Research article

Computing human to human Avian influenza $\mathcal{R}_0$ via transmission chains and parameter estimation

Omar Saucedo[1,*]**, Maia Martcheva**[2] **and Abena Annor**[3]

[1] Mathematical Biosciences Institute, The Ohio State University, Columbus, OH 43210, USA

[2] Department of Mathematics, University of Florida, Gainesville, FL 32611, USA

[3] Morsani College of Medicine, University of South Florida, Tampa, FL 33612, USA

* **Correspondence:** Email: saucedo.10@mbi.osu.edu.

Abstract: The transmission of avian influenza between humans is extremely rare, and it mostly affects individuals who are in contact with infected family member. Although this scenario is uncommon, there have been multiple outbreaks that occur in small infection clusters in Asia with relatively low transmissibility, and thus are too weak to cause an epidemic. Still, subcritical transmission from stuttering chain data is vital for determining whether avian influenza is close to the threshold of $\mathcal{R}_0 > 1$. In this article, we will explore two methods of estimating $\mathcal{R}_0$ using transmission chains and parameter estimation through data fitting. We found that $\mathcal{R}_0 = 0.2205$ when calculating the $\mathcal{R}_0$ using the maximum likelihood method. When we computed the reproduction number for human to human transmission through differential equations and fitted the model to data from the cumulative cases, cumulative deaths, and cumulative secondary cases, we estimated $\mathcal{R}_0 = 0.1768$. To avoid violating the assumption of the least square method, we fitted the model to incidence data to obtain $\mathcal{R}_0 = 0.1520$. We tested the structural and practical identifiability of the model, and concluded that the model is identifiable under certain assumptions. We further use two more methods to estimate $\mathcal{R}_0$: by the $\mathcal{R}_0$ definition which gives an overestimate of 0.28 and by Ferguson approach which yields $\mathcal{R}_0 = 0.1586$. We conclude that $\mathcal{R}_0$ for human to human transmission was about 0.2.

Keywords: avian influenza; transmission chains; identifiability; parameter estimation; basic reproduction number

1. Introduction

In the Netherlands, there was a report of an isolated incident of highly pathogenic avian influenza (HPAI H7N7) emerging from a poultry farm on February 28th 2003. As a result, 225 farms were culled in the affected area which meant that approximately 30 million chickens were killed. Weeks

after the first case was reported, H7N7 was diagnosed in 89 humans who had visited the poultry farms; however, 3 of these individuals did not have any contact with the infected poultry farms suggesting human to human transmission of avian influenza [1]. During 10 months in 2004, highly pathogenic avian influenza H5N1 was reported to infect eight countries in Asia. During this time period, there were 44 human cases that were documented, and 32 of them passed away from avian influenza [2]. In 2005, there was evidence that three clusters of HPAI H5N1 infections occurred in Indonesia resulting in several human cases, and limited human to human H5N1 transmission could not be ruled out among the clusters [3]. While there isn't sustained human-to-human transmission of avian influenza, these outbreaks are alarming and suggest the possibility that more effective transmission could occur in the future if the virus mutates [4].

It is rare for the avian influenza A virus to infect humans and sustain an effective human to human transmission. Avian strains that transmit to humans require genetic assortment in some sort of mammalian intermediary; however, there have been cases where the highly pathogenic H5N1 did not need such genetic assortment [5, 6]. Currently, the known human HPAI strains found in the outbreaks are H5N1, H7N3 or H7N7 [6]. In the human population, avian influenza virus causes flu like symptoms along with pneumonia. As with poultry, the low pathogenic human influenza A virus has a low mortality rate, but the high pathogenic human influenza A (namely, H5N1) has a mortality rate of up to 60 percent [7]. However, the current intra-species transmissions rate for humans is low. There have only been a few clusters of human to human transmissions and these have been among blood relatives who had close contacts, without any preliminary precautions, with an infected person [8]. Considering the high mortality rate of HPAI in humans, it is imperative that the process of transmission be better understood [5, 9].

The first step in our investigation is to construct a suitable model that describes the biological situation for transmission of avian influenza in the chicken and human population. Certain biological systems can be represented by a set of mathematical equations. Thus if a given biological system's data fits a particular model then that data can be used (in conjunction with the models' predetermined mathematical equations) to determine some of the system's descriptive values such as transmission rates and the critical threshold referred to as the basic reproduction number [10]. The reproduction number for an infectious disease gives us insight into the disease's ability to cause an epidemic by serving as a threshold value. The basic reproduction number of a disease is defined as the number of secondary cases an infectious individual causes in a fully susceptible population until the individual is no longer infectious [11].

The objective of this article is to explore the traditional methodology of computing the human to human H5N1 avian influenza basic reproduction $\mathcal{R}_0$ through parameter estimation based on differential equations versus calculating $\mathcal{R}_0$ using transmission chains. There have been several papers that have estimated the human to human reproduction number by estimating parameters from data [12, 13]. Xiao *et al.* developed a model with the intent of replicating the transmission dynamics of avian influenza H7N9. Using confirmed human cases in China, they fitted their model to data to obtain estimates for the transmission rate and estimated that the human to human reproduction number to be 0.467 [12]. Other approaches have been used to estimate $\mathcal{R}_0$. Chowell *et al.* used a Bayesian technique coupled with an SEIR model to find the reproduction number of H7N9 for the May 2013 China outbreak, and Boven *et al.* used final size distributions and household final size data to estimate parameters relevant to the reproduction number by maximum likelihood calculations. The reproduction number Chowell
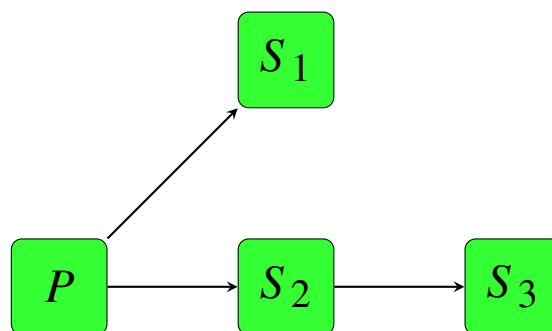
Figure 1. A transmission chain of size 4. P is the primary case and S_n represents the secondary infections linked to P. The arrows represent the transmission of the virus from one individual to another.

estimated was 0.1 while Boven concluded the secondary human to human transmission to be 0.21 [14, 15].

While the differential equation-based approach is a more common method of estimating $\mathcal{R}_0$, the approach of inferring $\mathcal{R}_0$ from transmission chains provides in addition avenue for estimating the reproduction number. Some zoonotic diseases (Monkeypox, Nipah virus, Measles) display subcritical transmission ($0 < \mathcal{R}_0 < 1$) such that infection happens as self contained chains that are too weak to cause an epidemic. H5N1 avian influenza is known to occur in small isolated pockets due to relatively weak transmissibility, and has the characteristics of subcritical transmission [16]. Being able to estimate $\mathcal{R}_0$ from transmission chain data is important for determining whether a disease that has $\mathcal{R}_0$ close to 1 would cause an epidemic, and it provides valuable insight on primary infections and secondary transmissions. Blumberg *et al.* used measles data from the United States and Canada to analyze what the impact of varying assumptions on chain size data would cause on $\mathcal{R}_0$ estimation [17].

This article is structured in the following fashion. In Section 2.1, we estimate $\mathcal{R}_0$ by using transmission chain size data and various maximum likelihood calculations. In Section 2.2, we establish an ordinary differential equation model describing the dynamics of avian influenza in chickens and humans. We use this model to derive the reproduction number and estimate the essential values represented in $\mathcal{R}_0$ by fitting to data. In Section 2.3 and Section 2.4, we explore the question of structural and practical identifiability of model. Finally, we conclude the manuscript by summarizing and discussing the results that we established.

2. Methods

In this section, we describe two methodologies in which we compute the basic reproductive number. The first method deals with computing $\mathcal{R}_0$ through methods described in [17, 18, 19]. This will be the primary method used to calculate the reproductive number for the family cluster data presented in [20]. In the second methodology, we will construct a system of differential equations and find the reproduction number via the next generation method [13].
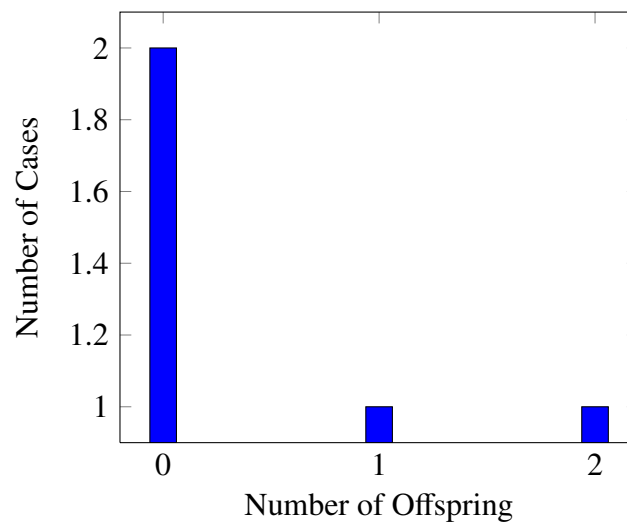
Figure 2. The transmission chain has four cases. Two cases produce zero offsprings. One case produces one offspring. One case produces two offsprings.

2.1. Estimating $\mathcal{R}_0$ through transmission chains

To calculate $\mathcal{R}_0$, we will use avian influenza (H5N1) transmission chain data gathered from [20, 21], and a subfield in probability theory called branching process theory [22]. We define a *transmission chain* as a set for which all secondary infections can be traced back to a primary infection, and the number of cases found in a transmission chain is referred to as the *chain size* of a transmission chain. An example of a transmission chain can be found in Figure 1. This single transmission chain has size four. From Figure 2, there are two cases which produce zero offsprings (S_1 and S_3), one case that produces one offspring (S_2), and one case that produces two offsprings (P). Note that a primary infection without secondary cases is considered as a transmission chain of size 1.

The growth of a population for which an individual generates offsprings can be described mathematically by the Galton-Watson branching process. This method starts with a single primary individual who produces a random number of offsprings for which each of those offsprings produces a random number of offsprings [23]. This mechanism can be characterized by the probability generating function $Q(s) = \sum_{i=0}^{\infty} q_i s^i$ of the offspring distribution. $Q(s)$ defines the probability distribution of the new offspring in the population that are produced by each individual in the population. In the context of this paper, $Q(s)$ can be used to calculate the probability distribution for the number of secondary cases generated by each infected case. q_i describes the probability that an infected individual produces i infections. Selecting the proper offspring distribution is critical to this analysis since we want to determine the link between heterogeneity and spread of the disease. Therefore, we employ the same assumption as in [17] by using a negative binomial distribution with mean $\mathcal{R}_0$ and dispersion parameter κ, which helps explain the degree of transmission heterogeneity in transmission chains. So, we have the following generating function derived from [24].

$$Q(s) = \left(1 + \frac{\mathcal{R}_0}{\kappa}(1 - s)\right)^{-\kappa} \tag{2.1}$$

Now, let $p_i(\mathcal{R}_0, \kappa)$ be the probability of a transmission chain having total size i and define

$$T_i(s) = \frac{Q(s)^i}{i} \tag{2.2}$$

Equation 2.2 is interpreted as the average probability of generating n infections caused by i individuals. From Theorem 1 in [25] and differentiating (2.1) with respect to s, we have

$$p_i = \frac{1}{(i-1)} T_i^{(i-1)}|_{s=0} \tag{2.3}$$

From equation (2.1), we have the jth derivative of $T_i(s)$ is

$$T_i^j(s) = \frac{\prod_{m=0}^{j-1}(\kappa i + m)}{i} \left(\frac{\mathcal{R}_0}{\kappa}\right)^j \left(1 + \frac{\mathcal{R}_0}{\kappa}(1 - s)\right)^{-\kappa i - j} \tag{2.4}$$

Substituting equation (2.4) into equation (2.3),

$$p_i(\mathcal{R}_0, \kappa) = \frac{\prod_{m=0}^{i-2}(\kappa i + m)}{i!} \left(\frac{\mathcal{R}_0}{\kappa}\right)^{(i-1)} \left(1 + \frac{\mathcal{R}_0}{\kappa}\right)^{-\kappa i - i + 1} \tag{2.5}$$

Recall the following special properties of the gamma function $\Gamma(x)$: $x\Gamma(x) = \Gamma(x+1)$ and $x! = \Gamma(x+1)$. Note from equation (2.5) that

$$\prod_{m=0}^{i-2}(\kappa i + m) = (\kappa i)(\kappa i + 1)\dots(\kappa i + (i-2)) \tag{2.6}$$

$$= \frac{\Gamma(\kappa i + 1)}{\Gamma(\kappa i)}\frac{\Gamma(\kappa i + 2)}{\Gamma(\kappa i + 1)}\dots\frac{\Gamma(\kappa i + (i-2) + 1)}{\Gamma(\kappa i + (i-1))} \tag{2.7}$$

$$= \frac{\Gamma(\kappa i + (i-1))}{\Gamma(\kappa i)} \tag{2.8}$$

Thus, the probability of a transmission chain having total size i for a negative binomial offspring distribution is:

$$p_i(\mathcal{R}_0, \kappa) = \frac{\Gamma(\kappa i + (i-1))}{\Gamma(\kappa i)\Gamma(i + 1)}\left(\frac{\mathcal{R}_0}{\kappa}\right)^{i-1}\left(1 + \frac{\mathcal{R}_0}{\kappa}\right)^{-\kappa i - i + 1} \tag{2.9}$$

It is important to know how many transmission chains of a certain size are found in data since they play an integral part in computing the likelihood. In Figure 3, we see the distribution of the chain size for the avian influenza data set. From the data set, there are 94 cases presented of which 36 are primary cases and 58 are secondary cases. From December 2003 to December 2006, there were a total of 263 cases of H5N1 [26]. From here, we observe that there are 169 transmission chains of size 1, 23 transmission chains of size 2, 9 transmission chains of size 3, two transmission chains of size 4, one transmission chain of size 5, and one transmission chain of size 8. We notice skewness to the right in the chain size distribution which suggest that there is a high degree of transmission heterogeneity. Therefore, it is imperative for us to understand how the number of isolated cases and the degree of transmission heterogeneity affects the estimation of $\mathcal{R}_0$.
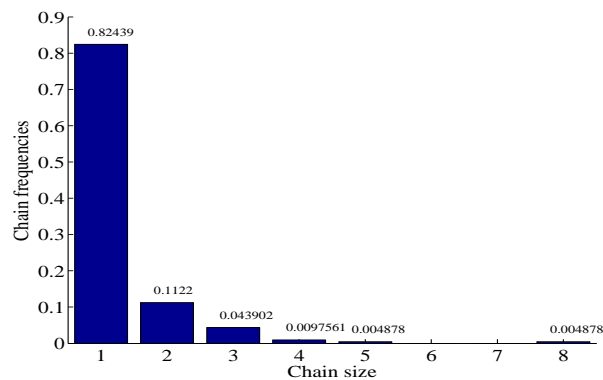
Figure 3. Distribution of chain size for family cluster data from December 2003 to December 2006.

The probability defined in equation (2.9) is the basis for which we will estimate $\mathcal{R}_0$. When imperfections are ignored in the data and no assumptions are made in regarding the data at hand [19], the likelihood for a given chain size distribution with probability $p_i(\mathcal{R}_0, \kappa)$ is defined as

$$L(\mathcal{R}_0, \kappa) = \prod_{i=1}^{\infty} p_i(\mathcal{R}_0, \kappa)^{n_i} \tag{2.10}$$

where n_i is the number of transmission chains of size i. We maximize the likelihood function with respect to $\mathcal{R}_0$ and κ to find the maximum likelihood estimate for these two parameters.

Let $C(i, \alpha) = \alpha \cdot i$ denote the number of transmission chains of size i. Then we define the average observed chain size, $\hat{\mu}$, by $\hat{\mu} = \frac{1}{\hat{N}} \sum_{k=1}^{\hat{N}} C_k(i, \alpha)$, where $\hat{N}$ is the number of transmission chains in the data. When perfect observation is assumed and we have the entire chain size distribution [19], the estimation of $\mathcal{R}_0$ is

$$\mathcal{R}_{0,MLE} = 1 - \frac{1}{\hat{\mu}} \tag{2.11}$$

We express the maximum likelihood estimation of $\mathcal{R}_0$ for this method by $\mathcal{R}_{0,MLE}$.

In order to account for various structures in the data, Equation (2.10) can be modified in two ways to represent a truncated likelihood and aggregated likelihood. When computing the truncated likelihood, we only take a look at the chains in the data that are of size 2 or greater to avoid any discrepancy in the number of isolated cases [19]. Thus, we define the truncated likelihood as:

$$L^T(\mathcal{R}_0, \kappa) = \prod_{i=2}^{\infty} \left(\frac{p_i(\mathcal{R}_0, \kappa)}{1 - p_1(\mathcal{R}_0, \kappa)} \right)^{n_i} \tag{2.12}$$

For the truncated estimate, we may alter the assumption of transmission heterogeneity to obtain three estimators for $\mathcal{R}_0$. By setting $\kappa = 1$, the negative binomial distribution simplifies to geometric offspring distribution, and by letting $\kappa \to \infty$, the negative binomial distribution simplifies to a Poisson offspring distribution for each of our estimators [19, 18]. We can produce a third estimator $\mathcal{R}_{0,\kappa=?}$ by

letting κ be a free parameter and assume that no prior information is provided for κ. This will allow us to determine the influence κ has on the $\mathcal{R}_0$ estimator.

In the aggregated likelihood calculation, intermediate sized chains found in the data are aggregated to account for small chains that may not be observed. The likelihood is constructed by considering the number of isolated cases, the total number of stuttering chains, the size of the largest stuttering chain, and the number of chains having largest size [19]. We define the aggregated likelihood as:

$$L^A(\mathcal{R}_0, \kappa) = p_1(\mathcal{R}_0, \kappa)^{n_1} \left(\sum_{\kappa=2}^{M-1} p_\kappa(\mathcal{R}_0, \kappa) \right)^{\hat{N}-n_1-n_M} p_M(\mathcal{R}_0, \kappa)^{n_M} \tag{2.13}$$

where M denotes the largest chain size. Note that $p_1(\mathcal{R}_0, \kappa)^{n_1}$ denotes the probability of observing n_1 isolated cases and $p_M(\mathcal{R}_0, \kappa)^{n_M}$ denotes probability of observing n_M chains of size M. As in the truncated estimator, the aggregate likelihood formula provides us with three estimators for $\mathcal{R}_0$. Note that the truncated likelihood $\mathcal{R}_{0,\kappa}^T$ and the aggregated likelihood $\mathcal{R}_{0,\kappa}^A$ are maximized over a single parameter $\mathcal{R}_0$ while $\mathcal{R}_{0,MLE}$ and $\mathcal{R}_{0,\kappa=?}$ are maximized over two parameters $\mathcal{R}_0$ and κ.

Table 1. $\mathcal{R}_0$ estimates for different maximum likelihood methods.

	$\mathcal{R}_0$ Estimates	95% Confidence Interval
$\mathcal{R}_{0,MLE}$	0.2205	[0.1625, 0.2955]
$\mathcal{R}_{0,\kappa=?}^T$	0.2655	[0.000, 0.4217]
$\mathcal{R}_{0,\kappa=1}^T$	0.2340	[0.1435, 0.3649]
$\mathcal{R}_{0,\kappa\to\infty}^T$	0.2862	[0.1831, 0.4219]
$\mathcal{R}_{0,\kappa=?}^A$	0.2455	[0.1697, 0.3564]
$\mathcal{R}_{0,\kappa=1}^A$	0.2301	[0.1642, 0.3125]
$\mathcal{R}_{0,\kappa\to\infty}^A$	0.2151	[0.1568, 0.2855]

Figure 4 represents the estimates for $\mathcal{R}_0$ and κ with respect to the likelihood calculation for the original data, truncated data, and aggregated data. The results of this analysis are summarized in Table 1. The confidence intervals were computed by using likelihood profiling. $\mathcal{R}_{0,\kappa=1}^T$ and $\mathcal{R}_{0,\kappa\to\infty}^T$ produce higher estimates than $\mathcal{R}_{0,MLE}$ which is expected since the truncated method assumes the chains of size 1 are under-represented in the data set. $\mathcal{R}_{0,\kappa=1}^A$ and $\mathcal{R}_{0,\kappa\to\infty}^A$ produce the lower $\mathcal{R}_0$ estimates when compared to the truncated method since this method relies on isolated cases which make up about 82% of transmission chain data. In contrast to the truncated method, the aggregated method assumes its transmission chains are homogeneous implying there will be fewer isolated cases which results in smaller $\mathcal{R}_0$ estimates. Observe that all likelihood estimates fall within 95% confidence interval of the full distribution maximum likelihood estimate $\mathcal{R}_{0,MLE}$. The optimized dispersion parameter for the maximum likelihood estimate was $\kappa = 0.751$.

In order to distinguish what method formulates the best likelihood estimate when compared to the initial data set, we compute the likelihood scores. Table 2 gives us the relative log likelihood scores for the avian influenza data. The likelihood scores are computed by taking the $\mathcal{R}_0$ estimates from Table 1 and computing the likelihood using the full data set and the truncated data set for each of the approaches. This is done to observe the variation between each of the estimation methods. As a reference point, we used $\mathcal{R}_{0,MLE}$ for ΔL and $\mathcal{R}_{0,\kappa=?}^T$ for ΔL^T since these estimates produced the smallest
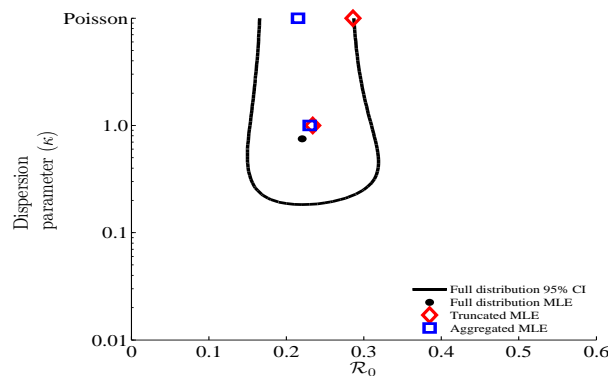
Figure 4. The estimates of $\mathcal{R}_0$ and κ for the family cluster data set.

likelihood scores for their respective methods. When we consider all chain size data in the likelihood calculation L, $\mathcal{R}_{0,MLE}$ gives the best likelihood score. This result makes sense because the $\mathcal{R}_{0,MLE}$ method is the only metric that takes into account all the information of the chain size distribution. The aggregate likelihood $\mathcal{R}_{0,\kappa=1}^A$ is the other estimator that is close to the likelihood score of $\mathcal{R}_{0,MLE}$ with -0.09 in relative log likelihood difference. From this analysis, we observe the models that assume a Poisson offspring distribution produce the lowest likelihood scores. When comparing the truncated method versus the aggregated method, we observe that the aggregated method produces better estimates than the truncated method. Since the likelihood calculation considers all the chain size data, and the truncated method ignores the isolated cases, this helps explain why the aggregated methods would yield better $\mathcal{R}_0$ estimates.

Table 2. $\mathcal{R}_0$ estimates for different maximum likelihood methods.

$\mathcal{R}_0$ Estimates	ΔL	ΔL^T
$\mathcal{R}_{0,MLE}$	0.0	-0.008
$\mathcal{R}_{0,\kappa=1}^T$	-0.13	-0.005
$\mathcal{R}_{0,\kappa\to\infty}^T$	-3.17	-0.003
$\mathcal{R}_{0,\kappa=?}^T$	-1.47	0.0
$\mathcal{R}_{0,\kappa=1}^A$	-0.09	-0.007
$\mathcal{R}_{0,\kappa\to\infty}^A$	-1.04	-0.827
$\mathcal{R}_{0,\kappa=?}^A$	-0.55	-0.555

If we remove the isolated cases from the likelihood calculation, we notice that the differences between the likelihood scores in the L^T calculation are smaller compared to the L calculation. Recall that the L^T calculation doesn't consider the isolated cases which means that there are less data points in the calculation. Hence, this leads to the smaller differences in the likelihood scores. In this scenario, the aggregated estimators perform worse than the truncated estimators due to the isolated cases being removed from the likelihood calculation. While the assumption of Poisson offspring distribution provides the best likelihood score for the truncated estimator when compared to $\mathcal{R}_{0,\kappa=?}^T$, the Poisson distribution assumption gives the worst likelihood score for the aggregated estimator when compared to $\mathcal{R}_{0,\kappa=?}^T$.

2.2. *Estimating $\mathcal{R}_0$ through Differential Equations*

We begin by defining the system of differential equations that describes the interaction between the human and the poultry population. The primary objective of our model is to capture the epidemiological dynamics of chicken-human interactions, and derive the basic reproduction number from the system.

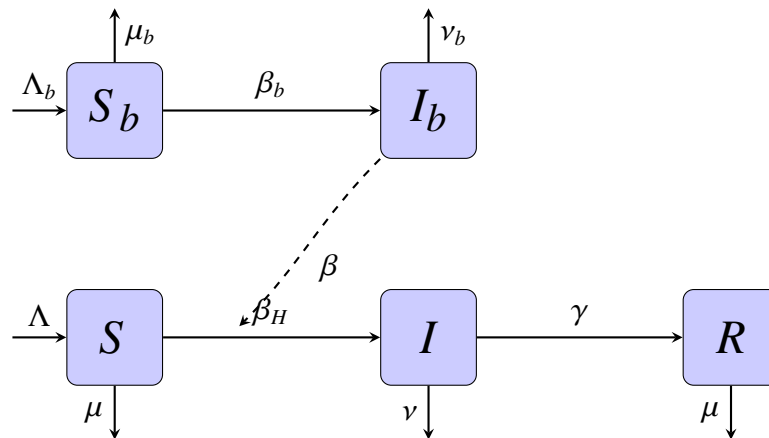### Dynamics of Avian Influenza in Chicken and Human Population

Figure 5. The dashed line represents the chickens infected with avian influenza that transmit the disease to humans.

$$S' = \Lambda - \beta S I_b - \beta_H S I - \mu S$$

$$I' = \beta S I_b + \beta_H S I - (\mu + \nu + \gamma) I$$

$$R' = \nu I - \mu R$$

$$S'_b = \Lambda_b - \beta_b(t) S_b I_b - \mu_b S_b$$

$$I'_b = \beta_b(t) S_b I_b - (\mu_b + \nu_b) I_b$$

We introduce the model characterizing the dynamics of H5N1 in the chicken and human populations. For the chicken population, we use a simple SI model since the disease kills the chickens or the infected chickens are culled to prevent the spread of avian influenza. S denotes the number of susceptible humans at time t, I denotes the number of infected humans with H5N1 at time t, and R denotes the number of humans who have recovered from H5N1 at time t. S_b and I_b defines the susceptible chickens and the chickens infected with avian influenza, respectively. The description of each parameter in the model can be found in Table 3. The demographic parameters for humans and chickens and the duration of the infectious period are pre-estimated from the literature.

To find the total population size for the system, N, we use the population size from the countries from which the family cluster data and the WHO data are derived. The countries that are represented in the data sets are Vietnam, Thailand, Cambodia, Indonesia, China, Turkey, Iraq, Egypt, and Azerbaijan [20, 26]. We estimated that the total population of the region is $N = 19225$ (in units of 10^5) individuals.

Table 3. Definition of each parameter in the model.

Parameter	Description	Value
$S_b(0)$	initial number of susceptible chickens	63505
$I_b(0)$	initial number of infected chickens	0.000436
$S(0)$	initial number of susceptible humans	19220
$I(0)$	initial number of infected humans	0.0000175609
$R(0)$	initial number of recovered humans	0.000004
$C(0)$	cumulative number of human cases	0.00012
$F(0)$	cumulative number of secondary human cases	0.00004
$D(0)$	cumulative number of human deaths	0.00008
γ	recovery rate of humans	0.15
β	transmission rate from chickens to humans	Fitted
β_H	transmission rate between humans	Fitted
$\beta_b(t)$	seasonal transmission rate of chickens	Fitted
Λ_b	recruitment rate of chickens	63505/730
Λ	recruitment rate of human	19225/(75*365)
ν	death rate of infected humans	Fitted
ν_b	death rate of infected chickens	0.1
μ	natural death rate of humans	1/(75*365)
μ_b	natural death rate of chickens	1/(2*365)

Since the China is the largest country in the data sets, we use their average lifespan which is 75 years [27]. Thus, we define the natural death rate for humans as $\mu = \dfrac{1}{(75 * 365)}$ days^{-1}.

The total human population size in the model is $N = S + I + R$ which satisfies the differential equation $N' = \Lambda - \mu N - \nu I$. Observe that the total human population size satisfies

$$N' \leq \Lambda - \mu N$$

This tells us that $\limsup\limits_{t \to \infty} N \leq \dfrac{\Lambda}{\mu}$. We will use this approximation for the total human population size to estimate the parameter Λ. An analogous argument can be constructed to estimate the parameter Λ_b.

The parameter Λ represents the recruitment rate of humans, and as noted above, we define as $\Lambda = N * \mu$. Given μ, we have that $\Lambda = \dfrac{19225}{(75 * 365)}$. The mean duration of infection in humans of the bird-to-human transmission estimated to be 6-7 days [27, 28] hence $\gamma = 0.15$ days^{-1}. The total chicken population in these region is about $N_b = 63505$ in units of 10^5 [29]. The lifespan for commercial poultry is about 2 years which means that $\mu_b = \dfrac{1}{(2 * 365)}$ days^{-1}. Thus, $\Lambda_b = \dfrac{63505}{(2 * 365)}$. The duration of infection in domestic birds is 10 days which means that $\nu_b = 0.1$ days^{-1} [13].

Studies have shown that seasonality has played a factor in the transmissibility of avian influenza between birds [30]. To incorporate that factor into the model, we assume the transmission rate among chickens $\beta_b(t)$ to be periodically forced. This transmission rate is assumed to have sinusoidal behavior.

Therefore, we define the transmission rate for chickens as:

$$\beta_b(t) = \kappa_1 \sin\left(\frac{2\pi}{365}(t + \omega)\right) + \kappa_2$$

This transmission rate assumes a 365 day periodicity. The first parameter in $\beta_b(t)$, κ_1, defines the amplitude, κ_2 represents the vertical shift, and ω the phase shift. All of these parameters in the transmissions rate $\beta_b(t)$ are fitted using the computer algebra program Matlab R2018a. Since the bird-to-human and human-to-human transmission events are sporadic and rare, we kept these transmission constant [31, 32].

The human reproduction number for this model can be explicitly computed and is given by

$$\mathcal{R}_0 = \frac{\Lambda\beta_H}{\mu(\mu + \nu + \gamma)}. \tag{2.14}$$

Similarly, the poultry reproduction number for this model can be explicitly computed by [17, 25]. Although our interest lies at estimating the reproduction number for humans, we nevertheless find the reproduction number for the chickens,

$$\mathcal{R}_b = \frac{\Lambda_b \int_0^{365} [\kappa_1 \sin(\frac{2\pi}{365}(t + \omega)) + \kappa_2]dt}{365\mu_b(\mu_b + \nu_b)}, \tag{2.15}$$

which simplifies to

$$\mathcal{R}_b = \frac{\Lambda_b\kappa_2}{\mu_b(\mu_b + \nu_b)}. \tag{2.16}$$

Figure 6 represents the cumulative number of human cases of H5N1 infections $C(t)$, the cumulative number of human deaths from H5N1 infections $D(t)$, and the cumulative number of secondary H5N1 human cases $F(t)$. The cumulative number of human cases and deaths is obtained from the World Health Organization (WHO) database, and the cumulative number of secondary cases are obtained from the family cluster data set [20, 26]. The time span of the data set is from December 25, 2003 to December 27, 2006 and each data set consists of 29 data points. We used this time frame since the family cluster data set is limited to this period.

We fitted our model to the data using initial guess values for $\nu, \beta_H, \beta, \kappa_1, \kappa_2$, and ω to acquire the best estimates for these parameters. The blue curve denotes the model fit to each respective data set. Since the data are given as cumulative number of cases, deaths, and secondary cases, we fit the cumulative number of human cases, deaths, and secondary cases defined as:

$$C(t) = \int_0^t (\beta S(\tau)I_b(\tau) + \beta_H S(\tau)I(\tau))d\tau$$

$$F(t) = \int_0^t \beta_H S(\tau)I(\tau)d\tau$$

$$D(t) = \int_0^t \nu I(\tau)d\tau$$

We determined the goodness of fit for the model by using the ordinary least squares method. Since we want to measure how well our model fits to the observed data, we examine the model output for
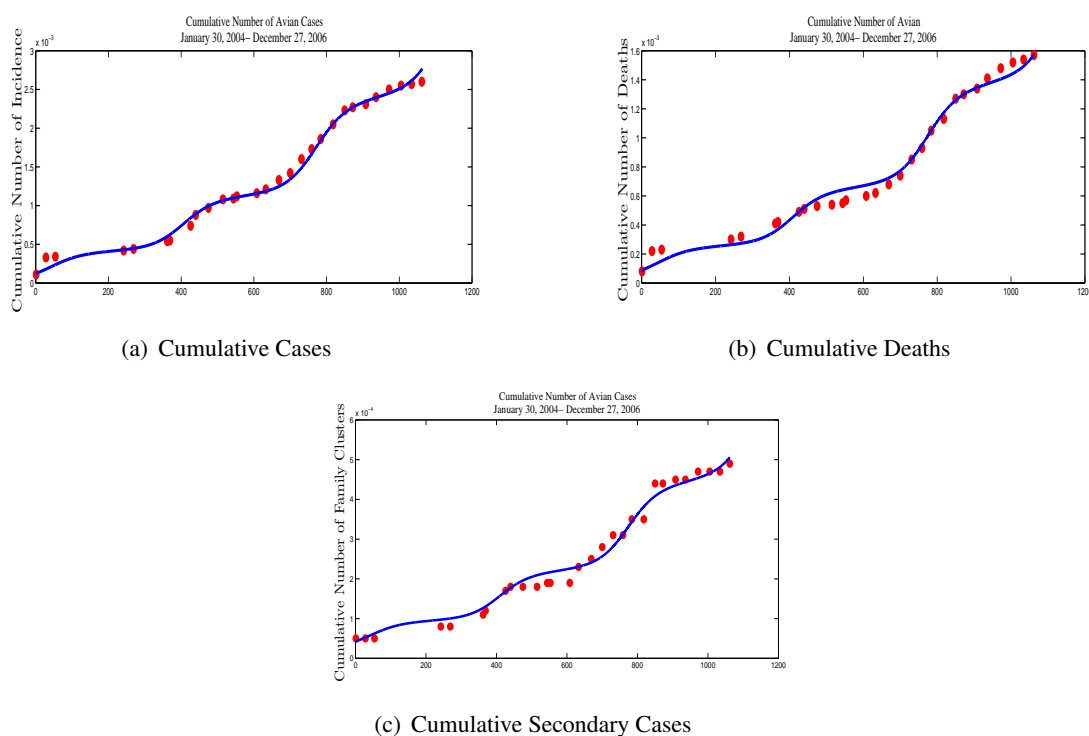
(a) Cumulative Cases

(b) Cumulative Deaths

(c) Cumulative Secondary Cases

Figure 6. Model output fitted to WHO cumulative data from December 2003 to December 2006.

the number of cases, deaths, and secondary cases. In the optimization process, we minimized the least square distance

$$SSR = \sum_{j=1}^{n} [(C(t_j) - y_{1j})^2 + (D(t_j) - y_{2j})^2 + (F(t_j) - y_{3j})^2], \tag{2.17}$$

where y_{ij} denotes the observed data points for each data set, y_{1j} is the cumulative number of human cases at time t_j, y_{2j} is the cumulative number of human deaths at time t_j, and y_{3j} is the cumulative number of secondary at time t_j. We used an ODE solver (ode45) in Matlab to solve numerically our system of differential equations and the built-in nonlinear solver *fminsearch* for the numerical optimization. Since the optimization method relies heavily on the initial values for each parameter, we started off by manual fitting using Mathematica 9 and the Manipulate feature to give us a good starting point. We then used these parameter values in Matlab and minimized the SSR. We repeated this procedure until we obtained the smallest value for SSR and no further improvements in the SSR were observed. The results for the optimization procedure were very sensitive to the initial conditions for $S(t)$ and $S_b(t)$. Therefore the estimated parameters that we obtained are solely dependent on these initial conditions, and are not a set of unique parameter values. In addition, if the initial conditions for $\nu, \beta_H, \beta, \kappa_1, \kappa_2$, and ω were marginally changed from the initial guess, the recovered parameters and the reproduction numbers would change slightly from what we estimated. As King *et. al.* indicated, the use of a deterministic model to fit to cumulative data could result in the violation of the independent errors assumption needed for the least square procedure [33]. Thus in Figure 7, we fitted the model to incidence data which yielded a reasonable fit.
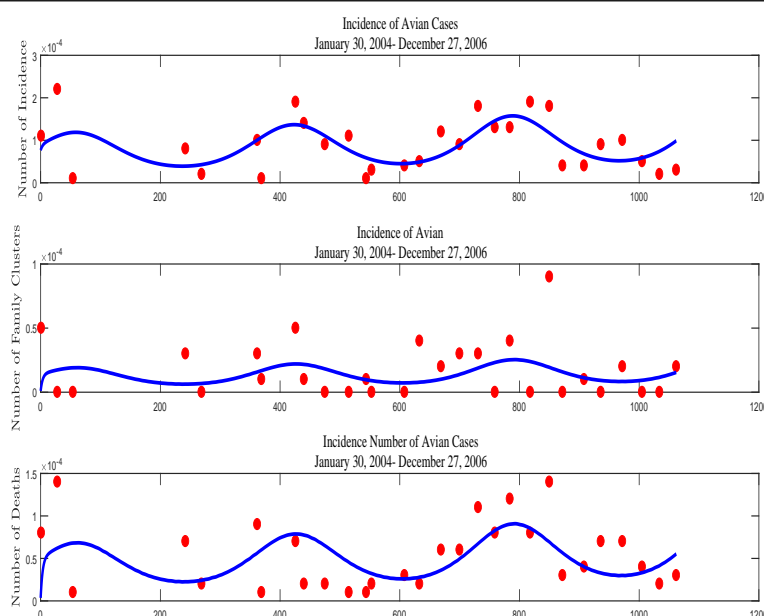
Figure 7. Model output fitted to WHO incidence data from December 2003 to December 2006.

According to our simulations, we obtain the following values for the estimated parameters for the cumulative data: ν=0.194, $\beta = 1.767 \times 10^{-7}$, $\beta_H = 3.168 \times 10^{-6}$, $\kappa_1 = 2.409 \times 10^{-7}$, $\kappa_2 = 1.621 \times 10^{-6}$, and ω=147.634. With these fitted parameters, we observe that the human and chicken reproduction numbers are $\mathcal{R}_0 = 0.1768$ and $\mathcal{R}_b = 1.0156$, respectively. Note that this human reproduction number differs slightly from the human reproduction number calculated in Section 2.1. From the incidence data fitting, the estimated parameters produced reproduction numbers ($\mathcal{R}_0 = 0.1520$ and $\mathcal{R}_b = 1.0039$) that are relatively identical to the reproduction numbers generated from the estimated parameters of the cumulative data fitting. The transmission chains and the next generation operator methods cannot be compared directly with each other due to the fact that the data are treated differentially in each scenario. In the differential equations scenario, we took into account all the cases, deaths, and secondary infections that occurred during December 2003 to December 2006; while in the likelihood approach, the transmission chain contained only information about primary and secondary cases.

2.3. Structural Identifiability

Determining whether the inverse problem is well posed for a given model and data set is the first step in parameter estimation. Since many parameters in a model are unmeasurable in practice, associating epidemic models with data to produce predicative results requires a plethora of parameter estimation and identifiability techniques. Structural identifiability analysis allows us to investigate whether the model parameters can be identified, provided that we have noise free data [34, 35]. Identifiability analysis provides an avenue for which the model can be reconstructed to determine which combinations of parameters can be estimated even when single parameters were deemed unidentifiable [36, 37]. If the ODE system is not structurally identifiable then the parameters estimated by a numerical optimization technique might yield unreliable results. On the other hand, a mathematical model which is structurally identifiable may not be practically identifiable. We note that in our model, all of the parameters are

Table 4. Estimates from cumulative and incidence data fitting.

Parameter	Description	Value
ν	cumulative fit estimate	0.194
ν	incidence fit estimate	0.241
β	cumulative fit estimate	1.767×10^{-7}
β	incidence fit estimate	8.951×10^{-6}
β_H	cumulative fit estimate	3.168×10^{-6}
β_H	incidence fit estimate	3.089×10^{-6}
κ_1	cumulative fit estimate	2.429×10^{-7}
κ_1	incidence fit estimate	1.596×10^{-7}
κ_2	cumulative fit estimate	1.621×10^{-6}
κ_2	incidence fit estimate	1.602×10^{-6}
ω	cumulative fit estimate	147.483
ω	incidence fit estimate	126.5973

constants except for $\beta_b(t)$ which may cause an issue in this analysis. However, Kelejian states that structural identifiability in a system with random parameters is identical to the constant coefficient version of the model [38]. Therefore, we may treat $\beta_b(t)$ as a constant, and proceed to conduct the identifiability analysis.

Mathematically, we say a parameter set θ is structurally globally identifiable if for every θ_0 in the parameter space, if $f(\mathbf{x}(t), \theta) = f(\mathbf{x}(t), \theta_0)$ implies $\theta = \theta_0$.
For our model, if

$$C(\mathbf{x}(t), \theta) = C(\mathbf{x}(t), \theta_0) \qquad (2.18)$$

$$D(\mathbf{x}(t), \theta) = D(\mathbf{x}(t), \theta_0) \qquad (2.19)$$

$$F(\mathbf{x}(t), \theta) = F(\mathbf{x}(t), \theta_0) \qquad (2.20)$$

implies $\theta = \theta_0$ then the model is structurally identifiable. Figure 8 illustrates how a parameter set is structurally identifiable. Figure 8 A is unidentifiable since $f(2) = f(8) = 10$, but $2 \neq 8$. Figure 8 B is identifiable since f is injective i.e. $f(5) = f(5) = 5$ implies $5 = 5$. It is important for us to verify the structural identifiability of the model since parameter estimation results rely on the predictive capability of the model, and there are several methods constructed to undertake this assignment. The general methods used are Taylor Series expansion, differential algebra, exact arithmetic rank (EAR), and implicit function theorem among others [39]. For our analysis, we will use the differential algebra approach and EAR to test the model. One of the strengths of the differential algebra approach method is that if the model is unidentifiable, the identifiable parameter combinations can be obtained. Using the identifiable parameter combinations, the model can be re-parameterized to obtain a structurally identifiable model [35].

The differential algebra approach builds upon deriving the input-output function which contains all structural identifiability information of the model. Using Ritt's algorithm, the input-output equations are determined from the characteristic sets [40]. To derive an input-output equation for the model, we used the Differential Algebra for Identifiability of Systems (DAISY) software [41]. The results from DAISY tell us that Λ_b, β_b, and β are unidentifiable. DAISY yields the following key parameter

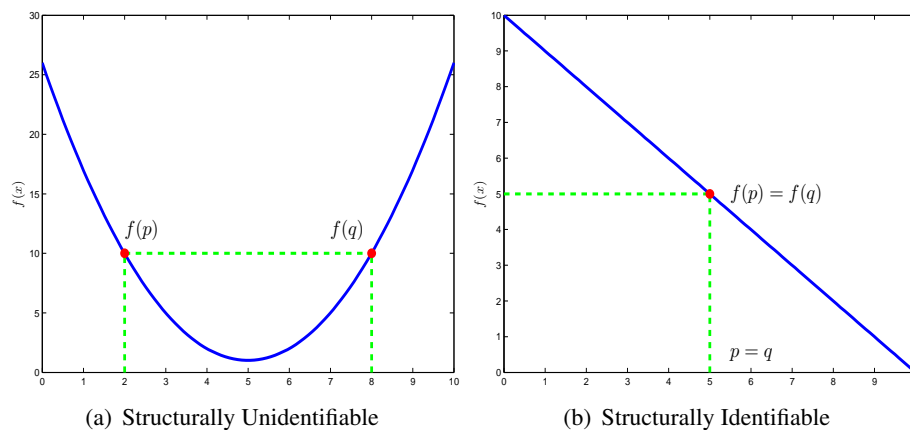(a) Structurally Unidentifiable   (b) Structurally Identifiable

Figure 8. Generic example describing whether or not a model is identifiable.

combinations that cause the model to be unindentifiable:

$$\Lambda_b = \frac{312}{\beta_b}, \quad \beta = \frac{11\beta_b}{24}$$

These combinations give us information about how we can reparameterize the model so that it can be identifiable. If we fix Λ_b, the parameter combination from DAISY informs us that the model becomes identifiable.

The EAR approach uses the inverse function theorem to the system of algebraic differential equations. The solvability of the system can be determined by examining the rank of the Jacobian matrix for this system of equations. If it's a rank-deficient matrix, the Jacobian would grant us insight into what parameters have an association with each other resulting in a non-identifiable model. The EAR method takes care of any rank scenario by efficiently computing a generic rank of the Jacobian matrix which allows a conclusive result about the identifiability [42]. We used the Mathematica package '*IdentifiabilityAnalysis*' created by Karlsson *et al.* to perform this procedure. The EAR approach concluded that Λ_b, β_b, and β are unidentifiable, which is consistent with the results from DAISY. Again if we fix Λ_b, the EAR approach tells us that our system is identifiable.

2.4. Practical Identifiability

In the previous section, we examine the structural identifiability of the model. This analysis deals with the design of the model itself, and is independent of empirical data. This potentially poses a problem in practice since a parameter that is deemed structurally identifiable can potentially be practically unidentifiable. There may be several underlying reasons why a parameter is unidentifiable, but knowing that a parameter is structurally identifiable will provide us the insight of why a parameter is not practically identifiable. Noisy data and the inability for the numerical optimization algorithm to locate the minimum SSE are the two possible causes of loss of practical identifiability from that scenario. In this section, we will undertake the issue of practical identifiability by performing a bootstrap algorithm and creating a profile likelihood of the fitted parameters.

The parameter estimates obtained in Section 2.2 are most likely an inaccurate description of the true parameter values, due to the small sample size and noise in the data. However, we will construct

the confidence intervals for our parameters estimates. We will use a bootstrapping algorithm which provides a way to construct a confidence interval with small sample size [43]. The algorithm will proceed as follows:

1. Begin by estimating the parameter set θ_0 from the data sets y_{ij} using the ordinary least squares method for $i = 1, 2, 3$ and $j = 1 \ldots 29$.

2. Define the standardized residuals from θ_0 for each data set: $r_{ij} = \sqrt{\dfrac{n}{n-p}}(y_{ij} - f(t_j, \theta_0))$, where n is the number of data points and p is the number of parameters in θ_0.

3. Create a bootstrap sample size using random samples with replacement from the residual set in step 2 to form a bootstrap sample of residuals r_{ij}^k.

4. Create new data sets from the residuals by adding the residuals to the model output. In this step, we must use the estimate θ_0 to evaluate the model: $y_{ij}^k = f(t_j, \theta_0) + r_{ij}^k$ where k is the iteration index.

5. Using the new data sets, find a new estimate θ.

6. Store values of θ into a matrix and repeat the algorithm 1000 times.

Table 5. 95% Confidence Intervals for each estimated parameter.

Parameter	95% Confidence Interval	Relative Error
ν	[0.1789, 0.2071]	0.00101
β_H	[2.9100, 3.2827]$\times 10^{-6}$	0.0229
β	[1.5185, 2.0697]$\times 10^{-7}$	0.0148
κ_1	[1.8854, 3.0190]$\times 10^{-7}$	0.0092
κ_2	[1.6189, 1.6234]$\times 10^{-6}$	0.00013
ω	[134.843, 159.671]	0.0015

Table 6. 95% Confidence Intervals for each estimated parameter in the incidence case.

Parameter	95% Confidence Interval	Relative Error
ν	[0.0465, 0.4613]	0.0556
β_H	[7.3443 $\times 10^{-7}$, 7.2807 $\times 10^{-6}$]	0.2974
β	[4.7958 $\times 10^{-6}$, 1.3109 $\times 10^{-5}$]	1.6198$\times 10^{-4}$
κ_1	[7.4659$\times 10^{-8}$, 2.5507$\times 10^{-7}$]	0.0329
κ_2	[1.5907, 1.6137]$\times 10^{-6}$	1.3812$\times 10^{-4}$
ω	[99.2198, 154.4013]	0.0017

Figure 9 represents the distribution of the parameter estimates using the bootstrapping algorithm on the cumulative data. Figure 11 represents the distribution of the parameter estimates using the bootstrapping algorithm on the incidence data. The red star in each figure represents the parameter estimate obtained using the optimization method. We see that the majority of the distributions for the parameters resemble a normal distribution with the exception of the distribution for β_H, κ_2, and ν in the incidence case. Although the histograms for these three parameters do not display a normal distribution and the histogram of κ_2 is erratic, the largest frequencies for β_H, κ_2, and ν center near the optimized value. In the case for the cumulative data, we notice that all of the true parameters fall on or
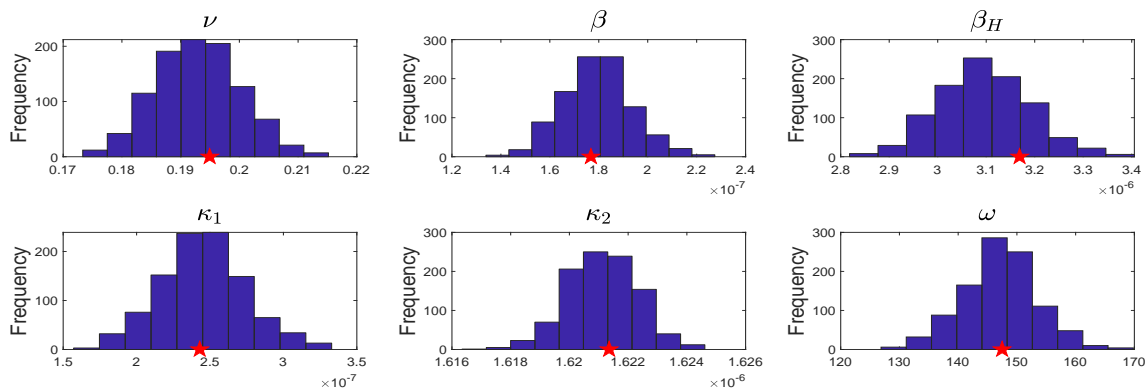
Figure 9. Distribution of parameter estimates for cumulative data: bootstrapping algorithm for 1000 iterations.
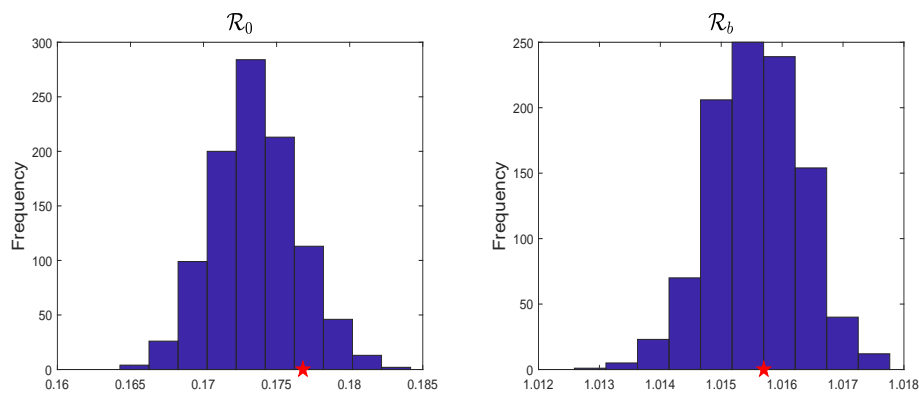

Figure 10. Distribution of $\mathcal{R}_0$ and $\mathcal{R}_b$ estimates for cumulative data: bootstrapping algorithm for 1000 iterations.

near the mean of the distribution except for β_H. Obtaining these type of results reveals the underlying distribution for our parameter space. Using the values for the parameters in the human and chicken reproduction number, we plotted the histogram for each respective reproduction number, and note that they too approximate a normal distribution (Figure 10 and Figure 12) with the exception of $\mathcal{R}_b$ in the incidence case. The distribution for $\mathcal{R}_b$ is attributed to the distribution of κ_2 and equation (2.16). The bootstrap results tell us that the data is noisy as excepted, but it isn't too noisy to obtain reliable parameter estimates.

In Tables 5 and 6, we see the 95% confidence intervals for each of the parameters that were estimated along with their relative error. We define the relative error as

$$RE = \frac{|\theta_{true} - \theta_{estimate}|}{\theta_{true}}$$

where $\theta_{estimate}$ is the mean value for the estimated parameter in the bootstrap algorithm. In the cumulative case, β and β_H yield the largest relative errors while the ν, κ_1, κ_2, and ω have low relative errors. In the incidence case, all the parameters had a small relative error with the exception of β_H. These results tell us that the parameters with low relative error can be practically identified while β_H is practically unidentifiable using the bootstrap algorithm due to the high relative error in both cases.

Another method of testing the practical identifiability of the model is to calculate the profile likeli-
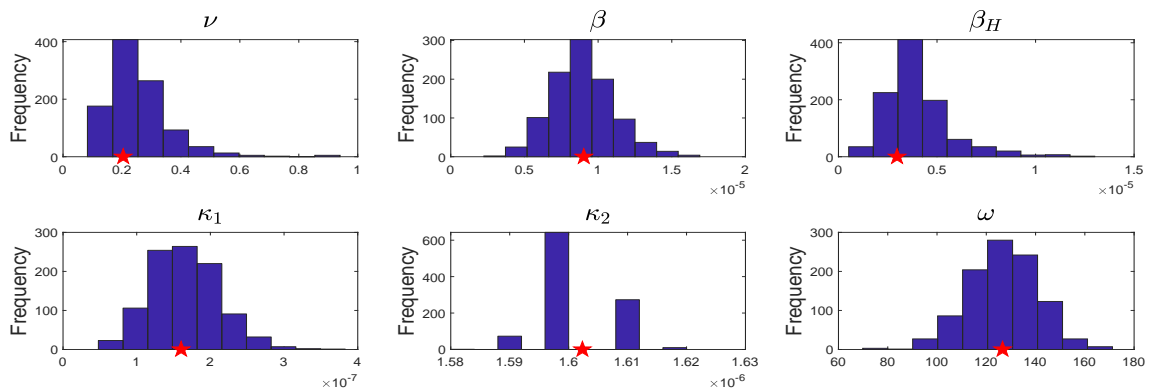
Figure 11. Distribution of parameter estimates for incidence data: bootstrapping algorithm for 1000 iterations.
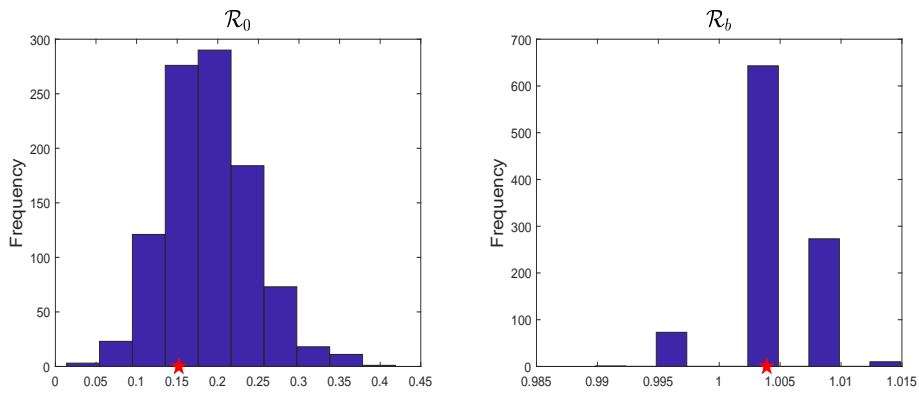

Figure 12. Distribution of $\mathcal{R}_0$ and $\mathcal{R}_b$ estimates for incidence data: bootstrapping algorithm for 1000 iterations.
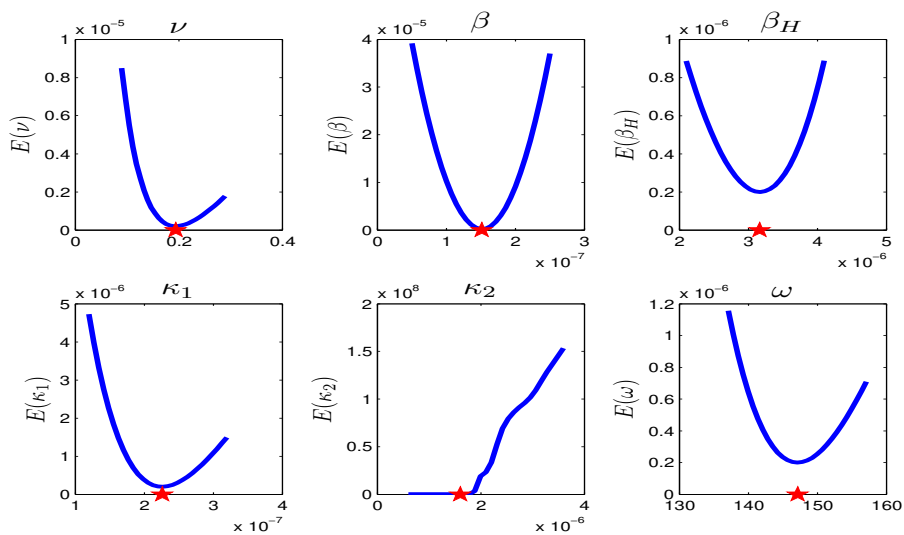

Figure 13. Profile Likelihood of the estimated parameters of the model for the cumulative data.

hood of each fitted parameters ($\mathbf{p}=[\nu, \beta, \beta_H, \kappa_1, \kappa_2, \omega]$). We define the profile likelihood of the fitted parameters $p \in \mathbf{p}$ is

$$E(p_i) = \min_{p_i \notin \mathbf{p}} SSE(\mathbf{p})$$

The purpose of this method is to determine whether a given parameter yields a unique profile likelihood value. If the graph of the profile likelihood is constant on a certain interval or stays below a fixed threshold, we classify that parameter to be practically unidentifiable [44]. For this simulation, $\mathbf{p}$ = $[0.1949, 1.767 \times 10^{-7}, 3.169 \times 10^{-6}, 2.429 \times 10^{-7}, 1.621 \times 10^{-6}, 147.483]$, and the local minimum is $E(p_i) = 1.99561 \times 10^{-7}$. In Figure 13, we see the profile likelihood of the parameters $\mathbf{p}$. Notice that the minimum of the profile likelihood is achieved at the fitted values for $\nu, \beta, \beta_H, \kappa_1, \omega$, but the profile likelihood for the parameter κ_2 is horizontal at the minimum. Therefore, this simulation suggests that κ_2 is practically unidentifiable while the rest of the fitted parameters are practically identifiable.

3. Discussion and conclusion

Constructing models that adequately describe the data from an outbreak is vital for public health. In this regard, parameter estimation plays a key role in developing such models. Testing the structural and practical identifiability for a model is imperative when it comes to estimating the basic reproduction number $\mathcal{R}_0$. In this article, we explored techniques that estimated $\mathcal{R}_0$ for human to human transmission from transmission chain data and parameter estimation via data fitting for avian influenza.

Incorporating transmission chain data and the likelihood function provides another mathematical process for estimating $\mathcal{R}_0$ instead of using the traditional differential equations procedure. Lloyd-Smith *et al.* considered a negative binomial distribution to be a more flexible method to employ when dealing with transmission data since it can describe a degree of transmission heterogeneity, and can be adjusted to a variety of infectious diseases [24]. The two parameters in the negative binomial distribution are $\mathcal{R}_0$ and the dispersion parameter κ. These two parameters are critical to understanding how a disease spreads since $\mathcal{R}_0$ represents the average secondary cases produced by an infectious individual in a susceptible population, and κ describes the variation in the levels of infectiousness in an individual. We implemented three likelihood methods that incorporate varying assumptions: the standard likelihood L, the truncated likelihood L^T, and the aggregated likelihood L^A. We found that the estimates for $\mathcal{R}_0$ were not equivalent under varying assumptions. When a data set has information about the full distribution chain size, it's better to use the L likelihood method over the truncated and aggregated method. However, if the data are lacking isolated cases, the truncated likelihood method performs the best in this scenario. The full distribution likelihood approach produced an estimate $\mathcal{R}_0 = 0.2205$. Ferguson and Fraser established $\mathcal{R}_0 = \dfrac{P_c}{1 - P_c}$, where P_c is the proportion of avian to human cases generating secondary cases. Out of the 263 cases that occurred between December 2003 to December 2006, 36 primary cases are known to produce secondary cases. Therefore, $\mathcal{R}_0 = 0.1586$ from Ferguson and Fraser's formula for $\mathcal{R}_0$ [45]. By the epidemiological definition of $\mathcal{R}_0$, we have that $\mathcal{R}_0$ should approximately equal the number of secondary cases divided by the number of primary cases which produces $\mathcal{R}_0 \approx 0.2829$.

We developed a system of differential equations to describe the dynamics of transmission of avian influenza from chickens to humans. Once we established the model, we computed the basic reproduction number $\mathcal{R}_0$ for the humans and $\mathcal{R}_b$ for the chickens. We saw that the human basic reproduction

number was composed of the parameters Λ, μ, γ, ν, and κ_2. In order to gain a proper estimate for $\mathcal{R}_0$, we estimated ν and κ_2 along with β and β_H from the cumulative number of H5N1 cases, the cumulative number of H5N1 deaths and the cumulative number of secondary H5N1 cases data set. The other parameters in $\mathcal{R}_0$ were pre-estimated from the literature. The bird to human transmission rate we estimated was within two orders of magnitude of what Hsieh *et al.* found for the H7N9 avian influenza outbreak in China. They determined that the mean bird to human transmission parameter to be 3.15×10^{-5} [46]. After finding the optimal parameter set, we discovered that the reproduction number with the best fitted parameters is $\mathcal{R}_0 = 0.1768$. When compared to reproduction estimates for H5N1 in poultry, we notice that our value for $\mathcal{R}_b = 1.016$ is smaller than the previously estimated reproduction numbers for in-poultry transmission. For instance, in the H5N1 outbreak in Thailand, the in-poultry reproduction number was estimated to be 1.27 [47]. We further note that our optimization procedure was highly sensitive to the initial conditions for the human and chicken populations. Thus, we fixed the initial conditions for the human and chicken populations as $S(0) = 19220$ and $S_b(0) = 63505$ (units 10^5). We assumed that the entire population where the data was obtain to be susceptible to the H5N1 avian influenza virus. If we considered the susceptible population to be just the farmers (or poultry handlers) that may have resulted in a higher $\mathcal{R}_0$ estimate with this framework. Therefore, this approach appears to underestimate $\mathcal{R}_0$.

Although our estimate is inconsistent with the estimates found the literature, we must acknowledge that the data used related to several different countries while most studies use data from a single country. We must be cautious when using fitting to the cumulative data and applying the sum of square errors to test the goodness of fit. Since the sum of squares relies on the assumption of independent and identically distributed errors, our deterministic model may lead to an over estimation of precision. To satisfy these assumptions i.i.d., we fitted the model to the incidence data (Figure 7) and obtained parameter values that are similar to the fitted parameters for the cumulative data. We may consider a stochastic model in the future since they account for real variability and they can handle uncertainty more easily than deterministic models [33].

As described in this study, it is important to know what type of data are available when estimating the reproduction number for H5N1. We observed about 20% difference among the reproduction values when using transmission chain data versus when we had cumulative number of cases, deaths, and secondary cases data. Given what is known about each method to under or over estimate, we can conclude that $\mathcal{R}_0 \approx 0.2$. The likelihood analysis pertained more to isolated clusters and transmission chain sizes for data accumulated in different regions while data fitting approach relied on cumulative data. Therefore, there exist more bias in one of the estimates due to the type of data being used. We must be aware when it is appropriate to apply these methodologies because the wrong application could lead to a severe over/under estimation of the value of $\mathcal{R}_0$. While collecting data for H5N1 presents a challenge, the understanding of the mechanics behind the potential human to human transmission, and the estimation of $\mathcal{R}_0$ play a vital role in understanding the risk of an H5N1 epidemic.

Acknowledgments

Conflict of interest

All authors declare no conflicts of interest in this paper.

References

1. M. D. R. van Beest Holle and A. Meijer, Human to human transmission of avian influenza /h7n7, the netherlands, *Euro Surveill.*, **10** (2005), 264–268.

2. K. Ungchusak, Probable person-to-person transmission of avian influenza a (h5n1), *New Engl. J. Med.*, **352** (2005), 333–340.

3. I. Kandum, Three indonesian clusters of h5n1 virus infection in 2005, *New Engl. J. Med.*, **355** (2006), 2186–2194.

4. S. Herfst, Airborne transmission of influenza a/h5n1 virus between ferrets, *Science*, **336** (2012), 1534–1541.

5. A. J. Hay, The evolution of human influenza viruses, *Philos. T. R. Soc. B*, **356** (2001), 1861–1870.

6. Avian influenza (2015).

7. Centers for Disease Control and Prevention, Highly pathogenic avian influenza a (h5n1) virus, (2015).

8. Y. Yang, Detecting human to human transmission of avian influenza a (h5n1), *Emerg. Infect. Dis.*, **13** (2005), 1348–1353.

9. J. K. Taubenberger and D. M Morens, The pathology of influenza virus infections, *Annu. Rev. Pathol.*, **3** (2005), 499–522.

10. Mathematical model, *Infect. Dis.*, (2015).

11. J. Hyman and L. Jia, An intuitive formulation for the reproductive number for the spread of disease in heterogeneous populations, *Math. Biosci.*, **167** (2000), 65–86.

12. Y. Xiao, X. Sun and S. Tang Transmission potential of the novel avian influenza a(h7n9) infection in mainland china, *J. Theor. Biol.*, **352** (2014), 1–5.

13. N. Tuncer and M. Martcheva, Modeling seasonality in avian influenza h5n1, *J. Biol. Syst.*, **21** (2013), 1–27.

14. G. Chowell, L. Simonsen, S. Towers, et al., Transmission potential of influenza h7n9 february to may 2013 china, *BMC Medicine*, **11** (2013), 1–13.

15. M. van Boven, M. Koopmans, M. Du Ry van Beest Holle, et al., Detecting emerging transmissibility of avian influenza virus in human households, *PLoS Comput. Biol.*, **3** (2007), 1–9.

16. M. E. Woolhouse and S. Gowtage-Sequeria, Host range and emerging and reemerging pathogens, *Emerg. Reemerg. Pathog.*, **11** (2005), 1842–1847.

17. S. Blumger and J. Lloyd-Smith, Inference of $\mathcal{R}_0$ and transmission heterongeneity from the size distribution of stuttering chains, *PLoS Comput. Biol.*, **9** (2013), 1–17.

18. S. Blumberg, S. Funk and J. R. Pulliam, Detecting differential transmissibilities that affect the size of self-limited outbreaks, *PloS One*, **10** (2014), 1–14.

19. S. Blumberg and J. Lloyd-Smith, Comparing methods for estimating $\mathcal{R}_0$ from the size distribution of subcritical transmission chains, *Epidemics*, **5** (2013), 131–145.

20. V. Pitzer, Little evidence for genetic susceptibility to influenza a from family clustering data, *Emerg. Infect. Dis.*, **13** (2007) 1074–1076.

21. S. J. Olsen, Family clustering of avian influenza a (h5n1), *Emerg. Infect. Dis.*, **11** (2005), 1799–1801.

22. T. Harris, The Theory of Branching Process, Dover, 2002.

23. I. Dumitriu, J. Spencer and C. Yan, Branching processes with negative offspring distribution, *Ann. Comb.*, **7** (2003), 35–47.

24. J. Lloyd-Smith, S. Schreiber, P. Kopp, et al., Superspreading and the effect of individual variation on disease emergence, *Nature*, **438** (2005), 355–359.

25. M. Dwass, The total progeny in a branching process and a related random walk, *J. Appl. Probab.*, **6** (1969), 682–686.

26. World Health Organization, Cumulative number of confirmed human cases of avian influenza a (h5n1) reported to who (2015).

27. S. Iwami, Y. Takeuchi and X. Liu, Avian flu pandemic: Can we prevent it?, *J. Theor. Biol.*, **257** (2009), 181–190.

28. W. H. O. Commmittee, The writing committee of the who consultation on human influenza a/h5 avian influenza a (h5n1) infection in humans, *New Engl. J. Med.*, **353** (2005), 1374–1385.

29. Food and Agriculture Organization of United Nations, Animal production and health division, (2015).

30. Centers for Disease Control and Prevention, Avian influenza a (h7n9) virus, (2014).

31. C. Hayden, Transmission of avian influenza viruses to and between humans, *J. Infect. Dis.*, **192** (2005), 1311–1314.

32. L. A. Reperant, T. Kuiken and A. D. Osterhaus, Influenza viruses, *J. Hum. Vaccin. Immunother.*, **8** (2012), 7–16.

33. A. A. King, M. Domenech de Cells, F. M. G. Magpantay, et al., Avoidable errors in modelling of outbreaks of emerging pathogens with special reference to ebola, *P. Roy. Soc. B*, **282** (2015), 143–151.

34. M. C. Eisenberg, S. L. Robertson, J. H. Tien, Identifiability and estimation of multiple transmission pathways in cholera and waterborne disease, *J. Theor. Biol.*, **324** (2013), 84–102.

35. N. Meshkat, M. Eisenberg and J. DiStefano, An algorithm for finding globally identifiable parameter combinations of nonlinear ode models using grobner bases, *Math. Biosci.*, **222** (2009), 61–72.

36. M. C. Eisenberg and M. A. L. Hayashi, Determining identifiable parameter combinations using subset profiling, *Math. Biosci.*, **256** (2014), 116–126.

37. N. D. Evans, L. J. White, M. J. Chapman, et al., The structural identifiability of the susceptible infected recovered model with seasonal forcing, *Math. Biosci.*, **194** (2005), 175–197.

38. H. Kelejian, Random parameters in a simultaneous equation framework: Identification and estimation, *Econometrica*, **42** (1974), 517–527.

39. H. Maio, X. Xia, A. Perelson, et al., On identifiability of nonlinear ode models and applications in viral dynamics, *SIAM Rev. Soc. Ind. Appl. Math.*, **53** (2011), 3–39.

40. N. Meshkat, C. Anderson, S. J. Rd, Alternative to ritt's pseudodivision for finding the input-output equations of multi-output models, *Math. Biosci.*, **239** (2012), 117–123.

41. G. S. Bellu, M. Audoly and S. D'Angio, Daisy: A new software tool to test global identifiability of biological and physiological systems, *Comput. Meth. Prog. Bio.*, **88** (2007), 52–61.

42. A. Raue, J. Karlsson and M. Jirstrand, Comparison of approaches for parameter identifiability analysis of biological systems, *Bioinformatics*, **30** (2014), 1440–1448.

43. H. T. Banks, J. E. Banks, C. Jackson, et al., Modeling the east coast akalat population: Model comparison and parameter estimation, Center for Research in Scientific Computation Report, (2013), 1–39.

44. A. Raue and C. Kreutz, Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood, Oxford University Press, **25** (2009), 1923–1929.

45. N. M. Ferguson and C. Fraser, Public health risk from the avian h5n1 influenza epidemic, *Science*, **304** (2004), 968–969.

46. Y. H. Hsieh, J. Wu, J. Fang, et al., Quantification of bird to bird and bird to human infections during 2013 novel h7n9 avian influenza outbreak in china, *Lancet*, **383** (2014), 541–548.

47. N. Marquetoux, M. Paul, S. Wongnarkpet, et al., Estimating spatial and temporal variations of the reproductive number for highly pathogenic avian influenza h5n1 epidemic in thailand, *Prev. Vet. Med.*, **106** (2012), 143–151.