

RESEARCH ARTICLE

Open Access



Estimating disease burden of a potential A(H7N9) pandemic influenza outbreak in the United States

Walter Silva^{1*}, Tapas K. Das¹ and Ricardo Izurieta²

Abstract

Background: Since spring 2013, periodic emergence of avian influenza A(H7N9) virus in China has heightened the concern for a possible pandemic outbreak among humans, though it is believed that the virus is not yet human-to-human transmittable. Till June 2017, A(H7N9) has resulted in 1533 laboratory-confirmed cases of human infections causing 592 deaths. The aim of this paper is to present disease burden estimates (measured by infection attack rates (IAR) and number of deaths) in the event of a possible pandemic outbreak caused by human-to-human transmission capability acquired by A(H7N9) virus. Even though such a pandemic will likely spread worldwide, our focus in this paper is to estimate the impact on the United States alone.

Method: The method first uses a data clustering technique to divide 50 states in the U.S. into a small number of clusters. Thereafter, for a few selected states in each cluster, the method employs an agent-based (AB) model to simulate human A(H7N9) influenza pandemic outbreaks. The model uses demographic and epidemiological data. A few selected non-pharmaceutical intervention (NPI) measures are applied to mitigate the outbreaks. Disease burden for the U.S. is estimated by combining results from the clusters applying a method used in stratified sampling.

Results: Two possible pandemic scenarios with $R_0 = 1.5$ and 1.8 are examined. Infection attack rates with 95% C.I. (Confidence Interval) for $R_0 = 1.5$ and 1.8 are estimated to be 18.78% (17.3–20.27) and 25.05% (23.11–26.99), respectively. The corresponding number of deaths (95% C.I.), per 100,000, are 7252.3 (6598.45–7907.33) and 9670.99 (8953.66–10,389.95).

Conclusions: The results reflect a possible worst-case scenario where the outbreak extends over all states of the U.S. and antivirals and vaccines are not administered. Our disease burden estimations are also likely to be somewhat high due to the fact that only dense urban regions covering approximately 3% of the geographic area and 81% of the population are used for simulating sample outbreaks. Outcomes from these simulations are extrapolated over the remaining 19% of the population spread sparsely over 97% of the area. Furthermore, the full extent of possible NPIs, if deployed, could also have lowered the disease burden estimates.

Keywords: Influenza, Influenza a virus -H7N9 subtype, Agent-based simulation model, Cluster analysis, Sampling studies

Background

A(H7N9) has infected humans in China in four waves, spring 2013, winter – spring of 2013–14, 2014–15, and 2015–2016. The fifth wave is currently in progress. As of June 2017, a total of 1533 laboratory-confirmed cases of A(H7N9) infections have been recorded in China causing 592 deaths [1]. The fifth (ongoing) wave has by far

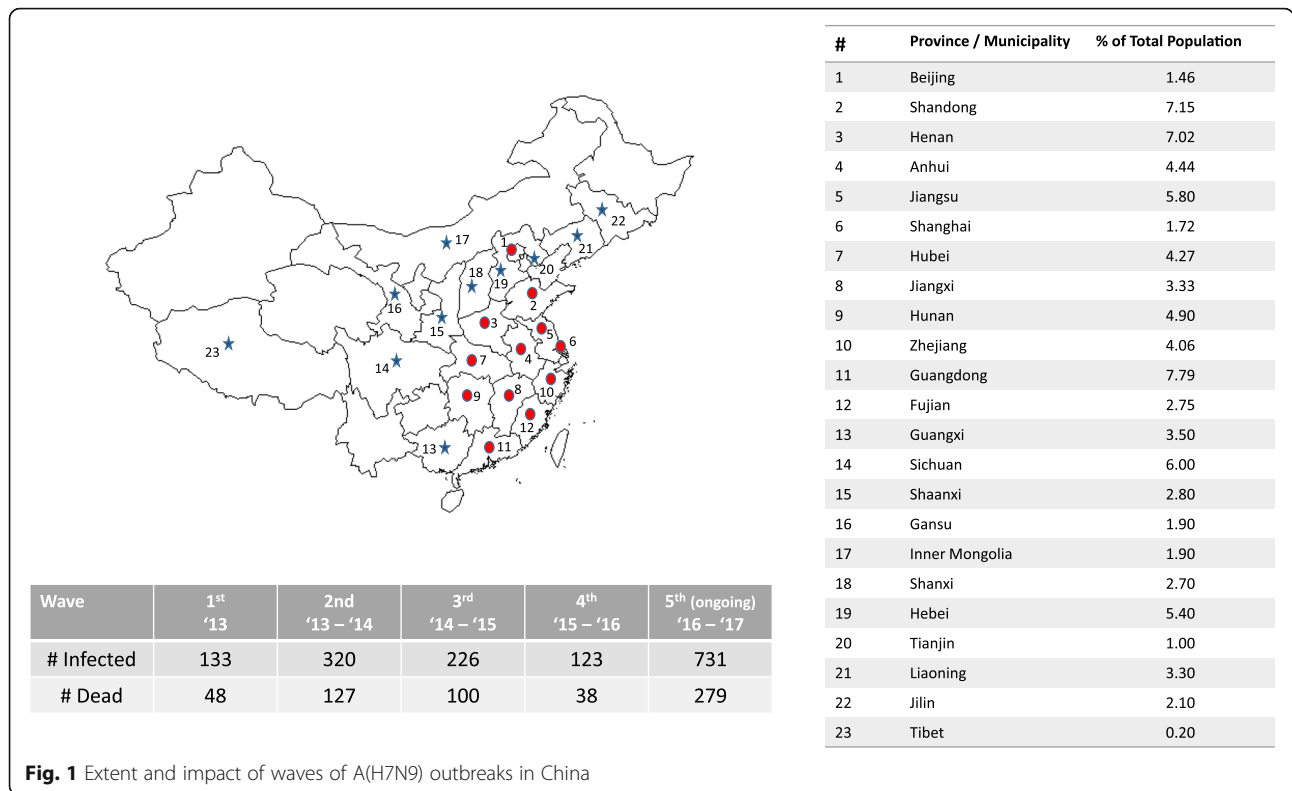
been the most widespread covering 23 provinces compared to 12 affected provinces till the fourth wave. It has infected almost the same number of people as the total of all four previous waves combined. Figure 1 depicts the outbreak locations and the numbers of infected and dead, for which information was collected from all applicable WHO reports on A(H7N9). We note that the affected provinces are relatively densely populated regions of China with over 84% of the population. The map in Fig. 1 was generated using the *mapdata* library from *R* software. Though most of the infections are

* Correspondence: silvasotillo@mail.usf.edu

¹Department of Industrial and Management System Engineering, University of South Florida, Tampa, FL 33620, USA

Full list of author information is available at the end of the article





known to be isolated cases, exceptions were noted where human-to-human transmission may have occurred. For example, there were at least 16 clusters of three infected family members and one cluster of two infected family members [2]. However, there is still lack of sustained evidence of human-to-human transmission [1].

A similar situation existed during the years 2003–2009 when the experts believed that a potential pandemic outbreak could be triggered by the H5N1 strain of the influenza virus. As during that period, H5N1 virus infected a total of 468 people with 282 deaths in 15 countries [3]. These numbers were updated in 2015 to 844 infected and 449 deaths [4]. Reports from critical examination of the impact of potential H5N1 outbreaks were presented to the literature in years 2005 [5] and 2006 [6]. Unexpectedly, however, instead of H5N1, the A(H1N1)pdm09 strain caused a worldwide influenza pandemic in 2009. This produced 60.8 million infections and 12,469 deaths in the U.S. alone [7].

Experts fear that A(H7N9) could become human-to-human transmittable and cause a pandemic. Hence, from a public health preparedness standpoint, it is essential to be able to assess the possible impact (disease burden) of a A(H7N9) pandemic. This paper addresses this need by developing a general methodology and applying it on a specific case concerning a pandemic covering all 50 states of the U.S.

Study of the data gathered from A(H7N9) infections reported in [8–14] present some of the characteristic epidemiological parameters for the virus. We have used these parameters in our disease burden estimation model. We note though that since A(H7N9) is not yet human-to-human transmittable, the actual parameters in the event of a pandemic may be different. Consequently, the true outcome of a pandemic may differ from the results presented in our paper. However, our AB simulation model and the estimation methodology are not affected by this limitation. As the new parameters are available for human transmittable scenarios, the model can be rerun and burden estimates can be refined. A comparison of A(H7N9) parameters with those for H5N1, both obtained from animal-to-human transmission cases, is shown in Table 1 [15]. A(H7N9) has been found so far to cause mild to severe disease in humans. In birds, generally, A(H7N9) has been low pathogenic (i.e., it did not cause clinical disease). However, recent observations suggest that the A(H7N9) virus has undergone some changes and it may increase its pathogenicity [16].

In what follows, we explain our methodology for estimating disease burden (measured by IAR and number of deaths) on the U.S. assuming that A(H7N9) becomes human-to-human transmittable and causes a pandemic. Our method is founded on an agent-based simulation

Table 1 ^bComparative parameters for H5N1 and A(H7N9)

Characteristic	H5N1	A(H7N9)
Incubation (days)	3.3 ± 1.5	3.1 ± 1.4
Latent Period (days)	2.15	< 3
Fatality risk	70%(China)	32%(China)
^a Admission to death	5.7 days	12 days
^a Admission to discharge	18.7 days	41.7 days
^a Median Age	26	62
^a Poultry exposure	71%	75%

^aPresented for information only; not used in our model

^bThe numbers are obtained from [15]

model that replicates the dynamics of the social and viral behavior during a pandemic. The simulation model being computing intensive, we apply it selectively on a few sample states in the U.S. for this purpose, all 50 states of the U.S. were first subdivided into smaller sub-groups using a clustering technique. Sample states for simulation are then chosen from each sub-group. A statistical method is used for calculating overall disease burden from the sampled data.

Methods

Agent-based (AB) simulation is a useful tool to emulate events that might occur in the future and thus support policy makers to prepare measures to address such events. Our AB simulation model incorporates four basic components: demographic information, human behavior, epidemiological characteristics of the virus, and non-pharmaceutical interventions (NPIs) to mitigate pandemic impact.

The demographic information includes the household composition (age, sex, work, and parental status), schools, workplaces, and communities. The human behavior describes the contact process in the mixing groups, compliance to quarantine, and isolation, and travel behavior. The epidemiological characteristics include the disease natural history, parameters affecting the force of infection, basic reproduction number (R_0), and the fatality rate. Selection of R_0 values was guided by similar studies (See “Model Validation” section). The NPIs are explained in details in “Non-Pharmaceutical Intervention” section.

Our AB simulation model is quite granular and uses a large computer memory. In its present form, the model is limited to run with up to five million people in an outbreak region. Given this limitation, in order to implement our methodology for a countrywide outbreak in the U.S. with 307 million people, we used a sampling approach. We divided the group of all 50 states of the U.S. into smaller clusters (sub-groups) of states. The clustering technique used to divide the states in smaller

sub-groups considered urban population size and density (pop/mile²) as attributes of the states. These attributes are highly correlated to the spread of human influenza virus. We applied the simulation model in a few selected states from each sub-group of states. Results from sample states are used to obtain disease burden for the sub-groups, which are then combined to estimate disease burden for the whole country.

AB simulation model

We used a previous version of our AB simulation model that was presented in [17–21]. We modified the model by incorporating a more detailed method for estimating the probability of infection for a susceptible using the measure of *force of infection* as defined in [5]. The force of infection measures the total daily viral load gathered by a susceptible individual from the infected contacts in the mixing groups. (More details for force of infection are presented in “Infection Model” section).

The model mimics the contact process and tracks each individual in an outbreak region using their scheduled hourly movements within the mixing groups: households, places (schools and workplaces), and community locations.

The model begins by generating the simulated individuals according to the U.S. census and demographic data [22] that gives population attributes including age, gender, and occupational status (school/work). Thereafter, we generate the households based on their composition (characterized by the number of adults and children) in the U.S. (see Tables 2 and 3). We randomly assign each individual to a household while maintaining the average household composition. We then generate the schools, workplaces, and other community locations. We assign each individual a daily (hour by hour) schedule, chosen randomly from a set of alternative schedules based on their attributes. The schedules also vary between weekdays and weekends. Simulation begins on the day when one or more infected people are introduced to the

Table 2 U.S. Household composition per census 2014

Household Composition		
# adults	# children	Proportion
1	0	28%
1	1	4%
2	0	31%
1	2	4%
2	1	13%
1	3	1%
2	2	13%
1	4	1%
2	3	6%

Table 3 U.S. Age distribution per census 2014

Age distribution of household members			
Children		Adults	
Age range	Proportion	Age range	Proportion
[0 – 5]	24%	[23 – 29]	16%
[6 – 9]	23%	[30 – 64]	67%
[10 – 14]	25%	[65+]	17%
[15 – 17]	13%		
[18 – 22]	15%		

region (referred to as day 1). Simulation model tracks hourly movements of each individual (susceptible and infected) throughout the day, and records for each susceptible the number of contacts with infected at each location. At the end of each day, the model calculates for each susceptible the total amount of viral load ingested (force of infection) from all contacts during that day. The severity of infection of the infected contacts and the place of contact (household, schools, workplaces, and community locations) play critical roles in determining the force of infection. This is used in calculating the probability of infection. The model updates the infection status of all individuals to account for new infections and disease progressions of the already infected ones. The key components of the AB model are described next.

Disease natural history

We adopted a similar disease natural history model was used in previous studies for other influenza A virus strains [5, 19]. Though an accurate disease natural history for human-to-human transmittable A(H7N9) virus is not yet known, we were guided by the disease natural history of other influenza viruses that have already caused pandemic outbreaks.

Figure 2 presents a schematic for the disease natural history. An infected individual simultaneously begins a latency and an incubation period (based on parameters given in Table 1). The individual displays symptoms

(unless asymptomatic) at the end of the incubation period, and becomes infectious after the latent period is complete. Following the infectiousness period, an infected either recovers or dies. The numerical values of the parameters characterizing the various elements of the disease natural history (e.g., length of incubation period, death rate) are given in Tables 1 and 4. Those who recover are considered to attain sterilizing immunity to further infections [23]. We made this assumption as we are not aware of the immune response to the A(H7N9) virus. Also, as no estimate is available for the proportion of asymptomatic cases for A(H7N9) infections, we assumed it to be 50%. Same proportion of asymptomatic cases was considered in previous studies [5]. The duration of infectiousness for each case is considered to be guided by a lognormal random distribution.

Infection model

An individual i is considered to accumulate force of infection λ_i in his/her home, places, and community locations. It is calculated using the following expression as given in [5].

$$\lambda_i = \sum_{k|h_k=h_i} \frac{I_k \beta_h \kappa(t-\tau_k) \rho_k [1 + C_k(\omega-1)]}{n_i^\alpha} + \sum_{i,k|l'_k=l'_i} \frac{I_k \beta_p^j \kappa(t-\tau_k) \rho_k [1 + C_k(\omega \psi_p^j(t-\tau_k)-1)]}{m_i^j} + \frac{\sum_k I_k \zeta(a_i) \beta_c \kappa(t-\tau_k) \rho_k f(d_{i,k}) [1 + C_k(\omega-1)]}{\sum_k f(d_{i,k})} \tag{1}$$

The first component in (1) expresses the force experienced by susceptible individual i at home from other infected household members k . The second component captures the force experienced at places (schools and workplaces) when a susceptible i is in the same place as infected k . The third component considers the force of infection gathered from all infected members of the

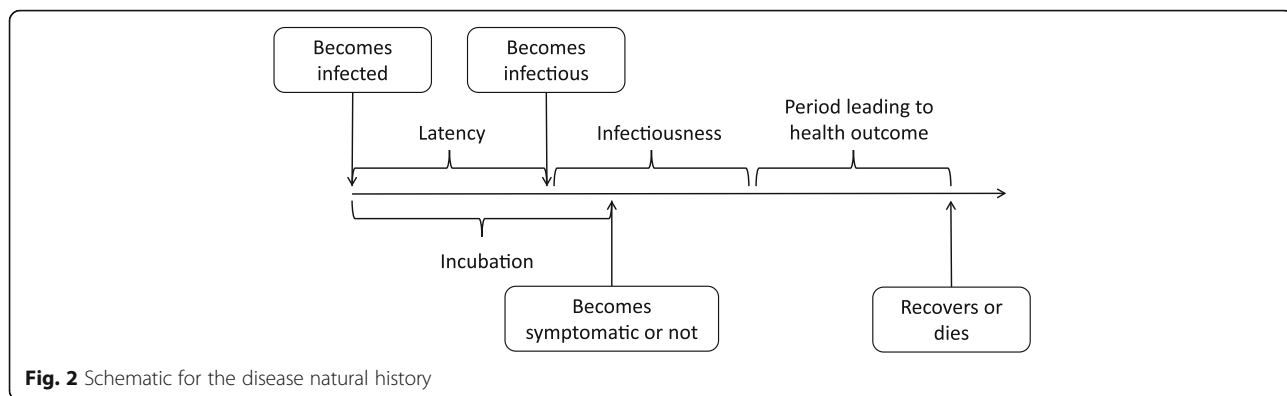


Fig. 2 Schematic for the disease natural history

Table 4 Parameter values used in the AB simulation model

Parameter	Values
β_h	0.47/day (for $R_0 = 1.8$); 0.39/day (for $R_0 = 1.5$)
β_p^j	0.94/day for schools and 0.47/day for workplaces (for $R_0 = 1.8$); 0.78/day for schools and 0.39/day for workplaces (for $R_0 = 1.5$)
β_c	0.075/day (for $R_0 = 1.8$); 0.06/day (for $R_0 = 1.5$);
$\kappa(t - \tau_k)$	lognormal distribution with: $\delta = -0.72$ and $\gamma = 1.8$
$\psi_p^j(t - \tau_k)$	0.2 (for schools) and 0.5 (for workplaces) only when the elapsed time since the onset of infection is greater than the latent period 0.25 days; the value of ψ_p^j is 0 otherwise
$f(d_i, k)$	$\frac{1}{[1 + \frac{d_i d_k^a}{a}]^b}$, Z with $a = 35$ km and $b = 6.5$
$\zeta(a_i)$	$\zeta(a_i) = 100\%$ if age $\in [20 - 65]$, 75% if age $\in [15 - 20]$ and $[65 - 70]$, 50% if age $\in [10 - 15]$ and $[70 - 75]$, 25% if age $\in [5 - 10]$ and $[75 - 85]$, 0% if age $\in [0 - 5]$
ρ_k	1
C_k	1 if individual k is a severe infection, 0 otherwise
ω	2
n_i^a	obtained from the households generated by the model
a	0.8
m_i^j	obtained from the population and places generated by the model
death rate	38.61%
symptomatic	50%

community visited by susceptible i . The parameters of (1) are defined in Table 5. λ_i is calculated at the end of each day for all susceptible i and the probability of infection is obtained as $1 - \exp^{-\lambda_i}$. It is assumed that if not infected by the end of a day, λ_i is reset to zero. That is, the force of infection is assumed to not accumulate from 1 day to the next.

Non-pharmaceutical intervention

We considered isolation of symptomatic infected individuals at home for a specific duration with isolation compliance of 53% for adult workers and 57.5% for non-workers [19]. A compliant infected individual is assumed to stay home all day. We consider an isolation threshold of 1 day (that is, on average an individual diagnosed with infection does not begin isolation until the day after) and isolation duration of 7 days.

We also considered household quarantine that restricts the movement of susceptible household members when one or more members are infected. Household quarantine parameters were considered same as for individual isolation. Children were assumed to fully comply with isolation. A partial school closure approach is considered. A classroom is closed when a threshold of newly infected children in the classroom is reached. A

Table 5 Parameters used in calculating the force of infection

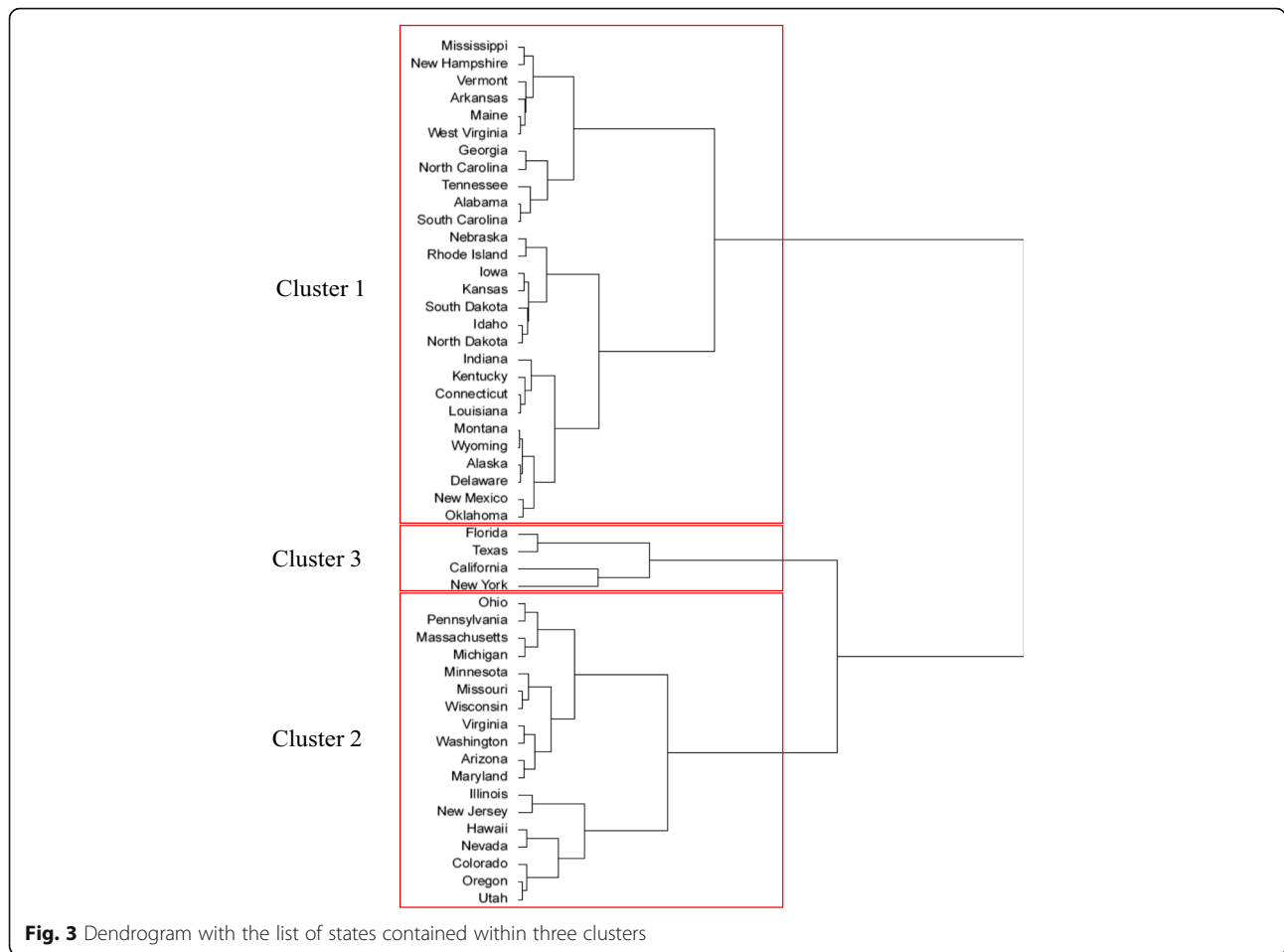
Parameter	Description
I_k	1 if infected and 0 otherwise
β_h	household transmission parameter
β_p^j	place transmission parameter
β_c	community transmission parameter
$\kappa(t - \tau_k)$	infectiousness at time $(t - \tau_k)$ since infection
$\psi_p^j(t - \tau_k)$	factor by which within-place contact rates change for symptomatic severe infection (reflecting sickness-induced absenteeism)
$f(d_i, k)$	a function of distance $d_{i,k}$ between individuals i and k
$\zeta(a_i)$	relative travel-related contact rate of an individual of age a_i
ρ_k	relative infectiousness of individual k
C_k	1 if infection is severe, 0 for mild (asymptomatic)
ω	2, infectiousness of a severe infection relative to a mild one
n_i^a	number of people in the household of individual i
a	power that determines the scaling of household transmission rates with household size
m_i^j	number of people in the place type j

threshold for the number of closed classrooms is used to close a school. We used a threshold value of one for both the classroom closure and school closure, and 21 days for the length of school closure. Workplace closure strategy was similar to that of school closure, with each department/group treated like a classroom. The thresholds were: three cases to close a department, 30% of the departments closed to close a workplace, and 7 days for closure duration.

Clustering technique

We adopted a commonly used hierarchical clustering (grouping) technique [24]. The clustering technique forms sub-groups of states such that the chosen attributes of the states within a sub-group are similar to one another and at the same time dissimilar from the attributes in other sub-groups. Higher the level of similarity within a sub-group and dissimilarity between sub-groups, better is the sub-group formation. A clustering method is hierarchical when it creates a set of nested sub-groups that are organized as a tree. At the lowest level of such a tree, each state is separate sub-group, and at the highest level, all states belong to one sub-group (see for example, Fig. 3). The user decides the number of sub-groups (clusters) to consider based on a desired level of similarity from the modeling application.

As stated earlier, we chose urban population size and population density (per square mile) as the two attributes of the states in the U.S. to be used by the clustering technique. We focused on urban population for two



reasons: 1) urban areas are more prone to pandemic spread, and 2) states with large population exceeded our simulation model capacity.

The attribute values were first normalized by subtracting from each the corresponding mean and dividing by the corresponding standard deviation. Such a normalization is generally recommended when the numerical values of one or more of the attributes vary significantly within the set. For example, in the U.S., the urban population sizes of the states vary between 1.62 and 36.8 millions, while the urban density ranges between 0.0019 and 0.0043 millions/sq. mile.

The clustering method begins by assigning each state to a separate cluster resulting in 50 clusters at start. Then it calculates the Euclidean distance between the attribute vectors (size, density) of all cluster pairs. (Euclidean distance is the length of the straight line joining two points in the two dimensional space representing the state pair.) The method then identifies the cluster pair with the smallest distance and combines them into one cluster. This reduces the number of original clusters by one, and for the combined cluster, its attribute vector is assigned as the centroid of the attribute vectors of the

constituent states. At the next step, the distances between the clusters are updated, and the process repeats until all clusters are combined into one single cluster. A line diagram (called dendrogram; see, for example, Fig. 3) is then used in selecting an acceptable number of clusters considering a desired level of similarity within each cluster. We implemented the clustering method by first using *preprocess* function within the *Caret* package of *R* library to normalize the data. Thereafter, we used the *predict*, *dist*, and *hclust* functions within the *stats* package of *R* library for the remaining steps of the hierarchical clustering method.

Disease burden calculation from estimates stratified by cluster and age-groups

We first calculated the mean and C.I. of the infection attack rates (IAR), for all three age-groups (indexed by *a*) in all sampled states (indexed by *i*) within each cluster (indexed by *j*), using replicated results of the AB simulation model with different seeds for the random variables. The 100(1 - α)% C.I. for IAR was calculated as $\hat{p}_{ij}^a \pm t_{\alpha/2, n-1} \frac{s}{\sqrt{n}}$, where \hat{p}_{ij}^a denotes the mean IAR, *n*

represents the number of simulation replicates, and s represents the standard deviation of the replicated IAR estimates. Mean IAR values for the clusters were obtained by combining the values of \hat{p}_{ij}^a into \hat{p}_j^a using the expression below [24].

$$\hat{p}_j^a = \frac{1}{n_j^a} \sum_{i=1}^{S_j} n_{ij}^a \hat{p}_{ij}^a, \tag{2}$$

where S_j denotes the total number of selected states that were simulated within cluster j , and n_{ij}^a represents the size of the urban population in state i within cluster j for age-group a . Note that $n_j^a = \sum_{i=1}^{S_j} n_{ij}^a$ is the total urban population for the selected states in cluster j for age-group a . The $100(1 - \alpha)\%$ C.I. on the IAR estimate for each age group within a cluster was obtained as $\hat{p}_j^a \pm t_{\alpha/2, n-1} \frac{s_j^a}{\sqrt{n}}$. The pooled standard deviation s_j^a was calculated from the estimates of standard deviation for each selected state within a cluster as the square root of $[(n-1)s_{j1}^a{}^2 + \dots + (n-1)s_{jk}^a{}^2] / [k(n-1)]$, where k is the number of selected states in cluster j , n is the number of simulation replicates, and s_{jk}^a is the standard deviation of replicated IAR estimates for age-group a in state k within cluster j . Hereafter, we combined the IAR estimates from all clusters into one value for each age-group using

$$\hat{p}^a = \frac{1}{N^a} \sum_{j=1}^C N_j^a \hat{p}_j^a, \tag{3}$$

where C denotes the number of clusters, N_j^a denotes the total population of all the states in cluster j in age-group a , and N^a denotes the total population in the country in age-group a . It may be noted that the estimate \hat{p}^a has a variance V^a from two sources of variability: 1) due to sampling: a sample population from each cluster is used

to estimate \hat{p}_j^a , which is then assumed to hold good for the whole cluster population, 2) due to simulation based estimation: \hat{p}_{ij}^a 's are obtained from \hat{p}_{ij}^a , which are estimated from simulation model with inherent variability. It can be argued that these two sources of variability are independent. We obtained the variance due to sampling V_1^a as follows [25],

$$V_1^a = \frac{1}{(N^a)^2} \sum_{j=1}^C (N_j^a)^2 \left(\frac{N_j^a - n_j^a}{N_j^a} \right) \left(\frac{\hat{p}_j^a (1 - \hat{p}_j^a)}{n_j^a - 1} \right). \tag{4}$$

The variance due to simulation V_2^a was obtained as the pooled variance from the variance estimates (s_j^a) of the three clusters as $V_2^a = [((n-1)s_1^a{}^2 + (n-1)s_2^a{}^2 + (n-1)s_3^a{}^2) / 3(n-1)]$, where n is number of simulation replicates per cluster. A $100(1 - \alpha)\%$ C.I. was calculated as $\hat{p}^a \pm t_{\alpha/2, n-1} \sqrt{V^a/n}$, where $V^a = V_1^a + V_2^a$. Finally, we obtained a single estimate of IAR (\hat{p}) for the whole U.S. across all age-groups $a \in \{1, 2, \dots, L\}$ using (3) and substituting in this equation N, N^a , and \hat{p}^a for N^a, N_j^a , and \hat{p}_j^a , respectively, and summing over $a = 1$ through L . The variance V on the overall IAR estimate was obtained by pooling variance values V^a from the three age-groups. A $100(1 - \alpha)\%$ C.I. on IAR was calculated as $\hat{p} \pm t_{\alpha/2, n-1} \sqrt{V/n}$. The number of deaths for each age group and also for the whole U.S. were calculated by applying the death rate on the corresponding numbers for infected persons.

Model validation

We validated our model by using it to replicate a H5N1 outbreak study for Southeast Asia [5]. Our choice of this validation approach was motivated by similarities

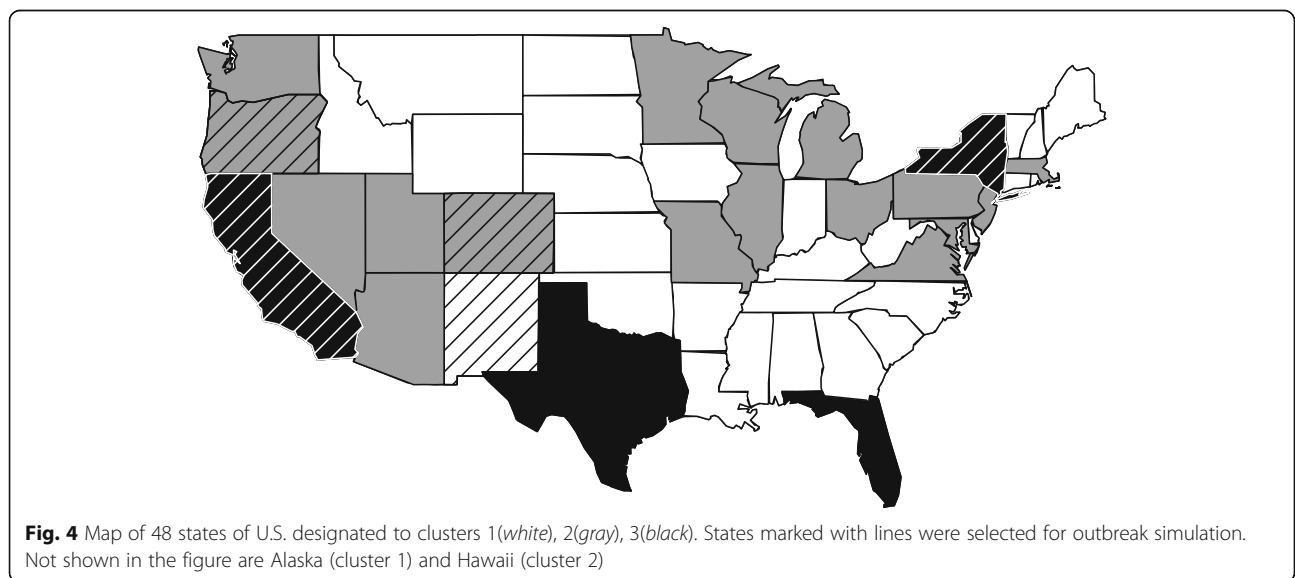


Table 6 Urban and Rural population distributions in the States selected for simulation

Region	Total pop (M)	Urban pop(%)	Urban pop (M)	Urban area(%)	Urban density pop/sq. mile
California	38.80	94.95	36.8	5.28	4304
Colorado	5.36	86.15	4.62	1.47	2836
New Mexico	2.09	77.43	1.62	0.68	1929
New York	19.75	87.87	17.35	8.68	4161
Oregon	3.97	81.03	3.22	1.15	2804

between the H5N1 and A(H7N9) virus strains. Both of these strains have still not acquired human-to-human transmission capability, but experts fear that they may mutate to that state. Also our modeling approach has much in common with that used in [5], and hence it provided an appropriate platform for validation. The study considered a population size of 85 million. We considered a subset of the population (5 M) and proportionately adjusted down the number of households, schools, workplaces, and community locations. As in [5], we considered two different cases of R_0 (1.5 and 1.8) values. We used ten replicates for each case, and did not deploy any NPIs since these were not used in [5]. For $R_0 = 1.5$, the average IAR is 34.58% with a standard deviation of 5.24, and 95% C.I. of [30.83–38.33]. The IAR reported in [5] for $R_0 = 1.5$ is 33%, which is within our C.I. Our corresponding results for $R_0 = 1.8$ are 55.7%, 8.16, and [49.86–61.54], respectively. The IAR reported in [5]

Table 7 Simulated population size and infection attack rates (IAR)

	Urban population (M)	Sample size (M)	IAR (\hat{p}_{ij}^a) ($R_0 = 1.5$)	IAR (\hat{p}_{ij}^a) ($R_0 = 1.8$)
< = 19 yrs				
California	9.7	1.32	0.3272	0.4197
Colorado	1.19	1.19	0.2782	0.3797
New Mexico	0.43	0.43	0.2461	0.3299
New York	4.18	1.21	0.3172	0.4176
Oregon	0.77	0.77	0.2572	0.3777
20 – 64 yrs				
California	22.4	3.04	0.1612	0.2113
Colorado	2.83	2.83	0.1431	0.1906
New Mexico	0.94	0.94	0.1385	0.1856
New York	10.62	3.06	0.1524	0.2007
Oregon	1.93	1.93	0.1419	0.1866
65 + yrs				
California	4.74	0.64	0.1607	0.2291
Colorado	0.59	0.59	0.1447	0.1926
New Mexico	0.25	0.25	0.1248	0.1615
New York	2.55	0.73	0.1589	0.2144
Oregon	0.51	0.51	0.1386	0.1872

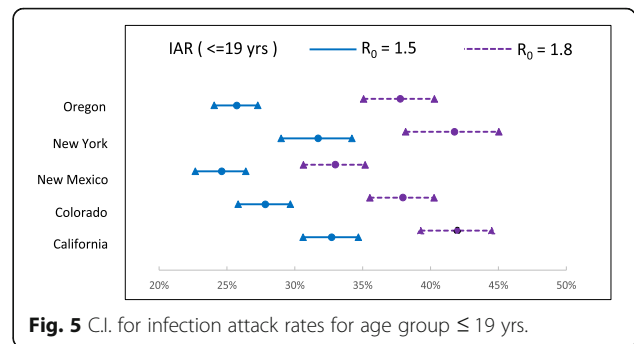


Fig. 5 C.I. for infection attack rates for age group ≤ 19 yrs.

is 50%. We note that in both cases of R_0 values, results reported in [5] lie in the lower half of our C.I.s.

Results

The dendrogram in Fig. 3 shows the outcome of clustering of the 50 states of the U.S. into sub-groups using population size and density as the characteristic features (attributes). Based on the selection criteria of the dendrogram, possible choices were either two or three clusters. We chose the three cluster option, which provided better similarity within the clusters; higher similarity was manifested by lower standard deviation of the attribute values within the clusters. We labelled the clusters with numbers 1, 2, and 3, which are composed of states with low, medium, and high values of the attributes, respectively. Figure 4 is a map of U.S. that depicts the cluster designation of all the states. For simulating sample outbreaks, we selected New Mexico from cluster 1, Colorado and Oregon from cluster 2, and California and New York from cluster 3. Selection of these five states was influenced by a recent paper [26] that presented disease burden estimates for seasonal influenza outbreaks in the U.S. However, choice of these particular states from the three clusters does not present a limitation of our study.

For outbreaks in California and New York with urban population sizes greater than five million, we selected a number of contiguous urban counties within each state with a cumulative population of up to five million. While for Colorado, Oregon, and New Mexico we simulated their total urban population, each being less than 5

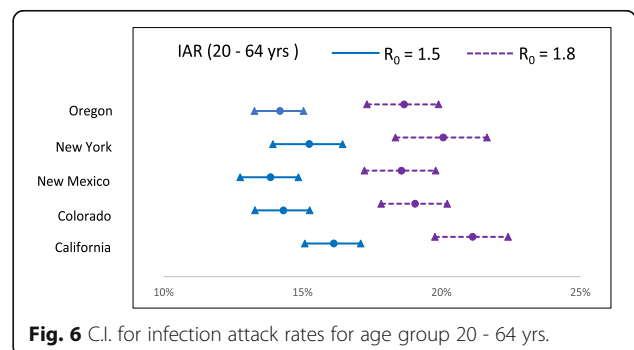


Fig. 6 C.I. for infection attack rates for age group 20 - 64 yrs.

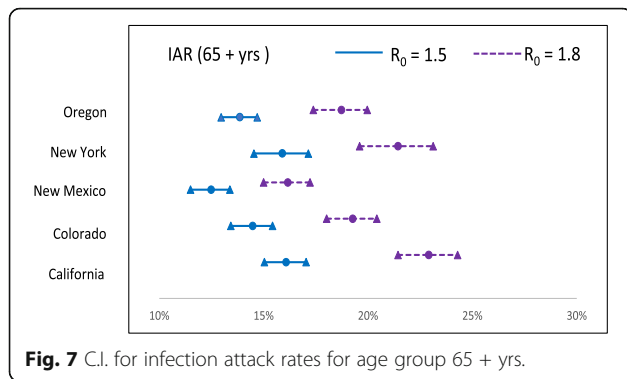


Fig. 7 C.I. for infection attack rates for age group 65 + yrs.

million. Our focus on urban population was guided by the fact that approximately 81% of the population of the selected states reside in dense urban regions constituting on average 3.4% of the land area (see Table 6). We used latest U.S. census data to extract information on households, workplaces, and schools. We implemented an ad-hoc non-pharmaceutical intervention (NPI) strategy comprising measures like isolation, quarantine, school and workplace closures. Pharmaceutical interventions (vaccines and antivirals) were not considered.

As shown in Table 7, the outbreak in the state of Colorado was simulated using the total urban population of 4.62 M as the sample size comprising 1.19 M for ages ≤ 19 , 2.83 M for ages 20–64, and 0.59 M for ages 65 and above. This approach was also used for New Mexico and Oregon. For California and New York, we adopted a proportional sampling approach. For example, California has 9.7 M people for age-group ≤ 19 years and 22.4 M and 4.74 M for age-groups 20–64 years and ≥ 65 years, respectively. The AB model used the family composition features from the U.S. census to randomly generate a total of $(9.7/36.8) \times 5M$ children, $(22.4/36.8) \times 5M$ adults up to age 64, and $(4.74/36.8) \times 5M$ adults 65 and above. Also using census data, our model generated a proportional number of households in a region and then populated each household following the average proportion of children and adults in various age-groups in U.S. families as presented in Tables 2 and 3. Thereafter, the model generated the places (schools and workplaces)

using census data in [25] and [27], and randomly assigned each individual to a place based on the age-group. Beyond households and places, the model also considered movements of the individuals in the community within the state for daily errands.

The AB model initiated each outbreak by introducing six infected individuals. The parameter values used to calculate the force of infection (λ_i) using (1) are shown in Table 4. To calculate the third component of λ_i , we assumed, for simplicity, that each day an individual (susceptible or infected who are not compliant with isolation) travel within or outside of their county of residence for errand or leisure. Travel related parameters used in our study were also used in [28].

The AB simulation model runs (with 10 replicates) in the selected states yielded the mean IAR values (\hat{p}_{ij}^a) as displayed in Table 7. We used the estimated values of \hat{p}_{ij}^a and n_{ij}^a to estimate \hat{p}_j^a using (2), the IARs per age-group within a cluster. These values were then combined to obtain estimate of IAR for each age-group (\hat{p}^a) in the whole U.S. Finally, IAR values for all age-groups were combined to obtain the overall IAR estimate (\hat{p}). Figures 5, 6 and 7 show the IARs and their C.I.s, which can be seen to be generally higher for states/clusters with higher population density.

Tables 8 and 9 show IARs and number of infected cases per age-group within each cluster. IAR estimates across all clusters within the age-groups (\hat{p}^a in (3)) and across all age-groups (\hat{p}), and the number of deaths are shown in Table 10. It can be observed in Table 10 that IAR for age-group ≤ 19 yrs. is approximately double that of for other two age-groups. Though no age-dependent virus characteristics were considered in the model, the increased IAR among younger population is the result of higher level of social interactions in schools. This can be attributed to the relatively short duration of school closure (21 days) in our NPI strategy.

It is also observed from Table 10 that the IAR for the two adult age-groups are similar. This is somewhat counterintuitive, as it may be expected that members of age-group 20–64 will have higher IAR resulting from higher work related social interactions. We believe that

Table 8 Infection attack rates (IAR) per cluster and age-group with 95% C.I.

	IAR (in % for $R_0 = 1.5$)			IAR (in % for $R_0 = 1.8$)		
	Cluster 1	Cluster 2	Cluster 3	Cluster 1	Cluster 2	Cluster 3
≤ 19 yrs	24.61 (22.67–26.55)	27.00 (25.17–28.83)	32.41 (29.98–34.84)	32.99 (30.63–35.35)	37.89 (35.31–40.47)	41.91 (38.74–45.08)
20–64 yrs	13.85 (12.76–14.94)	14.26 (13.29–15.23)	15.84 (14.66–17.02)	18.56 (17.23–19.89)	18.89 (17.61–20.17)	20.79 (19.24–22.34)
65 + yrs	12.48 (11.5–13.46)	14.18 (13.21–15.15)	16.01 (14.8–17.22)	16.15 (14.99–17.31)	19.01 (17.72–20.3)	22.39 (20.72–24.06)

Table 9 Number of infected cases per cluster and age-group with 95% C.I

	Number * (95% CI) R0 = 1.5			Number * (95% CI) R0 = 1.8		
	Cluster 1	Cluster 2	Cluster 3	Cluster 1	Cluster 2	Cluster 3
<=19 yrs	56.49 (52.04–60.95)	85.55 (79.75–91.35)	132.02 (122.1–141.94)	75.73 (70.31–81.15)	120.05 (111.89–128.21)	170.72 (157.79–183.65)
20–64 yrs	72.22 (66.55–77.88)	106.42 (99.18–113.67)	79.01 (73.12–84.9)	96.77 (89.84–103.71)	140.98 (131.41–150.55)	103.71 (95.96–111.45)
65+ yrs	16.16 (14.89–17.42)	26.15 (24.37–27.93)	23.67 (21.88–25.46)	20.91 (19.41–22.41)	35.06 (32.68–37.44)	33.10 (30.63–35.56)

*Numbers in 100,000

older people (65+) are accumulating viral load (force of infection) at a higher rate from age-group 20–64 by being at home with other infected family members.

The IAR estimates were compared with IAR estimates for other viruses as shown in Table 11. Simulation-based estimates of IAR for both H5N1 and A(H7N9) were found to be much lower than the field estimate for H1N1/2009. We conjecture that the lower IAR estimates are due to lower estimates of the force of infection, for which the parameters were estimated from only animal-to-human transmittable cases of outbreaks.

Discussion

Our paper is the first to estimate disease burden from A(H7N9) pandemic outbreak, hence we could not directly compare our results with other studies on A(H7N9). The disease parameter estimates used in our model were adopted from the recent reports on epidemiological studies of A(H7N9) [8, 15, 29, 30]. Other models (e.g., using differential equations) have been used to analyze A(H7N9) [30–32]. These models do not have the level of granularity offered by agent-based simulation models. However, such granularity comes with the cost of computation and memory usage, which resulted in our model capacity being limited to 5 million people per simulation run. We note that this limit can be increased with better computing hardware and

more efficient usage of memory. Among the five states that were selected for simulation, the size limit only applied to California and New York with urban population sizes much larger than 5 million. We used an age-based proportional sampling approach to select up to 5 million individuals from the urban areas.

A paper published in June 2016 [29] presented a comprehensive analysis of the laboratory-confirmed cases of A(H7N9) infection in mainland China. It offered renewed estimates for incubation period, fatality risk, hospital admission to death/discharge, median age, and poultry exposure. We note, however, that the parameter estimates that we used from an earlier study [15] do not differ significantly from those presented in [29]. Though it appears from the published data that A(H7N9) affects more people of higher age group, it is likely a function of the very high level of poultry exposure ($\geq 74\%$ [15]) for the older age group. Our AB model does not incorporate any age-dependent factor for calculating probability of infection. However, our model does consider age-based contact process, which in turn affects the infection probability.

We simulated outbreaks only in urban areas and extrapolated the results to the population in the remaining (rural) areas. Urban areas constitute on average 3.4% of the geographic area and approximately 81% of the population [22]. Hence, the disease burden estimates, for each

Table 10 Infection attack rate (IAR), number of infected cases and number of deaths per age-group U.S. with 95% C.I

	R0 = 1.5 (95% CI)			R0 = 1.8 (95% CI)		
	IAR(%)	# of Infected (million)	Death rate ^a	IAR(%)	# of Infected (million)	Death rate ^a
<= 19 yrs	28.74 (24.36–33.11)	27.41 (23.24–31.58)	11,094.72 (9405.4–12,783.77)	38.43 (33.7–43.15)	36.65 (32.14–41.16)	14,836.87 (13,011.57–16,660.22)
20–64 yrs	14.59 (12.05–17.12)	25.77 (21.29–30.24)	5631.32 (4652.51–6610.03)	19.33 (16.49–22.17)	34.15 (29.13–39.17)	7462.97 (6366.79–8559.84)
65+ yrs	14.29 (9.51–19.06)	6.60 (4.39–8.8)	5517.08 (3671.81–7359.07)	19.29 (13.94–24.64)	8.91 (6.44–11.38)	7447.97 (5382.23–9513.5)
Total U.S.	18.78 (17.09–20.48)	59.77 (54.38–65.16)	7252.30 (6598.45–7907.33)	25.05 (23.19–26.91)	79.70 (73.78–85.63)	9670.99 (8953.66–10,389.95)

^aRates per 100,000

Table 11 Comparison of infection attack rates among different influenza viruses

Description	H1N1 2009 (See Ref [7])	H5N1 (See Ref [6])	Seasonal Influenza (See Ref [33, 34])	A(H7N9) (Current Paper)
Data used	surveillance data from U.S. outbreak	simulated outbreak in U.S. and England	Surveillance data from U.S.	Simulated outbreak in U.S.
Method used	Extrapolation with Correction factors	AB simulation model	Proposed by CDC	AB simulation model and stratification
NPIs (school & workplace closure, case isolation)	yes (with antiviral)	yes (with vaccine and antivirals)	no	yes
Age-groups analysis	yes	yes	yes	yes
Estimated IAR	50%	28% for $R_0 = 1.7$ 34% for $R_0 = 2.0$	5% - 10% in adults 20% - 30% in children	18.78% for $R_0 = 1.5$ 25.05% for $R_0 = 1.8$

state, cluster, and age-group, as presented here, are likely to be upper bounds, since the rural areas are likely to yield less number of infections with less contacts and higher distances between individuals. Also, applications of vaccines and antivirals were not considered in our AB model, which increased the number of susceptible and the intensity of infection, respectively. Furthermore, the disease burden estimates could have been lowered by application of the full extent of NPIs.

For the NPI strategy implemented in our simulation model, we chose its parameters (e.g., length of school closure) somewhat arbitrarily. We note that these parameters could be optimized based on the virus and societal characteristics. In another study focused to assess in community resilience for influenza pandemic outbreaks, we tested NPIs with two different sets of parameters for a small A(H7N9) outbreak in a region with 1.1 million people in the state of Florida in U.S. [33]. Table 12 shows the parameters of these NPI strategies

and the corresponding IARs. Note that the strategy marked as NPI(1) is same as the strategy used in the study presented in this manuscript, and NPI(2) is the strategy that was recommended in [19].

Conclusions

Our AB simulation model has the following notable limitations. It does not assign specific geographic locations for the households, places, and community locations. As a result, we had to use estimated values for the distance between susceptible and infected ($f(d_{i,k})$ in Eq. (1)) in calculating the force of infection. Use of antivirals could have reduced the profile of infectiousness with a lower peak and shorter duration, in turn reducing the virus spread and the corresponding IAR. Also, we did not consider pre-existing immunity for any age-group in the population.

We reiterate the fact that A(H7N9) has not yet been found to be human-to-human transmittable. This paper

Table 12 NPI parameters

#	Measure	NPI (1)	NPI (2)
1	Global Threshold	10	10
2	Deployment delay	3 days	7 days
3	Case isolation threshold	1 day	1 day
4	Case isolation duration	7 days	10 days
5	Case isolation compliance for workers	75%	75%
6	Case isolation compliance for non-workers	84%	57%
7	Household quarantine threshold	1 day	1 day
8	Household quarantine duration	7 days	7 days
9	Household quarantine compliance workers	75%	53%
10	Household quarantine compliance non-workers	84%	84%
11	Cases to close a class in a school	4	1
12	Classes to close a school	6	3
13	School closure duration	10 days	21 days
14	# cases to close a department in a workplace	6	3
15	% of departments to close a workplace	60%	30%
16	Workplace closure duration	10 days	7 days

considers the hypothetical scenario of the virus mutating to a state capable of causing a worldwide human pandemic. Also, the disease natural history and the parameters used to calculate force of infection are extrapolated from the data gathered from recent cases of animal-to-human transmissions of A(H7N9) as well as data from previous pandemics caused by related viruses. Hence, the numbers presented in this paper are estimates at best.

Abbreviations

AB: Agent-based; IAR: Infection attack rate; NPI: Non-pharmaceutical intervention

Acknowledgements

"Not applicable".

Funding

"No specific funding was received for this study".

Availability of data and materials

The data analyzed during the current study are available from:

Epidemiological information: B. J. Cowling, L. Jin, E. H. Lau, Q. Liao, P. Wu, H. Jiang, and et al., "Comparative epidemiology of human infections with avian influenza a h7n9 and h5n1 viruses in china: a population-based study of laboratory confirmed cases," *The Lancet*, vol. 382, no. 9887, pp. 129–137, 2013.

Demographical information: <https://www.census.gov>

The datasets generated during the current study available from the corresponding author on reasonable request.

The R code for clustering can be requested from the first author.

Authors' contributions

WS did the data gathering, analysis and estimation of results. TD contributed with the analysis of data and was a major contributor in writing the manuscript. RI reviewed all the study and provided important feedback in all sections. "All authors have read and approved the final version of this manuscript."

Ethics approval and consent to participate

"Not applicable".

Consent for publication

"Not applicable".

Competing interests

"The authors declare that they have no competing interests."

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

¹Department of Industrial and Management System Engineering, University of South Florida, Tampa, FL 33620, USA. ²College of Public Health, University of South Florida, Tampa, FL 33620, USA.

Received: 28 March 2017 Accepted: 26 October 2017

Published online: 25 November 2017

References

- World Health Organization. Summary and assessment, 17 May 2017 to 15 June 2017. 2017. [Online; accessed 12-July-2017]. Available: http://www.who.int/influenza/human_animal_interface/Influenza_Summary_IRA_HA_interface_06_15_2017.pdf?ua=1.
- Center for Disease Control and Prevention. Avian influenza a(h7n9) virus. 2016. [Online; accessed 7-June-2016]. Available: http://www.who.int/influenza/human_animal_interface/influenza_h7n9/en.
- World Health Organization. Cumulative number of confirmed human cases for avian influenza a(h5n1) reported to who, 2003–2016. 2016. [Online; accessed 1-June-2016]. Available: http://www.who.int/influenza/human_animal_interface/EN_GIP_20160509cumulativenumberH5N1cases.pdf?ua=1.
- World Health Organization. Summary and assessment as of december 2015. 2015. [Online; accessed 1-June-2016]. Available: http://www.who.int/influenza/human_animal_interface/Influenza_Summary_IRA_HA_interface_14_Dec_2015.pdf?ua=1.
- Ferguson NM, Cummings DA, Cauchemez S, Fraser C, Riley S, Aronrag M, et al. Strategies for containing an emerging influenza pandemic in southeast asia. *Nature*. 2005;437(7056):209–14.
- Ferguson NM, Cummings DA, Fraser C, Cajka JC, Cooley PC, Burke DS. Strategies for mitigating an influenza pandemic. *Nature*. 2006;442(7101):448–52.
- Shrestha SS, Swerdlow DL, Borse RH, Prabhu VS, Finelli L, Atkins CY, et al. Estimating the burden of 2009 pandemic influenza a (h1n1) in the united states (april 2009–april 2010). *Clin Infect Dis*. 2011;52(suppl 1):S75–82.
- Liu Z, Fang C-T. A modeling study of human infections with avian influenza a h7n9 virus in mainland china. *Int J Infect Dis*. 2015;41:73–8.
- Zhou X, Li Y, Wang Y, Edwards J, Guo F, Clements AC, al. The role of live poultry movement and live bird market biosecurity in the epidemiology of influenza a (h7n9): a cross-sectional observational study in four eastern china provinces. *J Infect*. 2015;71(4):470–9.
- Husain M. Avian influenza a (h7n9) virus infection in humans: epidemiology, evolution, and pathogenesis. *Infect Genet Evol*. 2014;28:304–12.
- Nguyen VK, Hernandez-Vargas EA. Identifiability challenges in mathematical models of viral infectious diseases. *IFAC PapersOnLine*. 2015;48(28):257–62.
- Wang T. Dynamics of an epidemic model with spatial diffusion. *Phys A*. 2014;409:119–29.
- Li J, Sun G-Q, Jin Z. Pattern formation of an epidemic model with time delay. *Phys A*. 2014;403:100–9.
- Murillo LN, Murillo MS, Perelson AS. Towards multiscale modeling of influenza infection. *J Theor Biol*. 2013;332:267–90.
- Cowling BJ, Jin L, Lau EH, Liao Q, Wu P, Jiang H, et al. Comparative epidemiology of human infections with avian influenza a h7n9 and h5n1 viruses in china: a population-based study of laboratory confirmed cases. *Lancet*. 2013;382(9887):129–37.
- European Centre for Disease Prevention and Control. Mutation of avian influenza A(H7N9): now highly pathogenic for poultry but risk of human-to-human transmission remains low. 2017. [Online; accessed 5-July-2017]. Available: <https://ecdc.europa.eu/en/news-events/mutation-avian-influenza-h7n9-now-highly-pathogenic-poultry-risk-human-human>.
- Das TK, Savachkin AA, Zhu Y. A large-scale simulation model of pandemic influenza outbreaks for development of dynamic mitigation strategies. *IIE Trans*. 2008;40(9):893–905.
- Uribe-Sanchez A, Savachkin A, Santana A, Prieto-Santa D, Das TK. A predictive decision-aid methodology for dynamic mitigation of influenza pandemics. *OR Spectr*. 2011;33(3):751–86.
- Martinez DL, Das TK. Design of non-pharmaceutical intervention strategies for pandemic influenza outbreaks. *BMC Public Health*. 2014;14(1):1.
- Prieto D, Das TK. An operational epidemiological model for calibrating agent based simulations of pandemic influenza outbreaks. *Health Care Manag Sci*. 2014:1–19.
- Prieto DM, Das TK, Savachkin A, Uribe A, Izurieta R, Malavade S. A systematic review to identify areas of enhancements of pandemic simulation models for higher practical usability. *BMC Public Health*. 2012;12:251.
- United States Census Bureau. People residing in urban areas. 2010. [Online; accessed 21-May-2016]. Available: <https://www.census.gov/geo/reference/ua/urban-rural-2010.html>.
- Dutta A, et al. Sterilizing immunity to influenza virus infection requires local antigen-specific T cell response in the lungs. *Sci Rep*. 2016;6.
- Tan P, Steinbach M, Kumar V. Introduction to data mining. First ed. Boston: Addison-Wesley; 2006.
- National center for education statistics. Number of public school districts, by locale code (ccd) and state: 2003–04. 2003/04. [Online; accessed 21-May-2016]. Available: <https://nces.ed.gov/Surveys/RuralEd/Index.asp>.
- Reed C, Chaves SS, Kirley PD, Emerson R, Aragon D, Hancock EB, et al. Estimating influenza disease burden from population-based surveillance data in the united states. *PLoS One*. 2015;10(3):e0118369.
- United States Census Bureau. Number of firms, number of establishments, employment, and annual payroll by enterprise employment size for the united states and states, totals: 2013. 2013. [Online; accessed 21-May-2016]. Available: <https://www.census.gov/data/tables/2013/econ/susb/2013-susb-annual.html>.

28. Hyder A, Buckeridge DL, Leung B. Predictive validation of an influenza spread model. *PLoS One*. 2013;8(6):e65459.
29. Wu P, Peng Z, Fang VJ, Feng L, Tsang TK, Jiang H, et al. Human infection with influenza a (h7n9) virus during 3 major epidemic waves, china, 2013–2015. *Emerg Infect Dis*. 2016;22(6):964.
30. Liu Z, Fang CT. A modeling study of human infections with avian influenza a H7N9 virus in mainland China. *Int J Infect Dis*. 2015;41:73–8.
31. Kim S, Lee J, Jung E. Mathematical model of transmission dynamics and optimal control strategies for 2009 a/H1N1 influenza in the Republic of Korea. *J Theor Biol*. 2017;412:74–85.
32. Chen Y, Wen Y. Global dynamic analysis of a H7N9 avian-human influenza model in an outbreak region. *J Theor Biol*. 2015;367:180–8.
33. Center for Disease Control and Prevention. Prevention strategies for seasonal influenza in healthcare settings. 2015–2016. [Online; accessed 6-June-2016]. Available: <http://www.cdc.gov/flu/professionals/infectioncontrol/healthcaresettings.htm>.
34. World Health Organization. Seasonal epidemics and disease burden. 2014. [Online; accessed 21-May-2016]. Available: <http://www.who.int/mediacentre/factsheets/fs211/en>.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit

