Sequential analysis

Here we are interested in exploring how the prediction of the epidemiological parameters change through time. This is particularly valuable as it gives us an indication of the predictive power of our model during the early stages of future outbreaks. Our results are summarised in Fig. 2. As the observation time, \( T \), increases, the estimated value of the background term \( \delta \) decreases. Meanwhile, within subdistrict and between subdistrict transmission rates are both observed to initially increase as \( T \) increases, before decreasing again after the epidemic peak.

When the model is only fitted to the first 10 days on the outbreak, we predict significant spread of the virus over large areas of the country (Fig. 3, second column). The model also dramatically overestimates outbreak size. This overestimate appears to be due to the parameterisation of the model on day 10 - the background term \( \delta \) dominates (Fig. 2, fourth row) and the model predicts much lower within-subdistrict and cross border transmission (Fig. 2, fourth row). However, as the outbreak progresses, the model predicts that transmission is becoming more local, as seen by a reduction in the predictions of the value of \( \delta \). As the epidemic progresses, we see that the model performs much better at predicting the spatiotemporal profile of the outbreak. Indeed by day 30 (Fig. 3, third column), predictions of epidemic spread begin to be confined to regions where the true outbreak was observed. By day 50, the model is capable of producing accurate predictions of both the spatial and temporal profile (Fig. 3, fourth column), indicating that from this stage our modelling framework could be used to generate robust predictions that could inform policy in real time.

3.3. Evaluating the effectiveness of local control

We investigate the predictive ability of control policies for the preferred Model D. We therefore simulate our model from the start of the outbreak, using parameters that we have obtained on days 10, 15, 20, 25, 30, 35, 50, 100 and 200. We then explore the model predictions of the effectiveness of all competing ring culling and vaccination policies using model parameters obtained at each of these intervals. We observe that, even on day 10 when there is significant uncertainty in spatial spread parameters, Model D accurately predicts that 5km ring culling is the policy that minimises the total number of infected flocks (Fig. 4 A, top panel). Whilst the predicted epidemic sizes change for each control policy during the outbreak, the relative ordering of control policies remains consistent throughout. The same is true for total culled flocks, in that high capacity ring culling is always preferred (Fig. 4A, middle panel). However, we see somewhat different results for our preferred spatial (duck) model in this case there is no significant difference between a 5 km ring culling policy and a vaccination strategy with a 90% capacity (Fig. 4A, lower panel, Mann-Whitney test \( p = 0.14 \)). This result is valid when the delay from vaccination to immunity is 7 days. However, if we assume that it takes 14 days to acquire immunity, we find that the number of culled flocks increases by around 9% and therefore ring culling becomes the preferred intervention strategy.

If we are interested in minimising outbreak duration we see a somewhat different result. When the model uses only early outbreak data, predictions of epidemic duration are dramatically overestimated for IP only culling and all vaccination policies (Fig. 4A, lower panel). The dynamics of the simulated outbreak is mostly dominated by the background \( \delta \), so even under a vaccination capacity of 90%, this gives a high enough infection pressure to sustain the outbreak for a long period of time. This overprediction is largely resolved by day 25 when vaccination is used, possibly owing to the more accurate predictions of the spatial extent of the disease at this point in the epidemic.

Livestock epidemics like HPAI H5N1 have significant impact on the poultry industry, causing severe economic losses for governments, farmers and pose a serious threat to public health. We have performed a general economic analysis to explore the economic impact of control strategies by minimising the expected losses defined by Eq. (4). The optimal control strategy depends on the ratio between the relative costs of infection, culling and vaccination, as seen in Fig. 4. Interestingly, if the relative cost of vaccination and culling is much higher than the economic impact of infection, the best strategy is to cull only infected flocks. This strategy is consistent for parameters estimated at 200, 100, 50 or 20 days. This highlights a key tenet, that the optimal control policy is strongly dependent upon the objective of such an intervention.

4. Discussion

We have developed a set of mathematical models to investigate the level of complexity required to make accurate predictions regarding the spread of HPAI H5N1 in Thailand. When developing models, there is always a trade off between including enough model complexity to be able to capture the, often very complex, epidemiological and demographic characteristics that can lead to virus transmission, whilst at the same time keeping the model simple enough such that it is possible to parameterise. Here we explored that premise by considering five nested models of increasing complexity, where we accounted for random (non-distance based) spread, local spread of the disease, increased transmission owing to rice paddy density and increased transmission owing to presence of intensive duck farming.

Our results indicate that regions with a highly intensive duck farming industry are most likely to be infected during the H5N1 outbreak – our model that includes this factor provides the closest fit to the
observational data from the 2004 outbreak. Our analysis reveals locations where the probability of finding infected animals is substantially higher, which can be used to identify specific areas where surveillance would be more beneficial than others in terms of increasing the probability of detecting the infection at early stages.

The steady supply of poultry in the live birds markets has the potential to support disease circulation and may act as a potential source of genetic reassortment, emergence of new influenza viruses, zoonotic transmission, and disease spread at a national scale. Several influenza virus surveillance programs have found that HPAI H5N1 and other influenza A viruses has circulated in live bird and food markets in Thailand (Amonsin et al., 2008; Wisedchanwet et al., 2011; Jairak et al., 2016). However, outbreaks were predominantly concentrated in the centre of Thailand (Fig. 3), north from Bangkok, while the highest density of LBMs is in the vicinity of Bangkok Gilbert et al. (2014). No presence of influenza A virus was detected in any of the nine districts of Bangkok with LBM during the surveillance program in 2013 (Tantawiwattananon et al., 2017). We therefore believe that the addition of live birds markets into our model will not improve the power of our modelling to capture the 2004 Thailand outbreak.

We have explored the ability of models of this nature to be used in real time during ongoing outbreaks. Many infectious disease models are used retrospectively to determine the risks associated with spread and the impact of control for previous outbreaks. Whilst this has value, it is also crucial to determine whether these models can be used in real time to advise policy makers during the course of an epidemic. Our results suggest that, during the early stages of HPAI outbreaks, there is significant uncertainty, owing to partial reporting of cases and potential for undetected infections over large areas, in line with previous work on foot-and-mouth disease. Models such as the one we describe here therefore only have limited predictive power in the very early stages of HPAI epidemics. However, as more data are accrued, the uncertainty in predictions decreases, such that after the first few weeks, we are able to accurately predict both the size and the spatial extent of the outbreak. This is highly beneficial in terms of being able to inform targeted surveillance policies to improve detection.

Despite this uncertainty, the models are consistent in determining the control policies that should be deployed, even in the very early stages of the outbreak. This is promising, in that, despite over predictions of spread at epidemic onset, models are able to inform how to target interventions to reduce the risk of spread in the future. The caveat to this is the uncertainty in the impact of some control policies in
the early stages of outbreaks – whilst the model accurately predicts that high radius ring culling is the preferred policy to minimise the number of infected farms even when only using the first 10 days of the outbreak to parameterise the model, the duration of outbreaks is significantly overpredicted for IP only culling and vaccination policies. Additionally, we find that the strategy that minimises the overall cost of an outbreak is highly dependent upon the relative costs of culling, vaccination and other economic losses. We therefore conclude that it is important to determine the objective of control when deciding upon an intervention policy and that model predictions of the spread of the disease and the impact of interventions should be considered with caution during the early stages of any influenza epidemic.

Fig. 4. Effect of local controls for sequential parameter estimation using Model D with outbreak data censored at time $T$ days. (A) Projected number of infected flocks, number of culled flocks and duration of outbreak. (B) Effective control based on minimising the cost function Eq. (4) and varying the cost associated with culling of a single flock ($c_1$), the cost associated with vaccination of one hundred birds ($c_2$), and the cost associated with the monetary impact of a single infected flock ($c_3$).
There is a lack of studies that focus upon measuring the onset of immunity (i.e. the time necessary for a vaccine to induce clinical protection against the infection in transmission studies). The modeling framework outlined can be extended by including a stochastic variation in the delay from vaccination to immunity. We have assumed that there was no bias in detection of infected cases according to species or geographical attributes as roads. Data augmentation Markov chain Monte Carlo technique can be utilised to account for surveillance bias (Jewell et al., 2009a,b,c).

For the future, it is important to explore mechanisms to improve the predictive power of infectious disease models during the early stages of disease outbreaks. It is simply not possible for policy makers to wait for uncertainty to resolve after the first few weeks before employing an intervention and therefore seeking methods to reduce uncertainty in model predictions at epidemic onset is vital (Probert et al., 2018). It is also important to state that a model that has been fitted to a previous outbreak may not necessarily be able to predict the spatiotemporal dynamics of future outbreaks. We would always recommend that, for new outbreaks, even models that have been used previously should be parameterised to take into account the most up to date information from the current outbreak. One area for future research would be to use data from previous outbreaks to provide more informed priors for epidemiological parameters. This would enable us to explore whether this can better establish the likelihood of spread in the first few days of a new epidemic when the outputs of models could potentially have the most significant impact upon reducing the spread of the disease.

5. Supplementary material

5.1. Data

The demographic data describing poultry industry in Thailand during the 2004 outbreak has the following characteristics: number of subdistricts: 7,416; number of flock: 3,303,160; number of chicken: 180,725,929; number of ducks: 19,745,049. This gives that the national poultry density during the outbreak was 395 birds/km².

5.2. Mathematical models

We formulate four mathematical models with increasing complexity: a random process model (Model A), a spatial model (Model B), a spatial model incorporating rice density (Model C), a spatial model incorporating duck density (Model D), and a spatial model incorporating both rice density and duck intensity (Model E).

Any district k is described by a number of flocks, \( F_k \), number of susceptible flocks at time t, \( S_k(t) \), and number of in infectious flocks at time t, \( I_k(t) \). Then any flock i in a subdistrict k is described by a number of chicken, \( NC_{k,i} \) and number of ducks, \( ND_{k,i} \). For further calculations, we normalise number of chicken and ducks as follows:

\[
NC_{k,i} = \left( \frac{NC_{k,i}}{\lambda_{k}} \right)^{\frac{p_c}{2}},
\]

\[
ND_{k,i} = \left( \frac{ND_{k,i}}{\lambda_{k}} \right)^{\frac{p_d}{2}},
\]

where power laws \( p_c \) and \( p_d \) account for virus transmission differences in the two species.

We can write a general expression for the force of infection to a susceptible flock, k:

\[
\lambda_k(t) = (1 + \alpha_c H(r_{ck}, \varepsilon_c, n_{ci}) + \alpha_d H(d_{ck}, \varepsilon_d, n_{di}))
\times \left( \delta(t) + n_k(t) \right) + S_k(t) \left( \beta_{w} W_k \lambda_k(t) + \beta_{b} \sum_{i \in NC_{k,i}} B_{k,i} I_k(t) \right).
\]

(5)

\[
H(x, \varepsilon, n) = \frac{1}{(\varepsilon/n)^\alpha + 1}.
\]

(6)

Here \( r_{ck} \) is the fraction of the area of the subdistrict occupied by rice paddies; the intensity of duck industry is formalised as \( d_{ck} = \log_{(\max_{i=1}^{n}(\lambda_{ki})) + 1} \); \( \eta_k(t) \) is the number of infectious flocks at time t, whilst \( \beta_w \) and \( \beta_b \) are the rates of spatial transmission within and between subdistricts, whilst \( W_k \) and \( B_{k,i} \) are the rates of within-subdistrict transmission and local cross-border transmission. These terms are calculated using a local spatial kernel that implicitly incorporates the impact of distance on contact rates and takes a simple but flexible parametric form (Buherkerkeme et al., 2014). The impact of the local spatial kernel is determined by considering the distance between a randomly located susceptible premises and integrating over all possible locations of infected premises.

The rate of within-subdistrict transmission for flocks with a number of flocks \( F_k > 1 \) is given by

\[
W_k = \frac{d_k}{F_k^2} (F_k - 1) \sum_{i=1}^{F_k} \sum_{j=1 \neq i}^{F_k} (nc_{k,i} + \xi nd_{k,i}) (nc_{j,i} + \xi nd_{j,i}),
\]

(7)

where \( d_k \) is the density of flocks (i.e. number of flocks divided per subdistrict area) and \( F_k \) is a number of flocks in a subdistrict k.

The rate of local cross-border transmission between subdistricts with a number of flocks \( F_k > 0 \) and neighbouring subdistricts \( \{\| \| \in \mathbb{B}_k, F > 0 \} \) is given by

\[
B_{k,i} = \frac{d_k}{A_k F_k^2} \sum_{k \in \mathbb{B}_k} \sum_{i=1}^{F_k} \Omega(\Omega(nc_{k,x} + \xi nd_{k,x})\Omega(nc_{i,x} + \xi nd_{i,x})\Omega),
\]

(8)

where \( L_{i,k} \) is the length of a shared boundary between neighbour districts \( k \) and \( i \) and \( A_k \) is the area of subdistrict k.

We have adopted the analytical expressions for \( W_k \) and \( B_{k,i} \) from (Buherkerkeme et al., 2014). Both terms were derived to account for a probability of contact between premises within and between subdistricts.

The term \( \Omega_{k,i} \) is a scaling factor for between-subdistrict transmission informed by the mean distance between neighbouring subdistricts taking into account the size of each subdistrict, following the approach of Buherkerkeme et al. (2014) such that:

\[
\Omega_{k,i} = \int_{0}^{\infty} \int_{0}^{\infty} \alpha_\Omega (\lambda(t) + (\lambda(t)) dx dy dx,
\]

\[
\Omega(t) = \frac{1}{1 + \left( \frac{\alpha_\Omega}{\lambda(t)} \right)^2},
\]

where \( \alpha_\Omega \) and \( \theta_\Omega \) are the shape and scale parameters of the distance-dependent local spread kernel respectively (Buherkerkeme et al., 2014). The normalisation constant \( \zeta \) is defined such that

\[
\int_{0}^{\infty} 2\pi\Omega(t)dt = 1.
\]

5.3. Bayesian approach for parameter estimation

We use a Bayesian approach to fit the meta-population model to the outbreak data (Jewell et al., 2009a,b,c). The likelihood function has the following expression:

\[
L = \prod_{k \in \mathbb{K}} \prod_{i=1}^{n_k} \prod_{t \in T_k} \left( 1 - \exp(-\lambda(t)) \right)^{I_k(t)} \left( \exp(-\sum_{i=1}^{n_k} \lambda_i(t)) \right)^{0(I_k(t))} \times \prod_{i \in \mathbb{K}} \left( \exp(-\sum_{i=1}^{n_k} \lambda_i(t)) \right)^{I_k(t)} \prod_{i \in \mathbb{K}} \left( \exp(-\sum_{i=1}^{n_k} \lambda_i(t)) \right)^{I_k(t)}
\]

where \( n_k \) is the total number of infection events in a subdistrict \( k \) and \( I_k(t) \) is the time of the ith infection event in a subdistrict \( k \). We introduce \( I_k(t) \) to be the number of flocks that have been infected at time \( t \), and note that this is different from \( I_k(t) \) (i.e. the number of infectious flocks at time \( t \)). For each of the parameters, we update the parameter value using a random walk Metropolis algorithm with a Gaussian proposal: \( \theta^* = N(\theta, \sigma) \), where \( \theta \) and \( \theta^* \) are parameter vectors, and \( \sigma \) is a covariance matrix.

The proposed set of updated parameters \( \theta^* \) is then accepted with
probability
\[
\min \left\{ L(\theta^*) | q(\theta^*, \sigma) \right\},
\]
(9)
where \(q(x, y, s)\) is a normal PDF value at \(x\) with mean \(y\) and covariance matrix \(s\).

As a pre-emptive culling policy within a 1 km radius of flocks reporting infection was introduced during the outbreak, we simulate this by reducing the number of susceptible flocks as a fraction given by the ratio of the area of a circle with 1 km radius to the total area of the subdistrict in question. We assume that there was no re-population of any flocks after culling prior to the end of the epidemic.

5.4. Model simulation

Posterior samples of parameters are used to generate predicted outbreaks. At every time \(t\) of the simulation (time step \(\Delta t = 1\) day), every susceptible flock \(k\) has a probability of becoming infected:
\[
p_k(t) = 1 - \exp(-\lambda_k(t)),
\]
(10)
where the infection pressure depends on a particular model. We start the outbreak by seeding infection in the six subdistricts in Thailand that had reported infection during the first three days of the outbreak. The seed cases were the same for all simulations. The outbreak was assumed to end after seven days without new infected cases (or alternatively the simulation was terminated after 2000 days).

Modeling effect of vaccination

With vaccine capacity \(\gamma\), the demographic factors on the infection pressure then are reduced so that:
\[
nc_{v,k} = \left(1 - \frac{\gamma}{100}\right)n_{c,k} \left(1 - \frac{\rho_c}{100}\right),
\]
\[
n_{d,k} = \left(1 - \frac{\gamma}{100}\right)n_{d,k} \left(1 - \frac{\rho_c}{100}\right).
\]

Financial disclosure

MJT and RR are supported by the Engineering and Physical Sciences Research Council and the Biotechnology and Biological Sciences Research Council [Grant Numbers EP/P511079/1 and BB/K010972/4]. XX, MG and MJT are supported by the National Institutes of Health (NIH Grant 1R01AI101028-02A1). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Data availability

Data are available from the Department of Livestock Development in Thailand.

Author contributions

The author(s) have made the following declarations about their contributions: Conceived and designed the experiments: MT, RR, CJ. Performed the experiments: RR, MT. Analyzed the data: RR, WT. Contributed reagents/materials/analysis tools: TB, MG, XX, GZ, WT. Wrote the paper: RR, MT, MK, TB, MG, GZ, XX, CJ, WT.

Acknowledgments

We thank Will Probert and Ed Hill for useful comments on this manuscript.

References


