

Stochastic individual-based modelling of influenza spread for the assessment of public health interventions

J. K. Kelso^a and G. J. Milne^a

^a *University of Western Australia, School of Computer Science and Software Engineering*
Email: joel@csse.uwa.edu.au

Abstract: Infectious respiratory diseases that are spread by interpersonal contact, such as influenza, tuberculosis and measles, cause a significant burden of disease worldwide. Influenza causes annual epidemics resulting in the infection of approximately 5-10% of adults and 20-30% of children worldwide, causing 250,000 to 500,000 deaths annually as estimated by the World Health Organization [WHO 2009]; and the possibility exists that a pandemic strain may arise to which humans may have little immunity, which could be considerably more deadly and infect a high proportion of the global population.

Various pharmaceutical (vaccines and antiviral drugs) and non-pharmaceutical (personal hygiene practices, and social distancing such as closure of schools) interventions exist to mitigate the effects of influenza epidemics or pandemics. Although the individual-level effect of interventions can be estimated (vaccine efficacies can be estimated from clinical trials, for example, and an individual isolated at home can only infect other co-inhabitants), the interplay of infection spread and human behaviour makes quantitatively predicting the population-level outcome of interventions difficult. As experiments are clearly impossible, and as data on the application and effectiveness of interventions in past pandemics is sparse, modelling is an essential technique for assessing the effectiveness of proposed intervention policies.

One method of modelling infectious disease spread is to simulate the infection state and interpersonal interactions of individual humans. Our research group has constructed a stochastic individual-based simulation model for influenza spread based on a real community of approximately 30,000 people, and have used it to examine a wide range of epidemic and public health intervention scenarios such as vaccination, antiviral therapy and prophylaxis, and various types of social distancing. The key to a successful modelling exercise is to find a set of abstractions that capture all and only the relevant features of the system being studied. These abstractions need to be detailed and extensive enough to represent the significant features and processes of the real world while remaining simple enough to allow tractable analysis and/or simulation. In this presentation we describe our choice of abstractions for the influenza-modelling problem, including the scale (in terms of size of simulated population) and duration of simulated epidemics, temporal and spatial granularity, state information of individuals, and idealisations of human behaviour. We describe the software system that implements this abstract model, and also remark on some of the advantages and disadvantages of the individual-based approach to infectious disease spread modelling.

Keywords: *Influenza, infectious disease modelling, agent based modelling*

1. MODEL STRUCTURE AND ABSTRACTIONS

In this section we describe each of the real-world aspects that govern the spread of influenza, and explain how they are represented in the simulation model. We give a rationale for the particular chosen representation and, where possible, give reference to the data sources that informed that part of the model. Space precludes an exhaustive description of the simulation model; we refer the reader to published studies that utilized the model (see Section 4).

1.1. Population Model

We chose to model the population of the town and local government area of Albany, a town on the south coast of Western Australia having a population of approximately 30,000 people circa 2004. Census data from the Australian Bureau of Statistics (ABS) [Australian Bureau of Statistics 2004], including Confidentialised Unit Record Files (CURF) household data was used to assign an age and employment status to each individual, and to build a list of households. The model was further populated with a set of schools and workplaces, referred to collectively as contact hubs. Data from the Western Australian Department of Education were used to obtain a complete list of schools (including student age information), childcare facilities, and adult education institutions. Each child was assigned to a school or childcare centre, presuming that children attended school as close to their home location as possible, presuming that the known age structure of schools was maintained. Business survey information from the Western Australian Department of Planning and Infrastructure provided a comprehensive list of employers including their locations and nominal daytime populations. Workers and adult students were assigned to workplaces and adult education institutions respectively, with this assignment being made with reference to journey-to-work information survey data from the Western Australian Department of Planning and Infrastructure.

The scale of population was chosen because it represents a “complete” community, in that it included residents, schools, local employment and a hospital. The local government area also marks a level at which public health decision-making is likely to occur for school / college / employer based interventions.

Albany is effectively self-contained for epidemic modelling purposes. The town’s major traffic partner is the city of Perth – which is much larger (population approximately 1.6 million) and has much greater interstate and international traffic. Epidemic spread in Albany is therefore highly unlikely to influence dynamics of epidemic spread in Perth, so interaction can be modelled by a deterministic inflow of infected cases caused by travelers from Perth.

1.2. Temporal Resolution and Scale

Each simulation proceeded in a sequence of 12 hour day / night cycles. Time steps longer than 12 hours would have been computationally advantageous, but would have risked misrepresenting infection progression dynamics, in which significant changes take place on the timescale of approximately 12 hours. Individual behaviour does vary on shorter timescales – movement and meetings between people in workplaces, for example. However, not having any reliable data on the details of such behaviour, we model aggregate contact behaviour over a 12-hour period, as described below.

Simulations ran until no more infected individuals were left in the population, or until a fixed time limit (generally 1 year) expired.

1.3. Individual Movement and Contact Behaviour

Influenza spreads via interpersonal contact, with droplets containing virus particle exhaled or expelled from an infected individual being inhaled by another. This process will clearly be influenced by the situational details such as duration and proximity of contact; however it would be computationally prohibitive to model this process, assuming that data were to exist upon which a detailed contact model could be based. Instead, we define a nominal “contact” event to represent significant contact with an individual within a 12-hour period, and define a contact-context transmission probability factor that modifies the transmission probability depending on the context of contact (e.g. school, household, etc).

Three different types of contact between individuals were modelled. First, during every cycle all individuals in the same household came into contact. Secondly, during the daytime cycle, active individuals (a term defined below) came into contact with other active individuals who shared the same contact hub (school or workplace). Thirdly, active individuals made a random contact with other active individuals, potentially from

anywhere else in the population, but weighted towards individuals with nearby home locations. This “community” contact mechanism represented all other sources of contact that was not repeated on a daily basis – while out shopping, or on public transport for example. Interestingly, provided that there was either at least a small amount of hub mixing or non-local community mixing, the actual spatial pattern of community mixing did not influence the epidemic outcome.

Note that in the case of large schools and workplaces, individuals made infective contact with a fixed sub-group (the size of which was parameterized) rather than with the whole hub population.

Interestingly, the model parameters relating to contact behaviour are among those least constrained by observational data. Data from mobile phone movement data by confirms that human movement (at least in the developed world) occurs predominantly according to a 24 hour cycle [Gonzalez *et al.* 2008]. Some contact survey data that reports on number and “closeness” of daily contacts is available, which could potentially be used to further constrain assumptions about interpersonal contact [Wallinga *et al.* 1999; Glass *et al.* 2008; Mossong *et al.* 2008]. In calibrating these parameters we made use of seasonal influenza household survey data that estimated the proportion of in-household to out-of-house transmission [Monto *et al.* 1985; Ferguson *et al.* 2006]; more sophisticated calibration might be possible using similar additional data sources.

1.4. Infection Transmission

Transmission of infection between infectious and susceptible individuals who came in contact was resolved stochastically. The probability of transmission was calculated as a function of the state of the infectious (I_i) and susceptible (I_s) individuals involved at the time of contact, as given by:

$$P_{\text{transmission}}(I_i, I_s) = \beta * Susc(I_s) * Trans(I_i)$$

The basic transmission probability (β), capturing the infectivity of the virus strain, was calibrated to give epidemics with chosen target basic reproduction numbers R_0 or attack rates. The function *Susc* represents differential susceptibility of individuals of different ages, and was taken to incorporate both pre-existing immunity due to prior infection and differences in contact behaviour. *Susc* was calibrated to achieve age-specific attack rates recorded for various previous seasonal and pandemic influenza strains [Monto *et al.* 1985; ECDC 2009]. The function *Trans* represents the time-varying infectivity of an individual as their infection progresses (see below). This basic model was enhanced with parameters representing the efficacy of antiviral treatment, antiviral prophylaxis and vaccination.

1.5. Seeding Model

Epidemics were initiated by choosing one individual randomly from the population per day and infecting that person if they were susceptible, modelling infections imported into Albany via travelers (or returning residents) who had become infected elsewhere. The simulator could also be used to assess the probability of a sustained epidemic given particular seeding scenario; for example, a single infected traveler, or a returning household or school class with multiple infected members.

1.6. Infection Progression

The progression of influenza infection within an individual was modelled as a deterministic time-line, consisting of several phases. The duration of each phase was parameterized; the values given here were based on observed symptom and viral shedding data reported in [Carrat *et al.* 2008]. During the initial 12-hour latent phase the individual was asymptomatic, non-infectious, and behaved normally. From 12 to 36 hours post infection individual were deemed asymptomatic but partially (0.5) infectious. From 36 to 84 hours post infection individuals were fully infectious and symptomatic, from 84 hours to 144 hours individuals were symptomatic but partially (0.5) infectious (following the assumptions of {Ferguson, 2006 #18}). After this time individuals recovered, and were assumed to be immune to re-infection for the duration of the simulation. It was also assumed that infected individuals could (with an age-based parameterized probability) experience an effectively asymptomatic infection, in which case the transmissibility was fixed at 0.5 for the entire infectious (post 12-hour) period.

In addition to the changes in transmissibility due to virus shedding, infection was assumed to modify contact behaviour – symptomatic individuals were assigned randomly to either “soldier on” and behave normally, or withdraw to their household for the duration of their symptoms (with probability 50% for adults and 90% for, with a sensitivity analysis examining higher and lower values). Withdrawn individuals were thus “inactive”

and made no hub or community contacts. Additionally, whenever a child is at home during a daytime cycle (whether due to illness or school closure), an adult in the household is also assumed to stay at home.

1.7. Modelling Interventions

The model of the population, movement and mixing behaviour, infection transmission and infection progression described above is sufficient to allow the modelling of a wide range of public health intervention strategies in a straightforward way. The simulation model exposes the different mechanisms that each intervention targets, allowing the model of intervention to modify the relevant mechanism while the intervention is in effect. Full details of each intervention measure can be found in the references given in Section 4 – here we give an outline of each.

Increased Isolation

This intervention assumes that a public health education campaign (or, more simply, public awareness of a deadly pandemic) increases the chance that symptomatic individuals will stay at home rather than make hub or community contacts. Under this intervention the proportion of adults and children who stayed at home upon developing symptoms was 90% and 100% respectively - a sensitivity analysis examined compliance levels from 0-100%.

School Closure

During periods of school closure teachers and school children become inactive and withdraw to their household. High compliance is assumed, with no extra compensatory contacts being made. In addition, it was assumed that at least one adult would stay at home (and thus not attend a workplace) if school closure would otherwise leave a child unattended at home. Later extensions to the model allow for more complex models of school closure, including single classes sent home, single schools, or all schools in the district – all of which were used during the 2009 pandemic in various locations.

Workplace Nonattendance

During this intervention it was assumed that each workday each worker would randomly choose (according to a parameterized probability) to stay at home. This applied to all workers who would normally have attended work (not those who were already determined to be at home due to illness or child supervision).

Community Contact Reduction

During this intervention, active individuals reduced the number of random community contacts made each day, modelling a public health education campaign or possibly closure or cancellation of public venues or events.

Antiviral Medication

The application of antiviral interventions assumed that a certain proportion of symptomatic cases would come to the attention of the public health system (either through medical practices, hospital clinics, or pharmacy-related scheme), whereupon symptomatic individuals would be given antiviral treatment that reduced infection duration and transmissibility. In addition, if antiviral prophylaxis measures were in effect, the symptomatic individuals household and/or hub contacts would be given antivirals in a prophylactic capacity, which was assumed to reduce susceptibility while the course of drugs was being taken (generally for 10 days).

Vaccination

A sophisticated vaccination model was added to the model, capable of modelling age-specific vaccine efficacy, limited vaccine supply prioritization by age group, pre-vaccination versus reactive roll-out of vaccine, single-dose versus two-dose vaccines, and differing assumptions on the rate of acquisition of immunity following vaccination.

2. SOFTWARE SYSTEM

The simulation model has been realized as a custom C++ application. Each simulation is specified by three input data files: a file describing the demographic structure of the population, including individual ages, households, and mixing hub groups; a community contact matrix that specifies the probabilistic pattern of random community contacts based on household locations; and a parameters file that specifies the (numerous) simulation parameters governing all aspects of the model described previously.

Each simulation run proceeds in a straightforward way. For each 12-hour cycle, new infections are seeded into the population according to the specified seeding model (as described in section 2.5). The infection and behavioural state of each individual is calculated. The locations occupied by each individual during the cycle (the household, possibly a work or school hub, and possibly one or more locations for random community contact), and the possibility of transmission is resolved for all pairs of susceptible and infected individuals who come into contact.

Since the simulation model is stochastic, each simulation (with a fixed set of inputs and parameters) was repeated multiple times (usually 40) with a different random number sequence in order to obtain average outcomes and outcome distributions. The number of repeats was chosen to give 95% confidence that the multiple-run mean of the final attack rate was within 1% of the mean of the final attack rate distribution. For the scenarios with the widest distributions (which happen to be those with the lowest reproduction numbers) 40 runs was sufficient to achieve this. Further, the model was used to run series of experiments, systematically varying one or more parameters at a time. These repeated simulations are all independent and could be run concurrently.

The simulation model itself represents a highly concurrent system: infection transmission within each household could be resolved independently, for example. However as the repeated simulations mentioned above are already an embarrassingly parallel workload, parallelization within a single simulation was not attempted.

During each simulation run the simulator accumulates statistics on the ongoing epidemic, and it also writes a complete log of significant events (infection transmissions, work or school days lost due to illness or interventions, antiviral doses dispensed). Plain text file (comma separated value CSV) formats were used for input of simulation specifications and output of results. There was no requirement for a graphical user interface, and filter operation allowed the simulator to be easily incorporated into a scripting system.

The scripting system automated several parts of the experimentation process: virtual experiment series where one or more parameters were systematically varied, and the calibration processes whereby model parameters were calibrated to achieve epidemics with certain characteristics (basic reproduction numbers or age-specific attack rates). The calibration procedures used were fairly simple, resulting in point estimates from the parameter space. More sophisticated statistical inference techniques could be used to infer a (joint) probability distribution over the parameters, given the target epidemic characteristics.

As measure of performance, a single simulation run of the 30,000 member population for a simulated year (730 cycles) took approximately 10 seconds on a 3.8 GHz (Pentium D) CPU. The tasks were CPU-bound.

3. MODELLING OUTCOMES

The epidemic simulator has proven to be a valuable test bed for modelling a wide range of proposed public health interventions. This has allowed the interventions, as well as combinations of interventions, and various logistical details of intervention deployment to be assessed as responses to influenza pandemics with varying characteristics.

While space precludes an extensive description of the results obtained, some of the outcomes of these assessments are summarised below.

- Multiple social distancing measures applied early and continuously can be very effective in interrupting transmission of a pandemic virus for R_0 values up to 2.5 [Milne *et al.* 2008].
- The effect of different delays in implementing social distancing interventions was quantified; for example, for a highly transmissible pandemic virus ($R_0 = 2.5$) final attack rates could increase by a factor of six when intervention delays increased from 2 to 4 weeks [Kelso *et al.* 2009].
- The effectiveness of three different vaccination strategies was explored: pre-vaccination, reactive vaccination after the first cases were detected, and a split strategy of pre-vaccination followed by a reactively applied booster dose). It was shown that if pre-pandemic vaccine can achieve an immunological priming effect (which may be possible even with an imprecisely matched), the split vaccination strategy can speed the effective acquisition of immunity by 4 weeks, which may be invaluable given the time required for vaccine development [Milne *et al.* 2010].
- Various responses reported to have been enacted during the A/H1N1 2009 pandemic were simulated, including different antiviral strategies and different types and duration of school closure. It was found that, in general, interventions would have to be more rigorous (involving longer school

closure and more extensive use of antiviral prophylaxis) than those employed in 2009 to have a substantial effect [Halder *et al.* 2010a].

- It is hypothesised that if interventions of limited duration are employed, it may be optimal to delay the activation of interventions rather than to enact them immediately. The simulation model was used to quantify this phenomena and offer practical advice on how interventions could be activated based on daily diagnosed case counts [Halder *et al.* 2010b].
- The simulation model was used to quantify the degradation of effectiveness of antiviral interventions caused by diagnosis delay and low diagnosis coverage (the proportion of symptomatic cases that come to the attention of the public health system) [Kelso *et al.* 2010].
- By augmenting the simulation model with a health outcomes and economic costing model, the cost-effectiveness of various interventions employed in the A/H1N1 2009 pandemic were investigated. The mild severity (low hospitalization and case fatality ratio) meant that health care treatment costs were dominated by productivity losses arising from workplace absence due to illness and childcare requirements following school closure. Strategies involving antiviral treatment and prophylaxis accompanied by limited duration school closure were found to be the most cost effective amongst those simulated [Halder *et al.* 2011].

4. DISCUSSION AND CONCLUSIONS

The simulation model described in this paper has allowed significant research questions about the spread and control of pandemic influenza to be addressed. The scale and level of abstraction of the model has allowed the essential details of influenza spread (as well as various intervention measures intended to mitigate against influenza spread) to be captured while at the same time allowing comprehensive experiment series with stochastic repetitions to be run comfortably on desktop PC.

Having constructed, extended and utilized an individual-based epidemic spread simulation model, we can offer some comment on the potential advantages and disadvantages of this modelling approach, as compared to other approaches such as the more established techniques of modelling with differential equations [Anderson *et al.* 1991].

Firstly, the individual or (agent-based) approach comes with a higher computational cost. If D is the simulated duration of an epidemic (in days, for example), then a differential equation model can be numerically integrated in time $O(D)$, or it may have an analytical solution that (once found) allows outcomes to be computed $O(1)$. In contrast, an individual-based model may require $O(D * P)$, where P is the population size. As we have demonstrated, this is however not a significant limitation, as the problem is highly concurrent, and many (if not most) important phenomena can be modelled at a community scale and do not require, for instance, whole-of-country or global scale models.

Secondly, it is challenging to give a description of a complex simulation model that is both complete and comprehensible to a wide audience. While a differential equation model can be exactly described in several pages of differential equations, no such widely understood description language exists for simulation models. We have taken the approach of giving an informal description that should still allow a skilled software developer to produce a simulation model that is effectively equivalent. Alternative approaches might be to use a formal software modelling language (which unfortunately are not in widespread use in scientific modelling); or to publish simulation model source code, which would explicate the full model to at least a limited audience.

Thirdly, an individual-based simulation model can make use of detailed data sources where they exist – for example, detailed census or survey data. It might be argued that for practical purposes that fine details (the makeup of individual households, for example) will be irrelevant in the averaged outcomes of stochastic simulations, and that simpler model with aggregated details would be equivalent and more parsimonious. However, it is not clear *a priori* at what level aggregation can take place without changing simulation results, especially as this may depend on future extensions to the model. All other things being equal, having more accurate data in the simulation should not make the results less accurate.

Finally, we have found that once the basic objects of abstraction – individuals, households, workplaces, schools, CCD locations – had been chosen, there were straightforward ways of representing the real-world phenomena that drive influenza epidemic spread. The model also proved readily extensible: later elaborations to the initial model, including intervention triggering and relaxation conditions, vaccination, and antiviral treatment and prophylaxis, were incorporated in similarly straightforward ways.

REFERENCES

- Anderson, R. M. and R. M. May (1991). Infectious Diseases of Humans: Dynamics and Control. Oxford, Oxford University Press.
- Australian Bureau of Statistics (2004). "The 2004 Australian Census." Retrieved 10 November 2008, from <http://www.abs.gov.au>.
- Carrat, F., E. Vergu, et al. (2008). Time lines of infection and disease in human influenza: a review of volunteer challenge studies. *American Journal of Epidemiology* 167(7): 775-785.
- ECDC (2009). ECDC Risk Assessment 2009 influenza A(H1N1) pandemic Version 7. ECDC Risk Assessment. Stockholm, European Center for Disease Control.
- Ferguson, N. M., D. A. T. Cummings, et al. (2006). Strategies for mitigating an influenza pandemic. *Nature* 442: 448-452.
- Glass, L. M. and R. J. Glass (2008). Social contact networks for the spread of pandemic influenza in children and teenagers. *BMC Public Health* 8.
- Gonzalez, M., C. Hidalgo, et al. (2008). Understanding individual human mobility patterns. *Nature* 453: 779-782.
- Halder, N., J. Kelso, et al. (2010a). Analysis of the effectiveness of interventions used during the 2009 H1N1 influenza pandemic. *BMC Public Health* 10: 168.
- Halder, N., J. Kelso, et al. (2010b). Developing guidelines for school closure interventions to be used during a future influenza pandemic. *BMC Infectious Diseases* 10(1): 221.
- Halder, N., J. Kelso, et al. (2011). Cost-effective strategies for mitigating a future influenza pandemic with H1N1 2009 characteristics. *PLoS ONE* 6(7): e22087.
- Kelso, J. K., N. Halder, et al. (2010). The impact of case diagnosis coverage and diagnosis delays on the effectiveness of antiviral strategies in mitigating pandemic influenza A/H1N1 2009. *PLoS ONE* 5(11): e13797.
- 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.
- Milne, G., J. Kelso, et al. (2010). Strategies for mitigating an influenza pandemic with pre-pandemic H5N1 vaccines. *Journal of the Royal Society Interface* 7: 573-586.
- 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 pandemic. *PLoS ONE* 3(12): e4005.
- Monto, A., J. Koopman, et al. (1985). Tecumseh study of illness XIII. Influenza infection and disease, 1976-1981. *American Journal of Epidemiology* 121(6): 811-822.
- Mossong, J., N. Hens, et al. (2008). Social contacts and mixing patterns relevant to the spread of infectious diseases. *PLOS MEDICINE* 5(3): 381-391.
- Wallinga, J., W. J. Edmunds, et al. (1999). Perspective: human contact patterns and the spread of airborne infectious diseases. *TRENDS IN MICROBIOLOGY* 7(9): 372-377.
- WHO (2009). "Influenza (Seasonal) Fact sheet number 211." Retrieved Oct 5, 2011, from <http://www.who.int/mediacentre/factsheets/fs211/en/>.