

# **Evaluating Modeling Approaches that Forecast Infectious Disease Dynamics and Predict Disease Risks**

## **Graduate Research Project\***

*Graduate Student* Boloye Gomero

*Graduate Advisor* Dr. Suzanne Lenhart

February 20, 2011

\* This work was assisted through participation in the National Center for Medical Intelligence/National Institute for Mathematical and Biological Synthesis Infectious Disease Modeling Workshop held Jan. 23-25, 2011, at the National Institute for Mathematical and Biological Synthesis, sponsored by the National Science Foundation, the U.S. Department of Homeland Security, and the U.S. Department of Agriculture through NSF Award #EF-0832858, with additional support from The University of Tennessee, Knoxville.

## Introduction

Many recent studies that have formulated models to forecast infectious disease dynamics and predict disease risks have differed according to four themes: How the data were obtained, how the parameters were derived, type of model used, and what was forecast.

Some studies have attempted to collect data near-real time via use of Geographic Information System (GIS) software (Colizza, Barrat et al. 2007, Hay, Guerra et al. 2009) or via email correspondence. Two examples of the latter method are the studies conducted by Ong, et al. in Singapore (Ong, Chen et al. 2010) and by Ajelli et al. (Ajelli, Merler et al. 2010) in Italy during the 2009 H1N1 influenza outbreak, whereby the researchers received daily email or facsimile reports from general practice family doctors (GFPDs) on the counts of Influenza-like illnesses (ILIs) cases they encountered during their consultation with patients. Some other studies obtained their data retrospectively from archived information on previous epidemics in government databases (e.g., WHO, CDC, US Census bureau, etc.) or near-real time while these databases were being updated daily during an actual outbreak. Other studies have also obtained their data from laboratory sources. For example, Fraser et al. was able to determine the start date of the 2009 H1N1 outbreak in La Gloria, Mexico, by using a Bayesian coalescent method to analyze the diversity in the genetic sequences of viral samples collected from confirmed cases (Fraser, Donnelly et al. 2009).

With regard to how the parameters were derived, some studies have used parameter information that was estimated in other studies, or they have estimated the parameter values from data using Bayesian statistical analyses or by maximum likelihood estimates.

There have also been several different types of models used to study disease dynamics and attempt to forecast disease risk. Some of the models have been built on a SIR-compartmental framework using Susceptible, Infected and Recovered compartments, respectively, wherein members of the population are allowed to move from one compartment to the other using homogenous mixing and mass action laws and whereby the disease dynamics in each compartment are defined by coupled

differential or difference equations. Other models have incorporated additional features into the basic SIR-compartmental framework, and these features include additional compartments, age structure and other heterogeneous components, a seasonal forcing function, or the ability of the parameters to vary stochastically through time.

With network models, several studies have sought to simulate the transmission of disease in a defined social structure. Such models are usually large in scale and may be executed on a global or national level. Usually, the landscape is divided into grids and each grid is delineated by specified interaction rules, for example, a SIR interaction dynamic. A study by Jewell et al., for instance, built a network model that examined the transmission of avian influenza among poultry holdings in the UK. Their goal was to estimate the risk that an individual holding would pose in the event of an outbreak (Jewell, Kypraios et al 2009).

Agent-based or individual-based models are tools that have been used to garner insights into disease epidemiology. In such models, the members of a population are identified as agents with assigned decision rules about things such as movement or interaction with other members of the population. Simulations are intended to unveil emergent patterns and describe the transmission dynamics of the disease. Cooley et al., for example, designed an agent-based model to study the potential spread of Asian influenza virus in Allegheny County, Pennsylvania. Their goal was to examine the number of healthcare workers (HCWs) that may become infected, number of HCWs that will be available, number of patients each HCW should see, how quickly HCWs should be protected, and, the potential effect of varying HCW compliance with protective measures, in the event of an outbreak (Cooley, Lee et al. 2010).

Several studies have also explored the use of statistical tools to predict disease risk. Such models have been explored for their utility in capturing the dynamics of the disease during the first few days of an outbreak. They are often used to quickly derive parameters that measure the transmissibility of the disease during the first few weeks of an epidemic outbreak. The goal is usually to estimate such parameters as generation interval, total infected, epidemic start date, disease final size, peak of infection, etc.

The focus of various studies has also differed in what is being forecast. Some studies have attempt to measure the basic reproduction number,  $R_0$  (defined to be the

average number of secondary cases per typical case in a completely susceptible population), of disease agents. This has been done in different contexts, for example, how the value of  $R_0$  changes with respect to control measures implemented (for example, Milne, Kelso et al. 2008, Wood, Zamani et al. 2007), or, the best method for deriving  $R_0$  (Wallinga and Teunis 2004, Cauchemez, Boelle et al. 2006, and Wallinga and Lipsitch 2007 ). Some studies have also attempted to predict  $R_0$  retrospectively by fitting a specific model that best models the observed data, or, by using maximum likelihood estimates or extensions. There have also been attempts to predict  $R_0$  in near-real time using Bayesian techniques (for example, Cauchemez, Boelle et al 2006, and White and Pagano 2008). Note also that some of these studies have looked at  $R$  instead of  $R_0$  whereby for the former, the reproduction number is measured in a population that might be immune; its value is typically smaller than that of  $R_0$ .

The issue of complexity has differentiated epidemiology models, as with increased complexity comes increased computational time, number of parameters, and computing capacity. Arino, Brauer et al. in their 2006 paper, "Simple models for containment of a pandemic," addressed this issue by presenting a simple model as an alternative to more complicated models that for forecasting the dynamics of a pandemic. They claimed that many of the predictions of stochastic network models could be made with simple models in general, and that such simple models were a better method of planning for a pandemic when the locations of an outbreak and parameters values specific to the outbreak are not known with certainty in advance.

Complexity or scale is not the only issue at hand. Some other issues include: Notification bias (characterized by the need for better surveillance to increase case-finding time and the need for timely alerts to enhance efficient and early implementation of control measures), overestimation or underestimation of outcome measures (caused by factors such as unknown parameters, e.g., Generation interval, or model assumptions), and uncertainty (most studies do not rigorously quantify their level of uncertainty in their models).

The summaries that follow highlight the unique features of some published models since 2005 that have aimed at studying disease spread. The differences in how the data were obtained, how the parameters were derived, type of model used, and

what was forecast have also been highlighted for each of these summaries. (Please note that these articles have been posted on a Wiggiio group site set up for this research project.)

### Summary of Recent Studies

1. Modeling the worldwide spread of pandemic influenza: baseline case and containment interventions (Colizza, Barrat et al. 2007)

**What is being forecast:** Effectiveness of antiviral drugs (AV) as a containment strategy in the presence of air travel restriction during a pandemic outbreak.

**Disease agent:** Avian Influenza H5N1.

**Type of model:** A meta-population model that uses a Susceptible-Latent-Infected-Permanently Recovered framework (SLIR) in which the SLIR model is defined by a set of stochastic differential equations.

#### **Description**

The main input for the model is census data (obtained using the geostatistical software, ArcGIS) for all 3,100 urban areas studied and air traffic flow data obtained from the International Air Transport Association (IATA). The air traffic data is used to construct a weighted graph with  $V = 3,100$  vertices denoting airports in 220 countries and  $E = 17,182$  weighted edges,  $w_{i,j}$  with  $w_{ij}$  denoting passenger flow between airport  $i$  and  $j$ .

The model is a discrete SLIR model that is defined by a set of stochastic equations. The researchers adopt a meta-population approach whereby individuals are allowed to travel from city to city via airline transportation network and to move from one disease compartment to another as they travel. First, a baseline model is created which incorporates a stochastic transport operator (a function of traffic flow) and no intervention. Next, an intervention in the form of antiviral (AV) treatment is incorporated into the model. Based on whether or not the disease symptoms are detected, some individuals, with a pre-defined probability, are allowed to travel while others are not. Likewise, when AV treatment is available, some symptomatic individuals, with a pre-defined probability, are allowed to travel while others are not.

To analyze their model, the researchers use the basic reproductive number  $R_0$  as the main parameter to describe the spread rate of the disease. They also included seasonality in the infection transmission of the model. The latter was done by varying a transmission parameter that varied in time depending on the geographical location. Four scenarios for the basic reproductive number were considered. By pairing each of these scenarios with one of two start dates for the pandemic (corresponding to different seasons) and with one of four possible start locations for the outbreak, that is, Thailand, Vietnam, Europe or the United States, respectively, they analyzed disease prevalence by country and by major city in the presence of antiviral treatment and travel restrictions.

To quantify the uncertainty in their model assumptions, the researchers perform sensitivity analysis to study the effect of different temporal and geographical settings for the start of the pandemic, as well as for the role that the AV protocol assumed.

2. *Protecting healthcare workers: a pandemic simulation based on Allegheny County (Cooley, Lee et al. 2010)*

**What is being forecast:** Number of healthcare workers (HCWs) that may become infected, number of HCWs that will be available, number of patients each HCW should see, how quickly HCWs should be protected, and, the potential effect of varying HCW compliance with protective measures.

**Disease agent:** Influenza

**Type of model:** Agent-based model (ABM) that uses a Susceptible-Infected-Recovered (SIR) framework.

**Description**

The main input for the model is the population data of Allegheny County, PA. The ArcGIS Business Analyst software identified 48,595 businesses with total employment of 601,022. The researchers then synthesized workplaces of similar size and location to the actual hospitals and tagged them as hospitals, so the model could track HCWs through the simulations. The model parameters were obtained from a study by Longini

*et al* (Longini, Nizam *et al.* 2005) who worked with data on the 1957-1958 Asian influenza pandemic.

Their ABM uses agents to represent individual persons. Agents are assigned behaviors and activities within the population and are allowed to interact with each other. The ABM model is SIR-based and is able to track features such as age, sex, occupation, household location, household membership, school assignment of students, work location assignment of employed adults, work status as employed or unemployed, and disease status. Agents are assigned households using the US 2000 Census. The ArcGIS Business Analyst is then embedded with a segment of 35 Healthcare facilities (HCF) out of the 35,317 workplaces, with the HCFs comprising 43,300 persons with 19,508 HCWs who see patients on a daily basis.

The authors obtain ABM transmission probabilities from study by Longini *et al* (Longini, Nizam *et al.* 2005). They calibrate their model using the Ferguson *et al* (Ferguson, Cummings *et al.* 2006) approach (wherein 70% of all transmission occurred outside the household, with 33% of these occurring in the general community and 37% occurring in schools and workplaces) from historical influenza pandemics and with that, generate contact patterns that reproduce an epidemic similar to the 1957-1958 epidemics .

To analyze their model, the authors generate 100 calibrated epidemics using 100 distinct random sequences, each seeded with a single-infected adult. Next they investigate the attack rate for unprotected HCWs. The sensitivity of their calibration scheme is tested against an alternative calibration method described by Ferguson *et al.* They also test an alternative calibration rule that assumes equal transmission rates in schools and workplaces. Both schemes leave the principal study conclusions unchanged. In addition, they perform a sensitivity analysis of the HCW contact rate, and show that the attack rate varies very little in response to changes in the HCW contact rate.

In conclusion, they find that protecting hospital-based health care workers before or early in a pandemic has positive effects for the HCWs and the community as a whole.

3. Seasonal transmission potential and activity peaks of the new influenza A(H1N1): a Monte Carlo likelihood analysis based on human mobility (Balcan, Hu et al. 2009)

**What is being forecast:** Transmission potential of the H1N1 virus.

**Disease agent:** influenza A/H1N1v.

**Type of model:** A global structured meta-population discrete stochastic model, GLEaM (for Global Epidemic and Mobility) (Balcan, Colizza et al. 2009), that incorporates mobility and transportation data worldwide. The world population is partitioned into grids and the subpopulation in each grid is partitioned into Susceptible (S), Latent (L), Infected (I) (symptomatic and asymptomatic) and permanently Recovered classes, respectively. For each set of parameters generated by maximum likelihood estimates, the model generates quantities of interest such as the profile of the epidemic behavior in each subpopulation [subpopulation explained below] and the number of imported cases imported into the country.

**Description**

Two types of input data is fed into their GLEaM model: Census data obtained from 'Gridded Population of the World Project' of the SocioEconomic Data and Applications Center (SEDAC), and Transportation mobility data (that is, list of worldwide airport pairs connected by direct flights and the number of available seats on any given connection) obtained from the International Airport Transport Association (IATA) and Official Airline Guide databases. With these two data types, they divide the world into georeferenced census areas defined with a Voronoi tessellation procedure around transportation hubs. For their study, they identify 3362 subpopulations centered around IATA airports in 220 different countries.

The model is used to simulate mobility of individuals from one subpopulation to another by a stochastic procedure. Susceptible, latent, infectious symptomatic, infectious

asymptomatic and permanently recovered classes are defined for each subpopulation and transition between compartments occurs through a binomial and multinomial process.

For their analysis, they use their model to provide a maximum likelihood estimate of transmission potential by finding the set of disease parameters that best fit the data on the arrival time of cases in different countries. They also define the initial conditions of the epidemic by setting the onset of the outbreak near La Gloria in Mexico in February 2009 and then test different localizations of the first cases in census areas close to La Gloria. Following that, they perform sensitivity analysis on the starting date of the epidemic and they use a time-dependent modification of the reproductive number,  $R_0$ , in Mexico to model the control measures implemented in the country starting in April 2009 and May 2009. To account for seasonality, they model seasonality by using a standard forcing that rescales the value of  $R_0$  to  $R(t)$ . The parameter,  $R(t)$ , is used because the seasonality rescaling depended on the time of the year.

Following this step, for each set of parameter values, they generate  $2 \times 10^3$  stochastic realizations of the pandemic evolution worldwide for each  $R_0$  value. They then generate the distribution of arrival time of the infection in each country using a Monte Carlo method. Next, they use the estimates they obtain to simulate geographic and temporal evolution of the pandemic.

Uncertainty is handled by performing sensitivity analyses of the disease parameter values, on the starting date of the epidemic, as mentioned earlier, and to account for late/missed detection of symptomatic individuals.

Finally, they give outcome measures such as prevalence, morbidity, number of secondary cases and number and date of imported cases for each subpopulation.

#### 4. Real-time epidemic forecasting for pandemic influenza (Hall, Gani et al. 2007)

**What was forecast:** The timing of the maximum prevalence of a pandemic wave, along with its amplitude and duration.

**Disease agent:** Pandemic influenza of 1918-1919, 1957-1958 and 1968-1970.

**Type of model:** Deterministic SEAIR differential equations model.

### **Description**

The authors presented a deterministic SEAIR differential equations model with Susceptible (S), Exposed (E), Asymptomatic Infected (A), Symptomatic infected (I), and Recovered (R) compartments, respectively. They applied this model to the 1918 Spanish influenza pandemic, the 1957 Asian influenza pandemic, and the 1968 Hong Kong influenza pandemic. Note that instead of the entire population, initial conditions for the model were made to depend only on the proportion of the population with background immunity prior to the outbreak.

The authors obtained their data from the Royal College of General Practitioners (RCGP). The datasets for the 1918 and 1957 pandemics were of the reported case fatalities over the whole population while the dataset for the 1968 Hong Kong influenza pandemic was of raw Influenza-like-illnesses (ILI) consultation records from the first and second waves of the pandemic.

A total of nine parameters were used for the analysis. Five of these were parameter values for the disease kinetics and they were obtained from values reported in previous studies (the authors proposed that in the event of an actual epidemic, more precise parameter estimates be obtained from detailed epidemiological studies in the pandemic's early stages). Two of their parameters, population size and frequency of reporting, were data-dependent and specific to location. The remaining two parameters, the reproduction number and the proportion of cases that featured in the data, were unknown and so derived as part of a fitting process.

To analyze the data, the authors treated the data as if it was arriving in real time, without knowledge of future prevalence rates. The parameters were varied in the model to generate epidemic curves, including the maximum likelihood estimate (MLE) for the

anticipated underlying behavior of the epidemic. The MLE was then used to produce a family of epidemic curves that, within credible limits, fit the data. The parameter specifying the proportion of cases that featured in the data was bounded and the MLE of the true epidemic curve arising from its range of values was also plotted (they called it *formal MLE*).

The authors began the fitting process when the reporting rate had increased for three consecutive weeks. For the Hong Kong influenza epidemic, they began fitting when the weekly rate was above 200/100,000 cases.

The main outputs for the model were several predicted epidemic curves (i.e., incidence vs. number of weeks elapsed since the first rate used in the model was reported) for each pandemic wave of each of the three pandemics. These curves showed known incidence data, the unknown future prevalence rates for comparison of the derived fit, and fitting for the formal MLE. This fitting for the formal MLE was an envelope of the first two. It was unreliable in the early stages of the outbreak, since their chosen bound for it was unrealistic and it became tighter as more weeks of data became available. The authors proposed that the bound could be made more accurate using actual data obtained in the first few weeks of an outbreak.

With regard to the main results, the authors were able to use their model outputs to predict the timing of the peak of an epidemic wave within  $\pm 1$  weeks for the Hong Kong influenza data, and -2 to +1 weeks with the 1918 and 1957 mortality data.

In summary, the authors in this study used historical pandemic data to derive bounded MLEs and associated potential solution envelopes. They claimed that this approach yielded better predictors of future prevalence rates than did data obtained solely from past experience.

5. *A novel approach to real-time risk prediction for emerging infectious diseases: a case study in Avian Influenza (Jewell, Kypraios et al. 2009)*

**What was forecast:** Real-time predictions of quantities such as “the probability of individual poultry holdings becoming infected, the risk that individual holdings pose to the population if they become infected, and the number and whereabouts of infected, but not yet detected holdings.”

**Disease agent:** Avian Influenza H5N1.

**Type of model:** A continuous -time stochastic SINR model to which a reversible-jump Markov chain Monte Carlo algorithm is applied.

### **Description**

The authors applied a reversible-jump Markov chain Monte Carlo algorithm to a combined spatial and contact network continuous time stochastic Susceptible (S), Infected (I), Notified (N), Recovered (R) (SINR) epidemic model. Individuals were defined as poultry holdings. The notified group, in particular, comprised holdings that had been detected as having the disease, that were subject to government-imposed movement restrictions, but were still capable of infecting susceptibles. This group differed from the infected group which, on the other hand, was derived from holdings that had the disease, were capable of infecting susceptibles, but whose infectivity has not been detected.

The model was constructed in a Bayesian context to provide a real-time prediction during an epidemic. Unlike previous approaches that considered completed epidemics, using likelihood-based techniques and, hence, made assumptions about the time between infection and detection, the authors showed how a Bayesian approach to making inference on model parameters could be used to make risk predictions for epidemics in progress.

The authors obtain their data from Defra’s Great Britain Poultry Register (GBPR; set up in 2006). From this data, they obtained information on the major poultry production types present on each holding and the distance between any two holdings (the latter information was used in a spatial kernel). Using the data, they also characterized three main networks within the poultry industry: feed mills, slaughterhouses and company contacts. Based on information gathered from consultation with the networks, the

authors defined a frequency matrix for visits to and from the feed mill and slaughter houses. From consultation with the industry, they obtained information with which they defined a simple binary contact matrix used to represent the company network.

To analyze the model, the authors incorporated the concept of infectious pressure (i.e., idea that infected holdings 'push' on susceptible holdings with a pressure equal to the instantaneous rate of infection between them) into their model. This infectious pressure was used to define their transmission kernel. The authors also assumed a time-inhomogeneous Poisson model for the times of infections similar to the General Stochastic Epidemic, GSE (Bailey 1975). In addition, they superimposed another Poisson process on the notifications and following that, they defined the transmission rates for the infected holdings versus the notified holdings. The infected holdings included the presence of explicit contact networks while the notified holdings did not, due to the government restrictions imposed on them. Finally four parameters were fed into the model: A vector of length ten whose elements represented each of the production types, a vector of length 4 whose elements represented the different infection rates (i.e., background infection rate, infection rate due to company network, infection rate for an exponential spatial kernel for the infected holdings and another infection rate for the notified holdings), a vector of length two whose elements represented the contact frequencies (with the feed mills and with the slaughterer houses, respectively), and lastly, a decay parameter. The model was then used to make Bayesian inference on the epidemic.

The authors simulated the avian influenza epidemic using estimated parameter values and an epidemic of final size 375 (out of a total population size of 8636) was chosen. Analyses were performed at days 14, 25 and 50 and number of notified vs. occult (undetected) infections were noted. During the analysis, statistical information about the model parameters was built up during the epidemic. Authors demonstrated that uncertainty of the infection risk diminished quickly once the epidemic data began to arrive and they were able to see this as soon as 25 days into the epidemic progression.

The main outputs of the model were graphical representations of parameter learning for the parameters and several risk maps, including ones showing holding-level risk of infection and the probability of a holding-level reproduction number being greater than 1.

With regard to main results, suffice it to say that the model was rather robust in handling the problem of missing data and incorporating (as well as quantifying) parameter uncertainty into future predictions. Very importantly, steps were taken to shorten algorithm run time to just 7 hours, making the model valuable in the event of an actual outbreak. All the sums involved in calculating the likelihood were parallelized using a shared memory that showed good scalability on a 16-core Sun X4600 server.

6. *Effects of internal border control on spread of pandemic influenza (Wood, Zamani et al. 2007)*

**What was forecast:** The effect of time delay between epidemics in two population centers by imposing travel restrictions.

**Disease agent:** Pandemic influenza virus.

**Type of model:** Discrete time stochastic SIR model.

**Description**

The authors presented a discrete time stochastic SIR model with which they investigated the capacity of internal border control to mitigate the spread of an epidemic.

Australia was chosen as the country of study because of its unique demographic features and geopolitical boundaries. The main inputs for this study were the average volumes of domestic air travel for Sydney, Melbourne, and Darwin obtained from the Australian Domestic Airline Activity report and the authors only considered direct flights. Parameter values for the disease kinetics were obtained from existing literature.

Two scenarios were studied: one in which the authors assumed that the initial cases occurred in Sydney and spread to Melbourne in the presence of travel restrictions, and

one where they assumed that initial cases occurred in Darwin, a much smaller city than Sydney, and spread to Sydney. The latter was intended to simulate containing the epidemic within a smaller town through the use of travel restriction.

For the two scenarios, 80%, 90% and 99% travel restrictions were imposed and compared with the base case of unrestricted travel. They showed graphically how delay, defined as the median time between the day when the number of infected persons first reached 20 in Sydney (or, Darwin, respectively) and the day when the number of infected persons first reached 20 in Melbourne (Sydney, respectively), depended on reproductive rates (1.5, 2.5 and 3.5). They also showed how this delay depended on the form of the infectivity profile (constant infectivity vs. peaked infectivity) and the timing and severity of travel restrictions. On the output graphs, the authors indicated the timing during which the scenario 1 epidemic grows from 20-1000 cases in Sydney (Darwin, respectively for scenario 2).

The delay was found to be highly sensitive to assumptions about transmissibility of the influenza virus as well as to the ratio of city sizes, differences in travel rates, and the originating city. They suggested that if the origin of the epidemic were a smaller city, travel into that city from a larger one should be restricted. They also showed that when the growth rate is low, moderate delays in the pandemic could be achievable.

With regard to the main results, the model's simplicity (in comparison to large-scale network models on a global scale) is attractive, as it only examines 2-city routes. This allows for simpler computations and more detailed analyses of results. (Note that the only software needed for the analysis was MATLAB version 7.04 and the MathWorks Statistical Toolbox.) With this model, they were able to make inferences on how delay between epidemics in two connected locations depended on travel restrictions, population sizes, travel rates and residence of travelers, and the transmissibility of the influenza virus. As the authors point out, their study can be adapted to other countries with similar demographic characteristics as Australia.

The authors do note a few issues with their model. For instance, the model did not account for seasonal variations in the volume of air traffic and only direct flights were considered. The authors also assumed homogenous mixing within city. Asymptomatic infections were not accounted for, but the authors showed that excluding them did not greatly affect their results. Finally, the authors did not include additional importations or restrictions on overseas travel.

7. Simple models for containment of a pandemic (Arino, Brauer et al. 2006)

**What was forecast:** The relationship between the reproduction number and the number of doses needed in the course of a treatment during an epidemic.

**Disease agent:** Pandemic influenza virus.

**Type of model:** Deterministic SLIAR model.

**Description**

The authors presented a deterministic Susceptible (S), Latent (L), Infected (I), Asymptomatic infected (A), Recovered (R) ordinary differential equations model based on work by Brauer (Brauer 2006) but with the additional property that members of the population who are infected never develop symptoms but go directly from the latent stage to an asymptomatic infective stage and then to the removed stage, as suggested by Longini *et al* (Longini, Halloran et al. 2004). The authors then incorporate treatment into the model by adding new compartments for treated S, L, I and A individuals.

This model was presented as an alternative to the more complicated models that are often used to forecast the dynamics of a pandemic. The authors claimed that many of the predictions of stochastic network models could be made with simple models (such as the one they presented in the paper) and are a better method of planning for a pandemic when the locations of an outbreak and parameters values specific to the outbreak are not known with certainty in advance.

The main inputs the model used were disease parameters and mean attack rate comparable to that of Longini *et al* (Longini, Halloran et al. 2004) for the 1957 influenza epidemic. To analyze the model, they first computed the reproduction number as well

as the final size relation and the attack rate. Next, the authors showed, using contour plots, the relationship between the reproduction number and the number of antiviral doses used in the course of a treatment. For example, they showed that if the reproductive number was greater than or equal to 1, estimates of the number of antiviral doses needed during an outbreak were sensitive to parameter values. Thus, given this output, the authors could estimate various combinations of treatment rates of latent and symptomatic individuals required to bring the reproduction number down to 1. Nevertheless, the authors warned that the number of doses of antiviral treatment needed was proportional to the proportion infected during and at the start of the epidemic.

As it concerns main results, the authors proposed that their method (of using a simple deterministic compartmental model with a large range of parameter values) be used as a first response to an outbreak before implementing more complicated and computationally intensive stochastic network models.

Note that the authors do not state explicitly how or if uncertainty was handled in their model.

8. *Community-based measures for mitigating the 2009 H1N1 pandemic in China*  
(*Tang, Xiao et al. 2010*)

**What was forecast:** The effectiveness of implementing stringent Non-pharmaceutical Interventions (NPIs) during the 2009 H1N1 Pandemic in China. Also comparing the effect of *Fengxiao* and travel restrictions versus other NPIs.

**Disease agent:** 2009 influenza A/H1N1.

**Type of model:** A deterministic SEIR differential equations model extended to a meta-population model with network structures.

**Description**

The authors presented a spatially stratified SEIR model in which there were two E groups,  $E_1$  (exposed but not yet infected) and  $E_2$  (infectious but not yet symptomatic). The compartment, I, represented members that were infectious with symptoms.

Quarantine was implemented for the  $E_1$  and  $E_2$  classes as well as hospitalization for those that were infected with symptoms (thus giving an additional three compartments,  $Q_{E1}$ ,  $Q_{E2}$ , and  $H$ ). This baseline model was extended to a meta-population (or, patch) model with network structures. The patch model was formulated deterministically with the patches representing different demographic regions in China.

The data for the study was obtained from the province of Shaanxi on the initial laboratory-confirmed cases on H1N1 and the authors used it together with Markov chain Monte Carlo (MCMC) simulations to parameterize the model.

The goal of this study was to examine the effect of *Fengxiao*, “a tightly monitored measure of movement restriction put in place to prevent on-campus visits prohibit college and university students, faculty, and staff members from leaving their campuses, while maintaining essential services and normal scientific activities” which was implemented during the 2009 H1N1 epidemic.

The main outputs for the model were parameter estimates from MCMC simulations, Contour plots and sensitivity analysis of the reproduction number, plots demonstrating the effect of local quarantine, hygiene precaution and *Fengxiao* on the epidemic outbreak, and lastly, plots showing the effect of nationwide travel on the outbreak and on the mean attack rate.

The authors found from their study that if *Fengxiao* were implemented early in the outbreak, it could delay the outbreak peaks. However, if not appropriately managed, it could cause the outbreak to be more severe within the restricted areas. They also showed that local control strategies affected the peak magnitude while *Fengxiao* affected the peak timing.

Finally, no mention was made of the technology used in the model or of how uncertainty was handled in the study.

9. Model predictions and evaluation of possible control strategies for the 2009 A/H1N1v influenza pandemic in Italy (Ajelli, Merler et al. 2010)

**What was forecast:** Spatiotemporal spread of A/H1N1v influenza and potential effectiveness of mitigation strategies.

**Disease agent:** A/H1N1v influenza.

**Type of model:** A national transmission model coupled with a global homogeneous mixing SEIR model. The national model was a stochastic, spatially explicit, individual-based simulation model.

**Description**

The authors presented a real-time modeling analysis of the 2009 A/H1N1v influenza pandemic in Italy. Their model was a national transmission model coupled with a global homogeneous mixing SEIR model. The national model was a stochastic, spatially explicit, individual-based simulation model. The proportion of imported cases was determined by the number of internal passengers arriving daily at Italian airports. The effectiveness of four intervention strategies was examined. These interventions were: case isolation, school closure, vaccination and antiviral.

The input data used was the Italian 2009 A/H1N1v influenza surveillance data during the early phase of the A/H1N1v epidemic. This data was a record of influenza-like illnesses (ILI) reported by approximately 830 general practitioners in 21 regions and autonomous provinces of the country. Their parameters were largely obtained from published literature.

The authors examined the effectiveness of mitigation strategies in the absence of natural immunity and when natural immunity was assumed. In addition, they explored the impact on the pandemic spread of an early arrival of the vaccine.

The main outputs for the model comprised analyses of the effectiveness of the mitigation strategies considered as determined by the parameters: final attack rate, peak day, and peak day incidence.

It was observed that the control strategies were effective at slowing down the epidemic spread by at least 1 week. In particular, vaccination was found to be more effective if coupled with antiviral treatment. Vaccination had no effect on delaying the epidemic peak and their results showed that antiviral treatment would have been the most efficient strategy to reduce the impact of the influenza pandemic.

The model predicted the weekly number of confirmed cases at a 95% confidence level with respect to the original data. To further validate their model; the authors compared reported ILI estimates with the estimates from their model simulations. Their model peak was 3 weeks ahead of the observed ILI peak when the model assumed no immunity and it was 2 weeks ahead when natural immunity was considered. It was also found that when the vaccination distribution was started earlier, the attack rate was reduced further. The authors claimed that the misalignment of the simulated epidemic peak with that of the data was possibly due to the model parameterization. They explained that as more data was to become available later that year, they would achieve a better model calibration.

10. *Virtual epidemic in a virtual city: simulating the spread of influenza in a US metropolitan area (Lee, Bedford et al. 2008)*

**What was forecast:** Weekly variation of influenza cases and presentation to the emergency rooms and clinics.

**Disease agent:** Epidemic Influenza.

**Type of model:** Agent-based model of with the incorporation of city-level details (e.g., geographic, economic and climate characteristics of the city).

**Description**

The authors presented an agent-based model (ABM) of a theoretical epidemic in Norfolk, Virginia. The model incorporated city-level details. Also, every Norfolk citizen, as well as their behavior and social interactions was represented in the model. The authors stated that their model was an extension of the BioWar (Carley, Fridsma et al. 2006) simulation, a model where five different US metropolitan areas were modeled, using specific census, geographic, weather, school district, and business/entertainment

location data for each city. The epidemic was tracked for one year in the simulations and the results are analyzed.

The main input for their model was the U.S. Census Bureau data for the city of Norfolk. Model parameters, such as, duration of illness, incubation time of the virus, etc., were obtained from the Center for Disease Control (CDC), from the U.S. Army Medical Management of Biological Casualties Handbook, and from expert opinion.

The model had the following characteristics:

- Decisions, agent movement, and all other actions and changes (including movements around the city) occurred at four-hour intervals.
- The probability of an agent moving to a specified location depended on the time of day, day of the week and month of the year.
- Each agent was assigned a unique set of socioeconomic and demographic characteristics based on the U.S. Census bureau data.
- Each agent was assigned a social network in which he related to members of the network with varying degrees of intensity and frequency. The nature of the network was determined by agent's demographic status.
- Agent interactions with one another was based on the CONSTRUCT multi-agent model of group and organizational behavior, in which individuals were more likely to interact with agents who are in close proximity during a given time step.

To analyze the model, the authors introduced 200 cases of influenza into the population on Day 87 of the simulation. The authors performed sensitivity analysis for the transmission rate, incubation period, period of infectiousness, and the immunity rate. Agents infected with influenza had probabilities of developing different symptoms. Each symptom was assigned an evoking strength score based on INTERNIST-1 Quick-Medical Reference (QMR) scoring system, with higher scores corresponding to more severe symptoms. The sum of an agent's evoking strength scores was used to

determine its health-care seeking behavior. The authors ran a total of 10 simulations and each simulation for approximately 14 hours.

The main output for the model included the following:

- Plots showing the average, maximum, minimum prevalence of influenza of the ten runs done for each day.
- The period of peak prevalence.
- Unlike in compartmental models, the model presented by the authors showed variations in the disease prevalence by day of the week. The pattern was cyclical, increasing on weekdays and decreasing on weekends.
- The reproductive number expected for different transmission rates.
- The optimal incubation time, contagious period, and level of immunity required to lower the epidemic spread.
- Number of healthcare visits over time.
- Number of influenza cases by age and by race

The authors observed that Peak prevalence occurred around 185 days. Also, a transmission rate of less than 20% failed to generate a transmission. They observed that an epidemic occurred only when the incubation period was 5 or more days. Also, when different levels of immunity were examined, the authors found that the epidemic slowed down with an immunity level of 30% or more. The number of clinic visits varied by race and cyclically by day of the week. Moreover, they found that influenza may strike earlier and with a higher prevalence among individuals age 65 and over. The model showed no correlation between infection rates and race.

Note that the researchers did not vary immunity and transmission rates by age or other demographic characteristics. In addition, as the authors pointed out, the model was not validated with real-world data. (However, they did observe that their results were comparable to those derived from simpler conventional compartment models. Their claim was that very little data was available at the time of research \_\_ i.e., only data from the three previous pandemics. It would be interesting to see how their model

predictions might have held up in the face of the 2009 influenza pandemic.) Lastly, the authors noted that their model had limited ability to be generalized to other cities since each city has its own unique set of parameters. Nevertheless, given its successful attempt at portraying the city of Norfolk's true social interactions, the model, the authors claim, will be useful for city-level planning.

That said, the authors observed that the model was computationally intensive. The code for the program was written in C++ and the simulations were executed using a supercomputer at the Arctic Region Supercomputing Center (ARSC) at the University of Alaska-Fairbanks.

11. *A small community model for the transmission of infectious diseases: comparison of school closure as an intervention in Individual-based Models of an influenza pandemic (Milne, Kelso et al. 2008)*

**What was forecast:** The impact of social distancing as an intervention tool.

**Disease agent:** H5N1.

**Type of model:** Stochastic individual-based spatial model.

**Description**

The authors presented an individual based model designed to be representative of the epidemic dynamics in a small community in the developed world in the event of a global pandemic. The town studied was Albany, Western Australia, with a population size of about 30,000. The census data and state and local government data were used to construct a virtual population whose age and household structure matched that of Albany. Parameter information was obtained from studies previously done by other researchers. The goal of the study was to predict the impact of the four social distancing measures on the prevalence of the epidemic. These measures were school closure, increased home isolation of symptomatic individuals, workplace non-attendance and community contact reduction.

To apply the model to pandemic influenza, the authors assumed that the influenza pandemic was spread by the H5N1 influenza strain, and that the pandemic started in Southeast Asia and spread to Australia where spontaneous social distancing measures were subsequently implemented as the main intervention measures. In the case of school closure, interaction at school hubs were stopped and limited to households. In the cases of increased isolation, workplace non-attendance and community contact reduction, individuals were assigned a specific probability of remaining within their household.

In the model, it was assumed that an average of one new infection per day was introduced into the population for the duration of the simulations. Each simulation was run in 12 hour day/night cycles. During each cycle, based on whether the cycle was day vs. night or weekend vs. weekday, each individual's location, infection status and whether or not an adult needed to stay at home to supervise a child, were calculated. Individuals were required to occupy three types of locations; a household, a hub, or a random or untraceable community of contacts. The larger social hubs were divided into mixing groups comprising individuals more likely to interact with each other. When an infected individual interacted with a susceptible individual, the probability that the disease was transmitted was calculated according to a transmission function (a function of a transmission coefficient and the states of the infected and susceptible individuals). Constant infectivity was assumed during the period of infectiousness. It was also assumed that infected individual were immune after an infection.

For their analysis, the authors examined the four social distancing measures with reproductive numbers ( $R_0$ ) of 1.5, 2.0 and 2.5. The effect that each of the four social distancing measures (or a combination of any two of them) had on daily and final attack rates was also considered. The authors defined "Optimal timing" to be the time prior to the introduction of the first infected case, before any interventions are implemented. The optimal timing to apply each measure was simulated. Following this, the authors then went on to compare the final infection rates from five other IBM studies conducted by other researchers (Ferguson, Cummings et al. 2005; Longini, Nizam et al. 2005;

Ferguson, Cummings et al. 2006; Germann, Kadau et al. 2006; Glass, Glass et al. 2006) with the infection rates derived from school closure when school closure was used as the single mode of intervention.

From their model output, the authors found that the final attack rates ranged from 33% to 65%, corresponding to  $R_0$  values of 1.5 and 2.5, while peak daily attack rates ranged from 89 to 474 cases per 10,000. It was also observed that the final attack rate was reduced the most when school closure was combined with any of the other social distancing schemes. Note that unlike other models in which school closure had been modeled, the researchers assumed that school closure caused no additional increase in interpersonal contacts and that early and continuous school closure resulted in a decrease in infection rate from 65% to 55%. Sensitivity analysis showed that their model was most sensitive to assumptions about the sizes of mixing groups in schools, and the relative number of community contact (compared with household and hub contacts).

The authors noted that their results were close to those of the other five IBM simulations. They claimed that any differences observed may have been due mainly to differences in the underlying model assumptions.

In conclusion, the authors showed by their work that social distancing schemes, when applied early and in combination, had the potential to lower the spread of pandemic influenza, especially for lower reproductive numbers.

12. Simulation suggests that rapid activation of social distancing can arrest epidemic development due to a novel strain of influenza (Kelso, Milne et al. 2009)

**What was forecast:** The effect of the timing of social distancing intervention strategies on attack rate.

**Disease agent:** Pandemic influenza.

**Type of model:** Stochastic individual-based spatial model. This model is the same as the one presented by Milne et al. (Milne, Kelso et al. 2008).

**Description**

The authors presented an extension of the model presented by Milne *et al.* (Milne, Kelso et al. 2008) to allow delayed introduction of social distancing intervention measures. They also conducted a new series of simulations involving introducing the same four social distancing measures (or combinations of them) at different points in time, ranging from 0 to 8 weeks after the first infectious cases is observed.

The main output for their model included graphical representations of the effect of delayed activation of intervention on epidemic attack rates; the effect of the number of reported cases on the simulated attack rates; and, impact of interventions on age-specific attack rates.

The authors found that for each  $R_0$  value, a different proportion of case isolation was required for a single intervention measure versus intervention measures used in combination. The authors assumed the epidemic had been prevented if the final symptomatic attack rate was less than 10%. The delay period (i.e., between the occurrence of the index case and the commencement of the intervention strategy) that supported this condition was longer for smaller  $R_0$  values and for combined intervention strategies. There was also a greater reduction in the total attack rate for lower  $R_0$  values and for combined interventions.

The authors discussed some of the drawbacks in their models. Their model predicted that social distancing would have to be instituted early and for as long as 5 months, the latter of which, as the authors pointed out, was unrealistic. But the authors claimed that as a tool for providing extra time to develop and distribute vaccines, it would be a useful mitigation strategy. The authors also noted that their results were probably more applicable to developed countries than to developing countries, which have high population densities, and thus, may possibly experience a faster spread of the disease if social distancing were to be implemented.

13. A model for the spread and control of pandemic influenza in an isolated geographical region (Roberts, Baker et al. 2007)

**What was forecast:** The effectiveness of different control measures at reducing the influenza incidence in the event of an epidemic outbreak in an isolated population.

**Disease agent:** Pandemic influenza.

**Type of model:** A structured Kermack-McKendrick integral equation model of the spread of influenza virus in a population with SIR compartments.

**Description**

The authors presented a structured Kermack-McKendrick integral equation model of the spread of influenza virus in an isolated population. The epidemic is assumed to be spread by the inadvertent introduction of the virus to the population by an airline passenger undetected during border screening. The chosen location for the study was New Zealand, with Auckland (with population size approximately 1 million) as the point of entry of the virus. The parameters for the model were obtained from published literature on influenza.

The model comprised one homogeneous susceptible population and four locations for the infected population: the household, school, workplace and community. Accordingly, there were four categories for the recovered population. The rate at which an individual infected at one location infected someone else at a different location was stored in a 4-by-4 matrix whose norm equaled 1.

The model was solved numerically for the values of the basic reproductive number,  $R_0 = 1.1, 2.0$  and  $3.0$  and for each  $R_0$ , the duration of the epidemic was determined. The authors examined 12 potential control measures comprising house quarantine, social distancing, targeted antiviral treatment and antiviral prophylaxis (TATP) and various combinations of these three. For each simulation, the authors examined how each control measure modified the mixing matrix, they then computed the  $R_0$  value for the new matrix, and then they determined the "threshold" control effort which reduced  $R_0$  to

1. The authors assumed that if transmission was reduced at one location, there would be no accompanying increase in transmission elsewhere.

The main outputs of the model included graphical representations of the following: proportion of people infected in the population as a function of  $R_0$ ; incidence (infection per 1000 individuals) as a function of the number of days into the epidemic outbreak; proportion infected as a function of number of days into the epidemic outbreak; threshold values for the level of control required to prevent the epidemic; and, the  $R_0$  value for each control measure. All the strategies that successfully prevented the epidemic involved TATP. All the strategies apart from closing schools were sufficient to contain the epidemic.

When compared with other models (i.e., models by Ferguson *et al.* (Ferguson, Cummings *et al.* 2006), Germann *et al.* (Germann, Kadau *et al.* 2006), Longini *et al.* (Longini, Nizam *et al.* 2005) and Wu *et al.* (Wu, Riley *et al.* 2006)), their model overestimated the proportion of people infected in an epidemic. The authors suggested that this was possibly because the other models were more complicated and used equations that did not necessarily apply to their own model.

14. *A preliminary estimation of the reproduction ratio for new influenza A(H1N1) from the outbreak in Mexico. (Boelle, Bernillon *et al.* 2009)*

**What was forecast:** The Reproduction number.

**Disease agent:** 2009 H1N1 Influenza virus

**Type of model:** Statistical model

**Description**

The goal of the authors was to obtain timely estimates for the Reproduction ratio,  $R$  (average number of secondary cases per primary case), during the first few weeks of an epidemic outbreak. In this article, they detailed how they computed  $R$  and the generation interval (the time between the occurrence of a primary case and the time of appearance of the secondary case which the primary case produces, denoted  $GI$ ) for

the influenza virus during the first few days of the 2009 A/H1N1v influenza epidemic outbreak in Mexico. The main input for the authors' study was the daily incidence data from March 11 to May 2 as reported by the Mexican Health Authorities. The data consisted of 1,364 confirmed cases given as daily count.

Two methods were used to compute the reproduction rate. In the first approach, the intrinsic growth rate over time was estimated by a Poisson regression and the result was transformed to  $R$  using Laplace transform of the distribution of the generation interval. In this method, only time periods fitting exponential growth during the epidemic were analyzed.

In the second approach, on the other hand,  $R$  was obtained from real-time estimation by determining the daily values of the reproduction ratio, that is,  $R(t)$ , after averaging the number of secondary cases over all possible modes of transmission compatible with the epidemic curve. For this method, the authors assumed a known generation interval, an equal probability for all modes of transmission of the disease, and that there were no imported cases.

As both methods required use of generation interval and the authors did not have any information on the actual generation interval in Mexico, three plausible generation intervals were derived using different approaches: one (denoted as PAN) obtained from household studies from the 1957 and 1968 pandemics, one derived from viral excretion in experimental influenza infection (denoted as VIR), and a hypothetical distribution (denoted as ELV) introduced by Elveback *et al.* (Elveback, Fox *et al.* 1976).

For the first approach to finding  $R$ , the authors observed a reproduction ratio of 2.2 for the PAN GI; 2.6 for the VIR GI; and 3.1 for the ELV GI, with only slight differences in the goodness of fit for the three methods. The reproduction ratio was observed to decrease over time for all three methods, for instance, and the PAN GI yielded a maximum  $R$  value of 2.7 at 8 days and a minimum value of 2.0 at 17 days.

For the second approach to finding R, the authors observed that the plots of R against time showed similar profiles. All these GIs confirmed that the value of R stayed around 1 until around April 8, and then it increased rapidly for the two weeks immediately following April 8. The authors observed, however, that R appeared to vary more widely depending on the GI used. For instance, for the PAN GI, the maximum value of R was about 2.1 (April 18), for the VIR GI, it was 4.0 (April 11) and for the ELV GI, it was 3.2 (April 17).

The authors noted that compared to the real-time estimates of R, their results from their first method were an overestimate. One reason for this, they claim, is notification bias. They came to this conclusion by observing that as they repeated their analysis each time more data became available, the reproduction ratio estimates they obtained grew smaller, corresponding to the set up of enhanced surveillance starting on April 16, and hence, improved case finding over time. Another reason for the observed overestimation was the fact that the GIs used had to be derived since the generation interval was unknown for the outbreak in Mexico. From more detailed retrospective analysis, the authors observed that the GI derived used were much longer than the actual observed GI (a mean of approximately 1.9 days). A third reason for the limitation in their study, the authors pointed out, was in their deciding what part of the epidemic curve displayed exponential growth. Factors such as the chosen minimum duration of the infectious period (i.e., 5 days) and the start and end dates might have introduced stochastic variations in their analysis.

Despite the drawbacks of their analysis relating to uncertainty, the authors noted that their estimates for R were similar to estimates obtained for previous influenza pandemics.

15. *Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures. (Wallinga and Teunis 2004)*

**What was forecast:** The Reproduction number

***Disease agent:*** Severe acute respiratory syndrome (SARS)

***Type of model:*** Statistical model

***Description***

In this study, the authors attempted to derive the reproductive number from time-series of case counts. During the SARS outbreak of 2002, distinct temporal patterns in the number of SARS cases were observed. The authors studied the epidemic trend of four countries, Hong Kong, Canada, Vietnam, and Singapore, which all had very similar start dates for the epidemic and of implementation of control measures. The goal was to understand if the differences observed among the four countries were due to differences in the transmission potential for SARS or in the effectiveness of the control measures implemented.

The main inputs for this study were dates of symptom onset derived from epidemic curves (number cases by date of symptom onset) provided by the World Health Organization for Hong Kong, Vietnam, and Singapore, as well as from epidemic curve provided by Health Canada for Canada. To begin their analysis, the authors derived their generation interval (i.e., the time from symptom onset in a primary case to symptom onset in a secondary case; also called serial interval or generation time) using a Weibull distribution.

The authors used a new method to derive the reproduction number,  $R$  (the actual average number of secondary cases per primary case, and according to the authors, the value of  $R$  is typically smaller than the value of  $R_0$ ). They noted that previous researchers approximated  $R$  by assuming an exponential increase in the number of cases over time or by fitting a specific model that summarized assumptions about the epidemiology of the disease. In their study, on the other hand, the authors proposed use of a new version of the maximum likelihood method to estimate  $R$ . They pointed out that the traditional likelihood method was computationally intensive because it required consideration of all possible infection networks. However, in their new version, they considered pairs of cases rather than the entire infection network. They asserted

that this method allowed for a "finer temporal resolution under more general assumptions that was previously possible."

To test the performance of their estimation procedure, the authors constructed a stochastic, individual-based model that simulated epidemic processes with properties precisely specified. A time-varying effective reproduction number  $R_t$  as a function of symptom onset date,  $t$ , as well as the model parameters were estimated from observations on the SARS epidemic in Singapore and with the obtained values, epidemic curves were simulated. After this step, their estimation procedure was applied to the simulated curves. Their results showed that most of the estimates were close to the actual reproduction numbers, having a deviation of only less than 5 percent.

Next, the authors converted epidemic curves for Hong Kong, Vietnam, Singapore, and Canada, into the time course of effective reproduction numbers. They showed that there existed a direct relation between the epidemic curve and the time course of the reproduction number, and that this relationship was determined by the distribution of the generation intervals.

In conclusion, the researchers showed in this work how to derive likelihood-based estimates of effective reproduction numbers using only the observed time of symptom onset for the observed cases. They show that their method was more accurate than previous methods that assumed an exponential increase in the number of cases early in the epidemic.

With regard to deciding if the differences observed among Hong Kong, Canada, Vietnam, and Singapore were due to differences in the transmission potential for SARS or due to differences in the effectiveness of the control measures implemented, the authors found that the disease transmission potentials for the four countries were at similar levels. With regard to control, they found that timely alerts against a new infectious disease were crucial. They observed that "delaying the institution of control

measures by 1 week would have nearly tripled the epidemic size and would have increased the expected epidemic duration by 4 weeks."

16. Real-time estimates in early detection of SARS. (Cauchemez, Boelle et al. 2006)

**What was forecast:** The Reproduction number

**Disease agent:** Severe Acute Respiratory syndrome (SARS)

**Type of model:** Statistical model

**Description**

The authors noted that up to the time of their study, the reproduction ratio,  $R$ , had only been estimated retrospectively for periods from which all secondary cases had been detected. In this study, however, they proposed a Bayesian statistical framework for quickly estimating the reproduction number in an ongoing epidemic, which would account for yet unobserved secondary cases, and consequently, be useful for quickly investigating the effect of control measures.

The authors applied their method retrospectively to data from the 2003 SARS outbreak in Hong Kong. The data comprised dates of symptom onset of the 1,755 case-patients who were detected in Hong Kong in 2003.

In their statistical model, the number of cases with symptom onset at day  $t$  was denoted by  $n_t$ . The authors assumed the reproduction number  $R_t$  for day  $t$  was defined as the mean number of secondary cases infected by a case with symptom onset at day  $t$ . Supposing the number of secondary cases the primary cases infected was denoted as  $X_t$ , the reproduction number,  $R_t$ , was defined as the ratio  $X_t/n_t$ , defined for  $n_t > 0$ .

Next, using three steps, the authors showed that the daily counts of symptom onset available until day  $T$  were sufficient to estimate  $R_t$  (i.e., in the absence of a complete knowledge of the exact chain of transmission, and of secondary cases for whom clinical onset of the disease appeared after day  $T$ ). These three steps were:

1. They "predicted the eventual number of late secondary cases (as yet unobserved by day  $T$ ), for cases reported at day  $t$ , assuming the number of early secondary cases (reported before day  $T$ ) was known."
2. Using the method described by Wallinga and Teunis (Wallinga and Teunis 2004), they estimated the number of early secondary cases from the daily counts of symptom onsets.
3. Combining the above two steps, they obtained a predictive distribution of  $R_t$ .

The authors pointed out that their estimation procedure hinged on three assumptions: that the patients with symptoms appearing before day  $T$  were identified with certainty; that the transmission events were independent; and, that the generation interval had a known frequency distribution.

To explore the ability of their method to quickly detect the effect of control measures, the authors simulated  $R_t$  the temporal pattern of  $R_t$  based on an average over the 500 simulated datasets as a function of  $T$ . Each epidemic was characterized thus: "During the first 20 days of the epidemics, the theoretical reproduction number was 3. Control measures were implemented at day 20. In a first scenario, control measures were completely effective (no transmission occurred after day 20). In a second scenario, the theoretical reproduction number after control measures were implemented was 0.7." Next, authors noted that they investigated the bias and precision of the real-time estimator, the effect of the length of the generation interval on the results, as well as scenarios in which the theoretical reproduction number remained constant with time.

The authors plotted the expectation and 95% credible intervals of the real-time estimator of  $R_t$  calculated at the end of the epidemic and after a lag of 2, 5, 10, and 20 days. Their results showed that temporal trends in the expectation of  $R_t$  were well-captured after a lag of only 5 days. For a lag of 5 days, the authors observed that the credible interval of  $R_t$  was wide when less than 20 cases were detected but was relatively narrow when more cases were detected. The authors also observed that the width of the credible interval narrowed as the lag increased and more complete data

became available. They observed no difference between retrospective and 20-day estimates.

Next, using their method discussed above, the authors estimated the impact of control measures implemented on day 20 in the simulated datasets with completely effective or limited control measures. They observed that changes in  $R$  were qualitatively the same, both when control was limited and when it was not. They noted that even when control measures were completely effective, based on data available up to day 21, the average expectation of  $R_{20}$  was approximately 3. Based on data available up to day 25, a downward trend was apparent, whereas based on data available up to day 29, the average expectation of  $R_t$  was less than 1 from  $t = 27$  days. Based on data available up to day 40 (20 days after the implementation of the control measures), the estimates indicate that the threshold value 1 was crossed at day 22, which was 2 days after control measures were implemented.

The authors noted that in theory, the method could only be applied to communicable diseases with no asymptomatic cases; with no underreporting, and with knowledge of the generation interval. These criteria, thus, effectively excludes diseases like influenza from analysis using the authors' method. The authors also point out that their assumption that the distribution of the generation interval was known and remained unchanged during the course of the outbreak was unrealistic and future modifications to the model could take this factor into account.

17. *Pandemic potential of a strain of influenza A (H1N1): early findings. (Fraser, Donnelly et al. 2009)*

**What was forecast:** The transmissibility and severity of the influenza virus early in its spread via epidemiological analyses of the basic reproduction number ( $R_0$ ) estimates as well as studies of the viral genetic diversity.

**Disease agent:** 2009 H1N1 Influenza virus

**Type of model:** Statistical model

**Description**

By assuming that the reporting of the number of travelers to and from Mexico was complete, the authors used a model of the interval censored case counts to determine the number of infections that occurred in Mexico by late April, 2010.

The authors noted that one way to determine the transmissibility of the virus was to determine the magnitude of the epidemic by determining the epidemic's actual start date. One way to find the start date was to simply use the results of epidemiological investigations. For instance, in a place like La Gloria, a small, isolated community in Veracruz province, a large attack rate for the virus strain had been observed and the outbreak was reported to have begun on 15 February 2009. A second approach which the authors explored was to examine the diversity in the genetic sequences of viral samples collected from confirmed cases, assuming diversity accumulated according to a molecular clock model. To do this, the authors used a Bayesian coalescent method that assumed exponential growth of the viral population to examine twenty-three complete publicly available hemagglutinin (HA) gene sequences from cases not linked in epidemiological clusters. From this analysis, they found that the time of the most recent common ancestor to the 2009 A/H1N1v influenza virus was 12 January 2009.

Next, the authors derived maximum likelihood estimate of the reproduction number,  $R_0$ , by examining data on the age distribution of cases and the dates of disease onset. The parameter,  $R_0$ , was determined by fitting the data with parameter estimates from a model with heterogeneous mixing by age plus age-dependent susceptibility to infectiousness. The authors also used this method to derive an independent estimate of the mean generation time. Note that the authors did not determine an estimate of age-dependent infectiousness, or age-dependent mixing from these data.

Lastly, the authors determined the time-dependent reproduction number,  $R_t$ , from a time series of reported disease onsets among confirmed cases in Mexico. The authors pointed out that even though this data had many sources of uncertainty, they developed

methods for analyzing the data that accounted for substantial underreporting as well as a change in reporting rate on 17 April when surveillance within Mexico was intensified.

Given estimates of  $R_0$  and knowledge of epidemic size at the time, the authors also estimated the number of generations  $N_t$  of transmission of the virus among humans that is necessary to explain the epidemic. They noted that their  $R_0$  values were comparable to but perhaps on the low end of estimates obtained from the 1918, 1957 and 1968 pandemic influenza outbreak.

18. *A preliminary analysis of the epidemiology of influenza A(H1N1)v virus infection in Thailand from early outbreak data, June-July 2009. (de Silva, Warachit et al. 2009)*

**What was forecast:** The Reproductive number,  $R_0$  and the case fatality rate, CFR of the 2009 H1N1 virus in Thailand

**Disease agent:** 2009 H1N1 Influenza virus

**Type of model:** Statistical model

**Description**

The authors estimated the basic reproduction ratio for A(H1N1)v virus in Thailand during the 2009 A/H1N1v influenza outbreak and proposed a method to track the actual case count despite the exponential growth rate of the epidemic.

The input for their study came from two sources. First, they obtained counts of the cases by symptom onset date from the records at the WHO National Influenza Center. This data was used to calculate the intrinsic growth rate,  $r$ , the reproduction number,  $R_0$ , and the case fatality ratio, CFR. Second, they inferred data on the age distribution of the infected population up to 14 July from the daily incidence reports from the Bureau of Emerging Infectious Diseases, Department of Disease Control, Ministry of Public Health in Thailand. The influenza outbreak in Thailand was characterized by the reporting of the first two cases on 10 May, a two-week lapse afterwards, and subsequently, and exponential growth phase starting in early June.

According to the authors, the intrinsic growth rate was obtained by Poisson regression of the epidemic curve over the exponential growth phase. The reproduction number,  $R_0$  was derived by the equation  $R_0 = 1 + rT_c$ , where  $T_c$  is the mean generation Interval (GI). Finally, the final size of the epidemic (the proportion of the population that would be ultimately infected assuming 100% susceptibility at the outset and minimal control measures) was obtained by a Newton-Raphson numerical solution of the equation. Since information the generation interval for the epidemic was unavailable, the authors obtained their mean generation interval (GI) from two previous studies, denoted  $T_1$  and  $T_2$ . The  $R_0$  equation gave the Laplace transform of the GI distribution, assuming it was distributed exponentially.

Given that the authors observed significant deviations from the exponential curve toward the latter part of the period of 1-12 June, they performed their goodness of fit analysis of the model only using a valid combination of points that achieved a realistic goodness of fit. This goodness of fit of the model was assessed by a combination of the R-squared measure and Pearson's statistic.

Next, the authors estimated the CFR from early fatality counts and used it to extrapolate the number of infected cases at a later date, after laboratory testing of all suspected cases had stopped (20 June), causing significant underreporting. All reported deaths up to 14 July were analyzed to compare the CFR between age groups.

Their tabulated results and a plot of their output confirmed that the epidemic curve for the period 1-12 June without the counts for 8, 10, and 11 June yielded the best fit for exponential growth ( $R_2 = 0.9802$ ), giving  $r = 0.41$ . The corresponding  $R_0$  were 2.07 for  $T_1$  and 1.78 for  $T_2$ . The final-size were 81.5 for  $T_1$  and 72.5 for  $T_2$ .

The authors pointed out that their estimate of  $R_0$  for A(H1N1)v in Thailand was higher than the one estimated by Fraser *et al.* (Fraser, Donnelly *et al.* 2009) for the Mexican outbreak which used  $T_2$  as the GI, but it was lower the estimate by Boella *et al.* (Boelle,

Bernillon et al. 2009) also estimated for the outbreak in Mexico. For the CFR, on the other hand, the authors observed that their results were higher than for those obtained by Frazer *et al.* The authors claimed that this discrepancy may be largely due to the availability of health infrastructure and level of awareness in Thailand.

The authors cautioned that due to uncertainties in their model (for example, from how they modeled exponential growth and due to underreporting), their results ought not to be interpreted too generally.

19. *Estimating the effective reproduction number for pandemic influenza from notification data made publicly available in real time: A multi-country analysis for influenza A/H1N1v 2009. (Ong, Chen et al. 2010)*

**What was forecast:** The epidemic progression of the influenza virus

**Disease agent:** 2009 H1N1 Influenza virus

**Type of model:** Stochastic discrete SEIR infection model combined with an observation model that modeled the number of ILI cases based on the number of ILI consultations recorded by GPFDs. Also, the parameters of the model were modeled using Bayesian statistical techniques.

**Description**

This paper describes a new strategy used to monitor and forecast in real time the epidemic spread of the 2009 A/H1N1v influenza virus. The study was done during the 2009 epidemic outbreak in Singapore and the strategy employed in the study involved using General Practice Family Doctors (GPFDs) to report cases of Influenza-like illnesses (ILIs) daily via email communication or facsimile. Rather than ILIs, previous tracking of influenza in Singapore had only used a more general measure, Acute Respiratory Illnesses (ARIs). The researchers showed by their work that ILIs provided a higher resolution of the data and engendered more accurate predictions of the epidemic waves.

The authors' forecasts were generated from a process-based model refitted daily. The model used for the analysis was a stochastic discrete SEIR model. The movement of people to the exposed, infected and recovered classes was modeled by a binomial distribution. The authors then constructed an observation model that modeled the number of ILI cases based on the number of ILI consultations recorded by GPFDs. In this observation model, the number of cases reported on a given day was modeled by a Poisson distribution.

Next, the authors estimated the parameters of their model using the Bayesian statistical techniques. According to the authors, "semi-informative prior distributions were assigned to parameters and incoming data incorporated via the likelihood function to obtain a time series of posterior distributions for the parameters and unobserved state space." Following this procedure, the authors then assessed predictive error by taking the posterior distribution of absolute difference between forecasts and observations, averaged over a one-week time horizon, and then averaged to get the posterior mean prediction error. Given that the state space was unobserved, the authors used a statistical method called particle filtering to integrate over the possible realizations consistent with the daily observations.

Plots of ILI per GFPD per day showed a weekly periodicity to the data that the authors obtained, with consultations decreasing on the weekends and increasing again during the week. As a result, the researchers used weekly averages to plot the epidemic trend.

The authors found that earlier on in their study, their model was affected by the uncertainty in their parameters due to insufficient information for the subjective prior distributions. Nevertheless, after a few more weeks, when more data had become available, the model was correctly predicting when the next epidemic peak would occur (with the predictive error averaging one ILI per GFPD per day given a time interval of one week), albeit, the magnitude of the peak was overestimated.

The authors showed by their work that within a month, and with no budget, it was possible to establish and begin a protocol for daily data submission for ILI. They also showed that, in near real-time, it was possible to process the data obtained from their protocol by simply having the GPFs enter into public database information on the ILI cases they saw the day before. Upon submission, the data is then immediately processed and graphical trends of the influenza epidemic are output. Their statistical routines were implemented using the R Statistical Programming Language. Also, the results from their model were updated daily on a website that was publicly accessible.

20. *How generation intervals shape the relationship between growth rates and reproduction numbers. (Wallinga and Lipsitch 2007)*

**What was forecast:** The Reproduction number

**Disease agent:** Influenza A virus

**Type of model:** Statistical model

**Description**

The authors began by explaining that for new emerging infectious diseases, the value of the reproductive number was usually inferred indirectly from the observed exponential epidemic growth rate,  $r$ . However, several methods exist for determining the relationship between the reproduction number and the growth rate. For instance, there are several different equations that do this, including expressions derived in Wallinga and Teunis *et al.* (Wallinga and Teunis 2004) to estimate the reproductive number from time-series of case counts. Nevertheless, the authors pointed out that it was not easy to tell which equation was most appropriate for a particular epidemic of study. For instance, they pointed out that in the case of Hepatitis C, for example, the growth rate is estimated to be 0.96 per year and its generation time is 20 years. Using a linear equation, a reproductive number of 2.9 is obtained, but on the other hand, using the Euler-Lotka equation, a value of 6.8 is obtained. So in this paper, the authors addressed the issue of knowing what equations to use when inferring the reproduction number from growth rates.

To derive an expression for  $R$ , the authors first derived the Euler-Lotka equation using the human population as an example. Next, they derived a moment generating function expression for the reproductive number  $R$ . They noted that a moment generating function, if it existed, uniquely characterized the shape of the entire probability distribution. It also defined the distribution for the generation interval.

Following this, the authors proceeded to characterize a unique expression for the generation interval for several different scenarios, including the SIR model, the SEIR model, for epidemics characterized by a normal distribution, for epidemics characterized by a delta distribution, and for epidemics characterized by an empirical distribution. After this analysis, the authors provided insight into the change in reproductive number  $R$  after control measures have been implemented. To do this, they derived an expression for the time-dependent reproductive number,  $R_t$ , via the renewal equation for the birth process as derived from the Euler-Lotka equation.

Finally, using information on the growth rate and on the generation interval derived by other previous studies, they applied their method to the Influenza A virus to "illustrate the impact that various assumptions about the shape of the generation interval distribution may have on the estimated value of the reproductive numbers for a given growth rate." They analyzed both the SIR and SEIR models for the virus, as well as an empirical histogram, and derived several estimates for  $R$ .

In conclusion, the authors noted that their work showed that the shape of the generation interval distribution determined what relationship should be assumed between the epidemic growth rate and reproductive number. They also pointed out that their theoretical framework could be expanded to include discrete generation times, thus, accommodating diseases where the moment of infection is tied to discrete times.

21. A likelihood-based method for real-time estimation of the serial interval and reproduction number of an epidemic. (White and Pagano 2008)

**What was forecast:** The reproduction number and the serial interval

**Disease agent:** Various

**Type of model:** Statistical model

**Description**

As the authors pointed out, several methods have been developed for deriving the basic reproductive number. For instance, a method involving use of a branching process estimator (Guttorp 1991) for  $R_0$  required knowledge of the mean of the serial interval, or for the epidemic to have a long incubation time. The novel method developed by Wallinga and Teunis (Wallinga and Teunis 2004), on the other hand, also required knowledge of the number of new cases each day for an entire epidemic and knowledge of the serial interval. This method was improved upon by Cauchemez *et al.* (Cauchemez, Boelle et al. 2006) whose modification allowed for real-time estimation of  $R_0$  using Bayesian techniques to augment the data. In a subsequent publication, Cauchemez *et al.* (Cauchemez, Boelle et al. 2006) described a Bayesian method that "used a small subset of contact tracing data and daily case counts to determine the efficacy of the interventions by observing posterior probabilities of  $R_0 < 1$ . The serial interval, (that is, generation interval) was not estimated, but no information on it is required, except that provided by the contact tracing data."

In this article, the authors presented a method that will allow for the real-time simultaneous estimation of the basic reproduction number and the serial interval from surveillance data during the initial explosive phase of the epidemic. The authors pointed out that traditionally, the serial interval had only been estimated through detailed time-consuming and expensive contact tracing, but that their serial interval estimator was able to use information on the number of cases observed each day, much more readily available information than contact tracing.

In deriving their estimation of the reproduction number, they considered one method for when the serial interval is known (best-suited for post-epidemic analysis) and a different method for when the serial interval is unknown (good for real-time analysis).

For the scenario involving adequate knowledge about the serial interval, the authors point out that their method for deriving  $R_0$  could be derived as a maximum likelihood estimator (MLE) from the likelihood presented in Bauch *et al* (Bauch, Lloyd-Smith et al. 2005). They showed how this estimator was related to a branching process estimator and described its branching process features that were pertinent to their method. Then they showed the relationship between the Bayesian posterior mode and the MLE and described the properties of a Bayesian estimator.

For the scenario in which the serial interval is unknown, the authors used the example of a two-parameter gamma distribution which provided a rich family with sufficient flexibility to model a large number of infectious disease data sets. This gamma distribution was then discredited and the resulting probabilities from the distribution were normalized so that they summed to 1 and represented a probability distribution. The authors pointed out that one could apply a Bayesian approach to this problem but doing so would require intensive Markov Chain Monte Carlo methods for the analysis. The authors claimed that their method, whether with the serial interval known or unknown, was easier to implement.

Next, they performed a simulation study that validated their method as being reliable. They also confirm the reliability of their model in real-time estimation. Finally, they show the usefulness of their method by considering data from the Ebola outbreak in 1995 in Congo, the H7N7 Avian Influenza outbreak in 2003, and the Swine Flu outbreak in 1997.

22. Mitigation strategies for pandemic influenza in the United States. (Germann, Kadau et al. 2006)

**What was forecast:** The impact of various levels and combinations of influenza antiviral agents, vaccines, and modified social mobility (including school closure and travel restrictions) on the timing and magnitude of the spread of the disease.

**Disease agent:** H5N1 avian influenza A virus.

**Type of model:** A stochastic agent-based discrete-time simulation model.

**Description**

The authors sought to investigate the dynamics of the spatiotemporal spread of the disease by introducing a large-scale national stochastic simulation model to investigate the spread of a pandemic strain of influenza virus through a simulated U.S. population of 281 million individuals for  $R_0$  (the basic reproductive number) from 1.6 to 2.4. Their main goal was to capture the disease dynamics during the early stages of an outbreak. The main input for the study was the U.S. Census Bureau and Department of Transportation data on population demographics and mobility. To perform their analysis, the authors developed a national-level stochastic agent-based discrete-time simulation model.

The simulated 281 million people were distributed among 65,334 census tracts each of whose population distributions matched that obtained from the 2000 U.S. Census data. A 2000-person community was assigned to each tract. The model was made to run in cycles of two 12-hour periods ("day" and "night"), during which the authors identified seven contexts ("mixing groups") within which individuals could associate. Some of these contexts involved closed relationships while some others involved occasional, nonspecific interactions. Each class of mixing group was assigned its own set of age-dependent probabilities of person-to-person contact. Each of these contact probabilities were multiplied by the probability of transmission given contact, "a single multiplicative constant that can be varied to model different  $R_0$  values."

U.S. Census data on tract-to-tract worker flow was used to model the commute of working adults to their workplace, thus, accurately capturing the short- to-medium-distance population mobility important for disease spread. It was also assumed that, each individual took occasional long-distance trips (three per year on average), lasting between 1 day and 3 weeks, matching Bureau of Transportation Statistics data.

The authors assumed that influenza was introduced into the U.S. mainly by air transportation and that 0.04% incubating international passengers introduced the virus

to the U.S. each day at each of 14 major international airports in the continental U.S. The authors also assumed that a cumulative number of 10,000 symptomatic individuals nationwide were required to trigger a nationwide pandemic alert. The simulations were run for 180 days, the length of a U.S. influenza season. As soon as an index case was identified in a household, day care, preschool, school or workplace, neuraminidase inhibitors were used to treat the index case while targeted prophylaxis (TAP) was initiated for the rest of that household, day care, preschool, school or workplace.

The authors also explore the use of a ‘dynamic vaccination’ strategy, in which low-efficacy vaccine became available incrementally, starting as soon as 2 months before, to as late as 2 months after, the first individual in the U.S. was infected. Different production rates for the vaccine, different total production amounts, and different distribution policies (either uniformly throughout the population or preferentially to children) were also explored.

Social distancing strategies were investigated at the levels of schools, local communities and nationwide travel. During the epidemic, interactions became more concentrated in households and household clusters and long range travel was reduced by about 1%. Using  $R_0 = 1.9$ , the authors began their simulations in the absence of intervention strategies in densely populated areas. The pandemic peaked at 85 days with a final illness attack rate of 43%. A pandemic wave was observed after the first two months when more than 100,000 people became ill each day. This observation was similar to what was observed for the 1957-1958 and the 1968-1969 influenza pandemics.

Next, the authors studied the effect of using different combinations intervention strategies at different stages of the epidemic on the  $R_0$  value. For their analysis, the authors regarded strategies for mitigating pandemic influenza in the U.S. as successful when they limit the national attack rate to that of annual influenza epidemics, 10% of the U.S. population. It was observed that when intervention strategies were introduced, as  $R_0$  increases from 1.6 to 1.9, the outbreak transitioned from having the possibility of

being mitigated with moderate efforts to one of requiring vigorous application of multiple strategies to be abated. When  $R_0 > 1.9$ , the authors noted that a combination of behavioral changes, dynamic vaccination and antiviral drugs was necessary to reduce the spread of the disease. The authors also noted that a combination of vaccination of children and closing of schools did not offer improvements compared to implementation of just one or the other strategy. Lastly, they also observed that a combination of TAP, school closure and distancing without any vaccination was effective for  $R_0$  up to 2.4.

From their study, the authors noted that major efforts would be necessary to mitigate pandemic influenza, should it spread to the U.S., and consequently, they proposed that it would be most effective for influenza to be controlled at the source. They observed that if the virus did enter the U.S. population, a nationwide pandemic was likely to occur as soon as one month after the introduction of the virus into the population. The authors also observed that if a 90% restriction in air travel was enforced immediately after this, the reduction in the impact of the outbreak would only be minimal and delayed.

Nevertheless, the authors showed that their model was effective at capturing both the stochastic transmission processes that dominated the initial stages and final extinction of an outbreak and they detailed spatiotemporal dynamics of infectious disease spread. The authors did not model seasonal or environmental effects or viral evolution or incorporate disease-related mortality (because they assumed that deaths would occur only at the latter stages of the infectious period). They also assumed a constant contact, transmission, and disease course parameters throughout the U.S. for the entire duration of an influenza season. Also, the derivation of  $R_0$  was dependent on the generation time of the disease and the duration of the latent and infectious period. Because these values were unknown at the time of the study, the authors assumed particular values for them. Note that the analysis of the model presented by the authors involved use of parallel computing platforms.

23. Estimating the effective reproduction number for pandemic influenza from notification data made publicly available in real time: A multi-country analysis for influenza A/H1N1v 2009. (Hens, Van Ranst et al. 2010)

**What was forecast:** Real-time estimation of the effective reproductive number of the 2009 A/H1N1v influenza virus. The authors also analyzed the impact of regularization, time varying underreporting, delays in reporting, systematic reporting and the number of imported cases, using a simulation study.

**Disease agent:** 2009 H1N1 Influenza virus

**Type of model:** A Time series model.

**Description**

Using a method introduced by White and Pagano (*White and Pagano 2008*) and Fraser *et al.* (*Fraser, Donnelly et al. 2009*), the authors estimated the reproduction number from a time series of cases given an assumed serial interval distribution (the distribution of time periods between subsequent infections in a chain of transmissions). They also used a hybrid of the work done by both White and Pagano *et al.* and Fraser *et al.* to which they applied a fixed serial interval distribution and accounted for underreporting "while casting it in the likelihood (frequentist) framework." The authors estimated a global trend for the time-specific reproduction number,  $R_t$ , by estimating  $R_t$  without constraints "(in a so-called 'saturated' model where they obtained different estimates of  $R_t$  for each reported day) and compared this estimate to a regularized version using a penalized likelihood approach." (The regularization was done to overcome the effects of stochasticity.)

The main inputs for their model were data obtained of daily confirmed new influenza A/H1N1v cases, made publicly available on the websites of the World Health Organization (WHO), which recorded the number of accumulated cases for world, and the U.S. Centers for Disease Control and Prevention (CDC), which recorded the number of new daily cases in the U.S. These data were accessed daily during the duration of the epidemic. Data up to June 9th were included in the study and for the U.S., "probable" cases observed starting on 14th May 2009 were also included in the

study. The authors mainly focused on countries that reported cases since 11th May 2009 or earlier.

Given that the information on the disease course for the imported index case was likely to be unknown, assuming a 6-day truncated Weibull model with mean 2.67 and standard deviation 1.29 for the serial interval distribution, the authors used the first set of six non-zero new confirmed cases in a locality to start calculations for that locality (i.e., state, country or group of countries). When a burn-in period of 6 days (that is, the first 6 days at which more than 0 new cases had been reported in a country over a period during which each case was imported) and a time-independent underreporting factors were of 95% (i.e., assuming that only 5% of the total cases were actually reported) assumed, the authors examined the nature of the time-dependent reproduction number ( $R_t$ ) for Mexico, U.S., Canada and Australia and derived using the authors' saturated model vs. Fraser *et al.*'s piecewise-constant model. Plots of the time-dependent reproduction number ( $R_t$ ) revealed that the U.S. had the highest local maximum. They also showed that Mexico and Australia also had the steepest drop in  $R_t$  from the start of the calculation, probably owing to their linear growth curves.

To overcome the issue of irregular reporting of cases to the WHO and CDC, the authors pre-smoothed the cumulative number of cases based on a nonparametric model (a local quadratic smoother). They also used the nonparametric model to infer information on missing updates on the number of cases. Finally, to quantify the uncertainty associated with each estimation method, the authors used a parametric bootstrap approach to generate 95% confidence intervals.

To estimate the impact of delay in reporting on the estimation procedure, the authors simulated 100 epidemics with  $R_t = 1.2$  for  $t = 1$  to  $t = 50$  days for four different delay distributions.

A plot of the cumulative number of confirmed novel A/H1N1v cases per million over time (up to 11th June 2009) showed that countries in America and in the Pacific region had a higher incidence than countries in Europe and Japan.

Similar  $R_t$  plots for several states within the U.S. showed that after about day 30, the  $R_t$  value converged to a value slightly greater than 1, with Texas and Illinois showing large fluctuations.

The authors noted that their method was relatively insensitive to the magnitude of underreporting, but was sensitive to variations in the level of underreporting over time.

Their analysis of the effect of delay in reporting showed that the authors' non-pre-smoothed estimate of  $R_t$  showed a clear periodic effect for scenarios 1–3 whereas the pre-smoothed estimate of  $R_t$  resulted in a non-periodic smooth estimate at least after 10 days since the time of the first reported case. Following this first analysis of the effect of delay, the authors also assessed the impact on the estimated reproductive number when reporting happened every 2, 3 or 4 days. They found that without pre-smoothing, their estimation procedure broke down as soon as the reporting gap approached the mean serial interval of 2.67 days.

Next, the authors examined the impact of importations on the basic simulation without delay using a Poisson distribution with means 2, 5 and 15. Their results showed that when importations were high relative to domestic transmission,  $R_t$  was overestimated. They conclude that their results not reliable for localities with high numbers of importations combined with reporting delays.

The authors acknowledged that several of the assumptions they made might have affected their results. For instance, they assumed that there was no heterogeneity in age or space within a country. They also assumed that stochasticity did not play a significant role in the transmission process after the burn-in period. (Hens, Van Ranst et al. 2010)

## References

- Ajelli, M., S. Merler, et al. (2010). "Model predictions and evaluation of possible control strategies for the 2009 A/H1N1v influenza pandemic in Italy." Epidemiol Infect: 1-12.
- Arino, J., F. Brauer, et al. (2006). "Simple models for containment of a pandemic." J R Soc Interface **3**(8): 453-457.
- Bailey, N. T. J. (1975). The Mathematical Theory of Infectious Diseases and its Applications. New York, Macmillan Publishing Co., Inc.
- Balcan, D., V. Colizza, et al. (2009). "Multiscale mobility networks and the spatial spreading of infectious diseases." Proc Natl Acad Sci U S A **106**(51): 21484-21489.
- Balcan, D., H. Hu, et al. (2009). "Seasonal transmission potential and activity peaks of the new influenza A(H1N1): a Monte Carlo likelihood analysis based on human mobility." BMC Med **7**: 45.
- Bauch, C. T., J. O. Lloyd-Smith, et al. (2005). "Dynamically modeling SARS and other newly emerging respiratory illnesses: past, present, and future." Epidemiology **16**(6): 791-801.
- Boelle, P. Y., P. Bernillon, et al. (2009). "A preliminary estimation of the reproduction ratio for new influenza A(H1N1) from the outbreak in Mexico, March-April 2009." Euro Surveill **14**(19).
- Brauer, F. (2006). "Some simple epidemic models." Math Biosci Eng **3**(1): 1-15.
- Carley, K. M., D. B. Fridsma, et al. (2006). "BioWar: scalable agent-based model of bioattacks." Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on **36**(2): 252 - 265
- Cauchemez, S., P. Y. Boelle, et al. (2006). "Real-time estimates in early detection of SARS." Emerg Infect Dis **12**(1): 110-113.
- Cauchemez, S., P. Y. Boelle, et al. (2006). "Estimating in real time the efficacy of measures to control emerging communicable diseases." Am J Epidemiol **164**(6): 591-597.
- Colizza, V., A. Barrat, et al. (2007). "Modeling the worldwide spread of pandemic influenza: baseline case and containment interventions." PLoS Med **4**(1): e13.
- Cooley, P., B. Y. Lee, et al. (2010). "Protecting health care workers: a pandemic simulation based on Allegheny County." Influenza Other Respi Viruses **4**(2): 61-72.
- de Silva, U. C., J. Warachit, et al. (2009). "A preliminary analysis of the epidemiology of influenza A(H1N1)v virus infection in Thailand from early outbreak data, June-July 2009." Euro Surveill **14**(31).
- Elveback, L., J. Fox, et al. (1976). "An influenza simulation model for immunization studies." Am J Epidemiol **103**(2): 152-165.
- Ferguson, N., D. Cummings, et al. (2005). "Strategies for containing an emerging influenza pandemic in Southeast Asia." Nature **437**(7056): 209 - 214.
- Ferguson, N. M., D. A. Cummings, et al. (2006). "Strategies for mitigating an influenza pandemic." Nature **442**(7101): 448-452.
- Fraser, C., C. A. Donnelly, et al. (2009). "Pandemic potential of a strain of influenza A (H1N1): early findings." Science **324**(5934): 1557-1561.

- Germann, T., K. Kadau, et al. (2006). "Mitigation strategies for pandemic influenza in the United States." Proceedings of the National Academy of Sciences **103**(15): 5935 - 5940.
- Germann, T. C., K. Kadau, et al. (2006). "Mitigation strategies for pandemic influenza in the United States." Proc Natl Acad Sci U S A **103**(15): 5935-5940.
- Glass, R. J., L. M. Glass, et al. (2006). "Targeted social distancing design for pandemic influenza." Emerg Infect Dis **12**(11): 1671-1681.
- Guttorp, P. (1991). Statistical Inference for Branching Processes. New York, Wiley.
- Hall, I. M., R. Gani, et al. (2007). "Real-time epidemic forecasting for pandemic influenza." Epidemiol Infect **135**(3): 372-385.
- Hens, N., M. Van Ranst, et al. (2010). "Estimating the effective reproduction number for pandemic influenza from notification data made publicly available in real time: A multi-country analysis for influenza A/H1N1v 2009." Vaccine.
- Jewell, C. P., T. Kypraios, et al. (2009). "A novel approach to real-time risk prediction for emerging infectious diseases: a case study in Avian Influenza H5N1." Prev Vet Med **91**(1): 19-28.
- Kelso, J. K., G. J. Milne, et al. (2009). "Simulation suggests that rapid activation of social distancing can arrest epidemic development due to a novel strain of influenza." BMC Public Health **9**: 117.
- Lee, B. Y., V. L. Bedford, et al. (2008). "Virtual epidemic in a virtual city: simulating the spread of influenza in a US metropolitan area." Transl Res **151**(6): 275-287.
- Longini, I., A. Nizam, et al. (2005). "Containing pandemic influenza at the source." Science **309**(5737): 1083 - 1087.
- Longini, I. M., Jr., M. E. Halloran, et al. (2004). "Containing pandemic influenza with antiviral agents." Am J Epidemiol **159**(7): 623-633.
- Longini, I. M., Jr., A. Nizam, et al. (2005). "Containing pandemic influenza at the source." Science **309**(5737): 1083-1087.
- Milne, G. J., J. K. Kelso, et al. (2008). "A small community model for the transmission of infectious diseases: comparison of school closure as an intervention in individual-based models of an influenza epidemic." PLoS One **3**(12).
- Ong, J. B., M. I. Chen, et al. (2010). "Real-time epidemic monitoring and forecasting of H1N1-2009 using influenza-like illness from general practice and family doctor clinics in Singapore." PLoS One **5**(4): e10036.
- Roberts, M. G., M. Baker, et al. (2007). "A model for the spread and control of pandemic influenza in an isolated geographical region." J R Soc Interface **4**(13): 325-330.
- Tang, S., Y. Xiao, et al. (2010). "Community-based measures for mitigating the 2009 H1N1 pandemic in China." PLoS One **5**(6): e10911.
- Wallinga, J. and M. Lipsitch (2007). "How generation intervals shape the relationship between growth rates and reproductive numbers." Proc Biol Sci **274**(1609): 599-604.
- Wallinga, J. and P. Teunis (2004). "Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures." Am J Epidemiol **160**(6): 509-516.
- White, L. F. and M. Pagano (2008). "A likelihood-based method for real-time estimation of the serial interval and reproductive number of an epidemic." Stat Med **27**(16): 2999-3016.

Wood, J. G., N. Zamani, et al. (2007). "Effects of internal border control on spread of pandemic influenza." Emerg Infect Dis **13**(7): 1038-1045.

Wu, J. T., S. Riley, et al. (2006). "Reducing the impact of the next influenza pandemic using household-based public health interventions." PLoS Med **3**(9): e361.