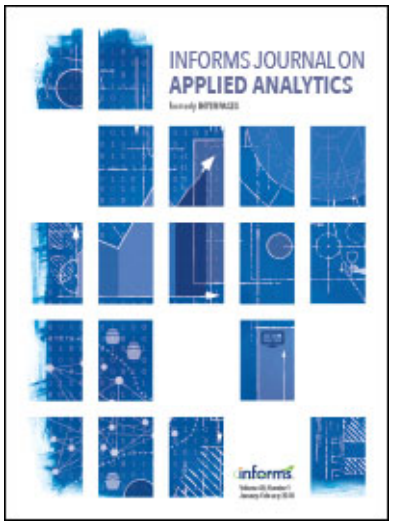

INFORMS Journal on Applied Analytics

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

A Nonhomogeneous Agent-Based Simulation Approach to Modeling the Spread of Disease in a Pandemic Outbreak

Dionne M. Aleman, Theodorus G. Wibisono, Brian Schwartz,

To cite this article:
Dionne M. Aleman, Theodorus G. Wibisono, Brian Schwartz, (2011) A Nonhomogeneous Agent-Based Simulation Approach to Modeling the Spread of Disease in a Pandemic Outbreak. INFORMS Journal on Applied Analytics 41(3):301-315. https://doi.org/10.1287/inte.1100.0550

Interfaces
Vol. 41, No. 3, May–June 2011, pp. 301–315
ISSN 0092-2102 | EISSN 1526-551X | 11 | 4103 | 0301

informs

DOI 10.1287/inte.1100.0550

A Nonhomogeneous Agent-Based Simulation Approach to Modeling the Spread of Disease in a Pandemic Outbreak

Dionne M. Aleman, Theodorus G. Wibisono
Department of Mechanical and Industrial Engineering, University of Toronto, Toronto, Ontario M5S 3G8, Canada
{aleman@mie.utoronto.ca, theo.wibisono@utoronto.ca}

Brian Schwartz
Ontario Agency for Health Protection and Promotion, Toronto, Ontario M5G 1V2, Canada; Department of Family and Community Medicine, University of Toronto, Toronto, Ontario M5S 3G8, Canada, brian.schwartz@oahpp.ca

To effectively prepare for a pandemic disease outbreak, knowledge of how the disease will spread is paramount. The global outbreak of severe acute respiratory syndrome (SARS) in 2002–2003 highlighted the need for such data. This need is also apparent in preparing for and responding to all disease outbreaks, from pandemic influenza to avian flu. Many previous studies of disease make simplistic assumptions about transmission and infection rates and assume that each member of the population is identical or homogeneous. We propose an agent-based simulation model that treats each individual as unique, with nonhomogeneous transmission and infection rates correlated to demographic information and behavior. The results of the model are output to geographic information system software to provide a map of the estimated disease spread area, which can be used as a policy-making tool for determining a suitable mitigation strategy. The Ontario Agency for Health Protection and Promotion (OAHPP) uses the model for pandemic planning for the Greater Toronto area in Ontario, Canada.

Key words: pandemic; influenza; SARS; disease spread; agent-based simulation; nonhomogeneous mixing model.
History: This paper was refereed.

Pandemic influenza outbreaks have been associated with both moderate and severe morbidity and mortality (Potter 1998); the severe acute respiratory syndrome (SARS) outbreak between November 2002 and July 2003 resulted in 8,096 infected individuals and 774 deaths (Public Health Agency of Canada 2003) worldwide. Diseases are spread by viral particle transmission from an infected individual to a susceptible individual by various mechanisms. Mitigation strategies to vaccinate susceptible individuals and treat infected individuals are critical for controlling the spread of a disease, particularly in urban areas where population density can expedite disease propagation.

Components of mitigation strategies include the determination of how to deploy finite prophylaxis resources, how to handle increased numbers of patients at hospitals, when to close schools, and when to enforce quarantines. Mitigation strategies must balance limited resources among members of the population and decide which individuals should be prioritized to most quickly stop the spread of disease. Without a model of how the disease will spread according to the specific individuals affected by the mitigation strategy, there is no way other than intuition to establish which strategy is the most effective.

Models to predict the spread of a disease have largely simplified the reality of populations (i.e., individuals' random contacts and behaviors) and the mechanisms by which diseases are transmitted. The most commonly used model, the homogeneous mixing model, assumes that the population is fully mixed (i.e., individuals with whom a susceptible individual has had contact are chosen at random from the whole population); all individuals have approximately the same number of contacts in the same period, and all

contacts transmit the disease with the same probability (Newman 2002).

Despite these generalizations, homogeneous mixing models are the traditional approach in predicting the number of individuals affected by disease outbreak and are used as the basis for mathematical models for the spread of the disease (Eubank et al. 2004). In these models, the propagation of disease transmission is based on the reproduction number (R), the number of individuals that are infected by a single case. In initial cases, it is known as the basic reproduction number, R_0, defined by $R_0 = kbD$, where k is the number of contacts for each individual per unit time, b is the probability of transmission per contact between infectious and susceptible individuals, and D is the mean duration of infectiousness (Lipsitch et al. 2003).

Although this model continues to provide useful insight and predictive capability (Burr and Chowell 2008, Halloran et al. 2008), together with the added benefit of closed-form solutions, in real life it is not necessarily a useful tool for designing mitigation strategies; however, some previous studies, such as the models presented in Ferguson et al. (2005), are capable of testing some containment measures. Consider a situation in which most infected individuals transmit the disease to only one other (or no others), and another situation in which small numbers of infected individuals transmit the disease to dozens or hundreds of others. The mean value of R_0 can be the same in both cases, although the epidemiological outcomes are vastly different (Meyers et al. 2005). The resulting predictions are "grossly inaccurate in at least some cases" (Newman 2002, p. 604) or may lead to spurious estimates that cannot justifiably be extrapolated from the specific settings in which they were measured to the broader community context (Meyers et al. 2005).

Three factors that allow for disease transmission are ambient environment, the infectivity of the affected individual, and the vulnerability of the susceptible individual (Cliff et al. 1986). However, homogeneous mixing models do not allow these factors to be considered. The real world is better represented by a nonhomogeneous mixing model (Longini et al. 2005, Del Valle et al. 2007, Meyers et al. 2005). Interactions are represented in a contact network, where members of the population are nodes and contacts between members are arcs. Based on the contact of each pair of individuals in a population, contact networks allow for the reproduction number to be individualized. This method is believed to provide a better prediction of the benefits of various mitigation strategies than the homogeneous mixing model (Burr and Chowell 2008).

Nonhomogeneous models can also account for environmental conditions. Low humidity and low temperature are conducive to virus survival and transfer (Cliff et al. 1986), and hard, nonporous surfaces (e.g., stainless steel and plastic) increase survival time more than porous surfaces (e.g., cloth, paper, and tissues) (Bean et al. 1982). Nonporous surfaces can be found in large quantity in work spaces and public transport vehicles (e.g., subways, buses, and streetcars).

Despite the versatility of nonhomogeneous models, they generally require complicated mathematics to obtain useful information about disease spread (e.g., the number of infections/deaths per day and the total number of infected/dead). Conversely, simulation methods are commonly used to generate hypothetical outcomes of complex systems by simulating events each period (e.g., daily, hourly, or other) according to the likelihood of each event. Outcomes are tallied throughout the simulation. Although simulations are simple to implement and modify, and almost any metric can be recorded, they potentially can require too much time to complete to be useful in a real-world setting. However, advances in computing technology have greatly increased simulation speeds and the size of processes that can be modeled. Simulation models have been applied to disease outbreaks (Das et al. 2008; Ferguson et al. 2005, 2006; Germann et al. 2006; Halloran et al. 2008; Lee et al. 2008; Longini et al. 2004; Patel et al. 2005; Shi et al. 2010; Wu et al. 2006). However, many of these approaches are based on R_0 (Ferguson et al. 2005, 2006; Germann et al. 2006; Longini et al. 2005; Shi et al. 2010; Wu et al. 2006), whereas the approach by Lee et al. (2008) assumes homogeneous transmission rates.

Eubank (2005) provides a detailed representation of a specific urban area in a disease outbreak, including steps such as population synthesis, activity assignment, location choice, and travel-time estimation. We use these steps as the foundation in our research.

We use a nonhomogeneous agent-based simulation, as Das et al. (2008), Halloran et al. (2008), Shi et al. (2010), and Stroud et al. (2007) describe, to provide a realistic model of disease spread. An agent-based simulation will allow each member of a population to be simulated individually, thereby providing a means to consider the unique individual characteristics that affect transmission and infection probabilities. Individual behaviors that dictate type and length of contact between two individuals can also be modeled, as well as the possibility of infected individuals recovering and becoming immune, as Kilbourne's model (Anderson and May 1991) explains.

The Ontario Agency for Health Protection and Promotion (OAHPP) has two major concerns about preparing for a disease outbreak: knowing how the disease will spread and which mitigation strategies will be most effective. Although many types of disease-spread models already exist, none is exactly like the one we present; the specific purpose of our model is to facilitate the testing of mitigation strategies. Every aspect of the model, from the selection and construction of agents to the generalization of contact between individuals to parallelized C++ implementation, is designed to incorporate the effects of potential mitigation strategies. The model is designed for use as a tool to compare the relative differences in disease spread between different mitigation strategies.

In the mitigation-strategy scenarios, a model that uses groups to represent the population would have to implement groups for every possible combination of factors. As new factors are considered in mitigation strategies, the groups would have to be revised, and the number of groups would increase exponentially with each new factor. Because our major concern is to devise a framework that can test any mitigation strategy, we use individuals, rather than groups, as agents in our model. Each individual is assigned an amount of contact time with each other individual based on demographic information (e.g., using the same public transportation route or living in the same household). Additionally, a rate of disease transmission per unit time is assigned according to the type of contact two individuals have. The amount of contact time and rate of disease transmission can change from day to day as individual behaviors change (e.g., being under quarantine or seeking medical care).

The benefit of our model is that because it focuses on individuals and the unique contact between any two members of the population, it is flexible. If the epidemiology of the outbreak suggests that ethnicity or gender has an impact on vaccine effectiveness, or that socioeconomic status has an impact on compliance with quarantines, those factors can be incorporated with minimal effort. As long as census data or expert opinion on behaviors are available, the model can address any consideration. We also record individuals' locations in the model, and results from the model are output to geographic information system (GIS) software for visualization. The visual maps of the disease spread can be used to develop and test appropriate mitigation strategies.

The EpiSimS model that Stroud et al. (2007) present is similar in conceptual design to our model. EpiSimS emphasizes work and commerce interactions; unfortunately, such detailed data on population movement in Ontario are not available. One aspect that our model considers that EpiSims does not yet address is the use of public transportation, which is heavily used in Toronto, unlike in California where the EpiSimS model was implemented.

The model presented in Das et al. (2008) is also conceptually similar to our model, although the authors do not specifically state that they employ agent-based simulation. Despite the similarity of our model to this model, our variation on the concept outperforms their model because we can consider a larger population in more detail and in less time. The largest population size tested in Das et al. (2008) had approximately 1.1 million individuals in 400,000 households and required 50 minutes for one simulation day. Our model is currently being tested in the Greater Toronto Area, which consists of almost five million individuals in 1.8 million households, and runs one 60-day simulation in less than two minutes (1,000 simulations require about 13 minutes when we run them using our parallel computing infrastructure). The models presented by Halloran et al. (2008) and Shi et al. (2010) also consider large populations; however, it is unclear how Halloran et al. (2008) obtain the transmission probabilities and address the interactions among the population. Moreover, Shi et al. (2010) use R_0 as the basis of disease transmission rather than individual transmission probabilities.

One clear shortcoming of both the EpiSimS model and our model is that rates of transmission per unit time are not readily available for most diseases. However, models relying on R_0 suffer from a similar shortcoming, because R_0 is not readily available during an outbreak. Even if we assume that R_0 is an acceptable one-size-fits-all measurement of disease spread, this number is very difficult to calculate, as demonstrated by the fact that for the 2009 H1N1 pandemic, R_0 required more than two months to determine. In addition, it differs in different locales, that is, one model relying on a reproduction number cannot represent disease spread if the environment (e.g., urban or rural) changes within the area being modeled.

Simulation Model

According to the susceptible, infectious, and removed (SIR) model, each individual is classified as either susceptible (i.e., not infected), infectious, or removed. In the model, each individual can transition to another state in each period. We can think of this as a unique Markov chain (see Figure 1) for each individual. Each susceptible individual has a unique possibility of transitioning to an infected state, or staying susceptible.

Similarly, each infected individual has the probability of transitioning to a removed/recovered state, or of staying infected; these probabilities are determined by the individual's age, health, and vaccination status. The removed classification can indicate that the individual was sick, but is now recovered and immune, or that the patient died from the disease. Alternatively, in the case of a rapidly mutating virus, individuals could be allowed to transition back to susceptible after recovering from the infection. Infected individuals who die are removed from the model because their state is fixed and they no longer have contact with other members of the population.

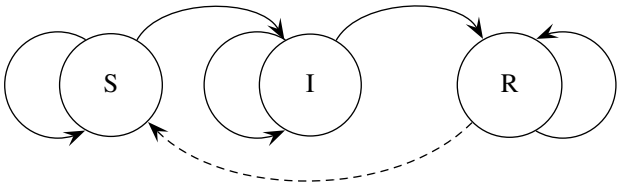

Figure 1: An individual's status as susceptible, infected, or recovered (removed) can be thought of as a Markov chain.

Like a Markov chain, in each unit of time, an individual has a certain probability of transitioning from one state to another. In real life, these probabilities are determined by a number of factors, for example, age, vaccination status, contact with infected individuals, and the rate of disease transmission per unit time of contact. These probabilities also change on a daily basis as the number of infected individuals with whom a person comes into contact changes.

To account for these changing and interrelated probabilities, an agent-based simulation model is used to individually monitor the state of each member of the population. Each individual is an object in the simulation with various characteristics, including age, vaccination status, home location, work location, and household membership. Household membership indicates which members of the population live in the same dwelling. In addition, once infected, each individual will be contagious for a randomly generated number of days, which is calculated as a function of age. Transportation routes used for daily commutes are also assigned to individuals.

Contact networks are used to establish contact leading to disease transmission between individuals. Each individual in the population has a certain level of contact with each other individual (the level of contact could be zero—no contact). The time and type of contact between two individuals can vary. The "quality" of contact refers to the intensity (closeness) of the contact. Therefore, if person i is infected, the probability that person i transmits the disease to person k is much higher than the probability that person i transmits the disease to person j.

Using the concept of contact networks, but without using actual networks, we replace a uniform reproduction number with individualized probabilities for each person j transitioning from a susceptible state (S) to an infected state (I) in period n:

$$\Pr_j^n(\mathrm{S},\mathrm{I}) = \sum_{i\in\mathcal{I}} t_{ij}^n b_{ij}^n + c_{ij}^n \quad \forall\, j \in \mathcal{S},$$

where $\mathcal{S} \subseteq \mathcal{N}$ is the set of susceptible individuals in population $\mathcal{N}$, $\mathcal{I} \subseteq \mathcal{N}$ is the set of infected individuals in population $\mathcal{N}$, t_{ij}^n is the time of contact between person i and person j in period n, and b_{ij}^n is the probability of disease transmission from person i to person j per unit time in period n. Values for t_{ij}^n and b_{ij}^n for

contact under social situations (not including contact while on public transportation) can be found in Haber et al. (2007) and Del Valle et al. (2007). The term c_{ij}^n accounts for transmission from person i to person j via indirect contact. For example, if infected person i uses a particular bus route, anyone else using that bus route that day will have an increased likelihood of infection because of the potential contact, given the ability of the virus to survive on inert objects for up to several hours.

Figure 2 shows flowcharts of the daily processes for two types of individuals. The flowcharts are simplified for clarity; the primary simplifications are the omission of probability of death and random community contact. In the model, an infected individual can die at any time according to the death rate specified by the disease. Random community contact also occurs at any time throughout the day.

Each individual has a flowchart; individuals whose daily activities cause them to be at the same place at the same time will have direct contact; individuals who are at the same place at different times will have indirect contact. Contact can lead to infection if one of the individuals is infected.

The generality of this formulation allows for nonhomogeneous modeling of the population; that is, all individuals are unique. Children and the elderly can be given higher probabilities than young adults for becoming infected. Health-care workers, because of the increased duration and degree of contact with infected individuals as the simulation progresses, will also have higher rates of transmission per unit time.

In addition to the nonhomogeneous nature of the model, the model also allows for any type of contact or changes in type of contact to be simulated by manipulating the t_{ij}^n and b_{ij}^n values. As behaviors of individuals change during the simulation (e.g., infected individuals voluntarily isolate themselves or are admitted to hospitals in period n), their corresponding t_{ij}^n and b_{ij}^n can be modified accordingly. If

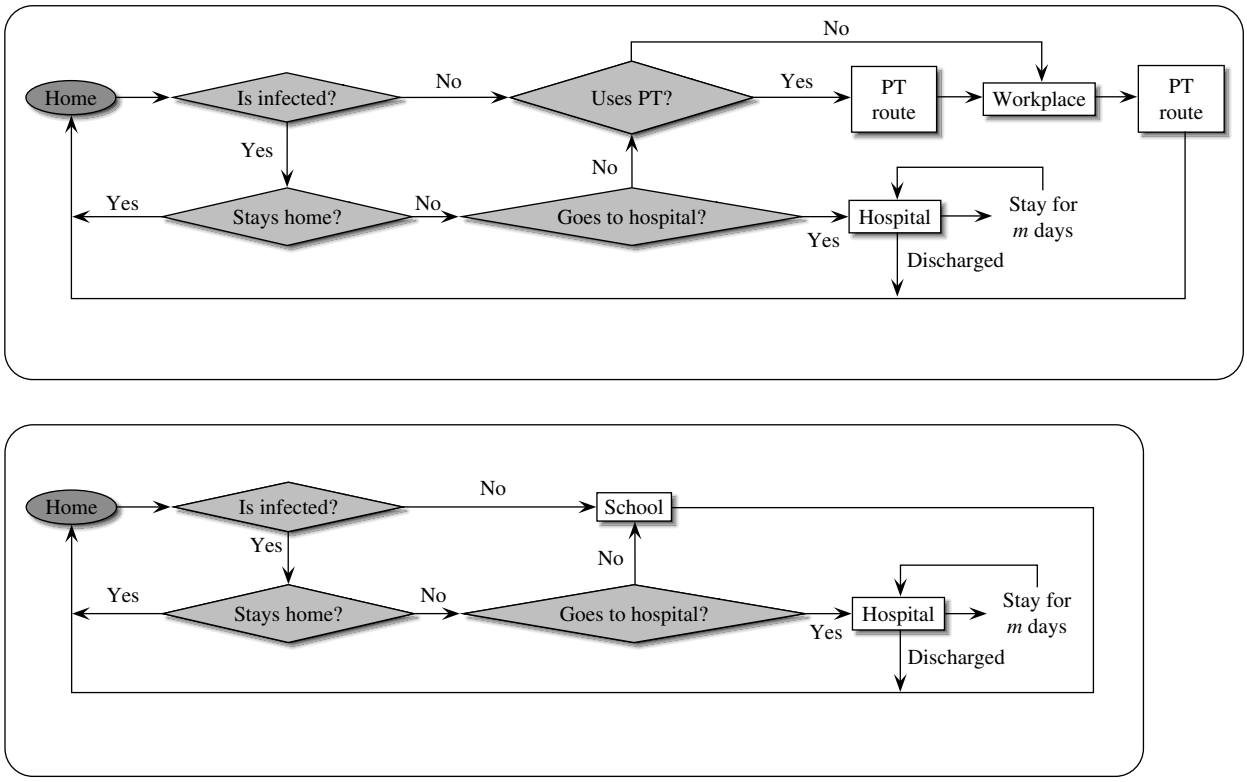

Figure 2: The top flowchart shows the process for an adult; the bottom flowchart shows the process for a school-age individual.

more social, demographic, or behavioral data become available, they can be easily incorporated into the model.

Factors Affecting Disease Transmission

Social and Demographic Aspects

Social and demographic aspects include both individual and population-wide characteristics. Individual characteristics include age, residence, and the level of general interaction with other members of the population. Although gender may play a role in susceptibility among individuals, this aspect has not been widely researched. Population-wide characteristics include population size, population density, age distribution, and distribution of household size.

Wallinga et al. (2006) analyzed the transmission parameters for respiratory-spread infectious agents using social-contact data according to age-specific groups. This research used conversations among the individuals as the mode of contacts because this mode can be used as a proxy measure for exposure to infectious respiratory-spread agents (Edmund et al. 1997). The results of this study showed that the probability of transmission increased in proportion to the frequency of conversation between two individuals; different age-group interactions yielded different transmission rates. Haber et al. (2007) included the length of contact as a significant factor in disease transmission and also showed that the disease transmission differs among age groups for the same length of time. Previous studies that addressed age-specific transmission include Ferguson et al. (2005), Longini et al. (2005), and Patel et al. (2005).

Transmission probability is significantly higher when a susceptible individual is in close contact with infectious individuals for a cumulative period (Edmund et al. 1997, Haber et al. 2007, Wallinga et al. 2006). This implies that family members of infectious individuals who are in close proximity and experience repeated exposure will have a higher risk of becoming infected. Data from the 2003 SARS outbreak, in which many family members of the initial cases were infected (Public Health Agency of Canada 2003, Poutanen et al. 2003), supported this. Therefore, the number of individuals in each household is an important factor in estimating an outbreak's size. A distribution of household sizes can be obtained from census reports and statistical data (Statistics Canada 2006a, b; Toronto City Planning, Research and Information 2006).

Behavioral Aspects

Examples of behavioral aspects include daily commutes and attendance at schools, workplaces, and hospitals. Most respiratory diseases have an asymptomatic period during which infected individuals are not yet aware of their illness, and therefore continue about their daily activities (Anderson and May 1991). During this time, the disease can be transmitted in their environments, including the workplace, school, or public transportation. The significance of disease transmission in public facilities is evidenced by research indicating that closure of such facilities could reduce the contact rate between infectious and susceptible individuals by up to 50 percent (Degli Atti et al. 2008, Haber et al. 2007, Halloran et al. 2008).

Individuals can be randomly assigned daily transportation routes based on building density reports for major commerce districts (e.g., downtown areas). Disease transmission is believed to occur only within a two-meter radius (Brankston et al. 2007). Thus, an infected individual has a significant probability of transmitting the disease only to persons traveling the same transit route and who are located within a two-meter radius of this infected individual. Detailed data regarding bus, streetcar, and subway passenger density and ridership can be obtained from the Toronto Transit Commission (TTC) (Toronto Transit Commission 2007).

Transmission occurring in hospitals, long-term care facilities, day-care facilities, and schools are also a major concern; however, homogeneous and nonhomogeneous mixing models cannot easily capture these environments.

Simulation Implementation and Parameters

Population

The simulation model was tested on a pandemic influenza outbreak in the GTA, which consists of 4.99 million people living in 1.8 million households

(Statistics Canada 2006a). The model randomly generates a population using probabilities obtained from this census data. Random population generation is used to ensure robustness of the model in the face of potential inaccuracies in the census data and population demographics that fluctuate from year to year. For lack of more information, the number of households is considered to be fixed at 1,801,250, and the population is generated by first randomly creating households of varying sizes. The number of households of each recorded size and the probabilities of households being of particular sizes as determined by Statistics Canada (2006a) are given in Table EC-1 in the electronic companion, which is part of the online version that can be found at http://interfaces.pubs.informs.org/ecompanion.html. Specific probabilities of a household being located in each of the 1,093 census tracts can also be found in the *Census Track Probabilities* table (Table EC-9) in the electronic companion.

After the number of households of each size has been determined, the total population size (p) is calculated, according to the household sizes, as $p = \sum_{i=1}^{M} ih_i$, where h_i is the number of households of size i and M is the maximum size of a household.

Once the population size has been determined, p individuals (agents) are generated. Each individual is sequentially assigned to a household according to the household sizes that were previously determined. At this time, each individual is also assigned an age. The census data were used to determine age distributions (see Table EC-2 in the electronic companion). Any other pertinent demographic or behavioral attributes, such as whether the individuals use public transportation (and the routes they use) or choose to stay home when sick, are also assigned.

For simplicity, the model assumes that the maximum age of individuals is 100 (and thus, the age range of "85 years and over" is treated as 85–100), and that the maximum household size is six individuals. Although different assumptions regarding maximum ages and household sizes could possibly affect the simulation outcomes, insufficient data are available for the GTA to estimate what these maximum values should be. Ultimately, the purpose of the model is to compare the potential efficacy of different mitigation strategies, rather than strictly predict the exact spread of disease; therefore, these assumptions are deemed insignificant.

Contact and Disease Transmission

Although the model is capable of addressing unlimited forms of contact between two individuals, this early version of the model considers only two types of contact: close and casual. Close contact exists between members of the same household and commuters on the same public transportation routes, because the subway lines are densely packed during rush hour. Casual contact occurs among random individuals living in the same census tract.

Pandemic influenza was chosen as the disease because of the availability of disease transmission data. Haber et al. (2007) provide data regarding transmission rates (see Table EC-3 in the electronic companion), contact times for contact between individuals in households (see Table EC-4 in the electronic companion), and randomly within the community (see Table EC-5 in the electronic companion). These values are also delineated according to interage and intra-age group transmission and contact, thereby allowing us to exploit the heterogeneous design of the model. The data in Haber et al. (2007) are for casual contact; our model assumes that the rate of transmission for close contact is slightly higher and is tuned based on expert opinion from the OAHPP and infectious control experts at local hospitals. Specific data from Haber et al. (2007) is given in the electronic companion.

To address public transportation, the simplifying assumption is that every individual of working age (18–65) used the nearest subway line to travel to the main downtown hub station (Union Station) twice a day (to and from work). Public transportation data were obtained from the TTC, and only subway routes during rush hours were considered in the model. The travel time of each subway car, passenger density, and total ridership for both rush hour and nonrush hour usage are available at the TTC website (http://www.ttc.ca).

Initial Outbreak

Initially, all individuals in the population are marked as being in the "susceptible" state. The outbreak is initialized by setting the state of a small number of

individuals to "infected." In all simulations, the initial number of infected individuals is constant to facilitate comparisons between the different mitigation strategies tested. The number of initial infected persons is set to seven because the number of initial cases in the SARS outbreak in Toronto was believed to be seven (Public Health Agency of Canada 2003).

Mitigation Strategy Assessment

The model's capacity to assess the relative effectiveness of mitigation strategies is important. As an early example, we tested the effectiveness of a social-distancing strategy (Glass et al. 2006)—an advertising campaign that urges individuals who show symptoms of the disease to stay home until they are well, rather than force themselves to go to work or school; infected individuals would continue to have contact with household members. Clearly, such a campaign would not discourage people with severe symptoms from seeking medical attention; it would also not create a panic situation in which hospitals would be overrun with uninfected individuals displaying minor symptoms, for example, a runny nose. Note that this campaign is not a voluntary quarantine; in a quarantine, the targeted individual is confined to the house and must not have any contact with household members; in addition, other household members must use personal protective equipment as necessary. Such a measure is generally applied only to healthcare responders with known exposure to the disease, likely because of an inability to force members of the general population to stay home.

If we arbitrarily assume that the campaign will successfully convince x percent of the population to stay home when they become ill, how will the spread of disease be affected? Will the difference in disease spread be significant enough to warrant the cost of the campaign?

We present tests of social-distancing effectiveness levels from 0 to 100 percent in 10 percent increments, for 11 possible realizations of campaign effectiveness. If the individual is assigned to stay home, then all casual contact with the community for that individual ceases, as does any use of public transportation. Future versions of the model that incorporate work/school environments will also prevent the individual from having contact with others in those environments. During the infectious period, the individual still has contact with household members, as is common when someone in a household is sick with, for example, the common cold.

Common random numbers are used to seed the random number string in each simulation across each scenario to ensure that we can later compare the simulations fairly. For example, the first simulation in each scenario is seeded with number a_1, the second simulation in each scenario is seeded with number a_2, and so on. The random number seeds are generated using a Mersenne twister algorithm (Matsumoto and Nishimura 1998).

User Interface

At present, our model does not have a graphical user interface. Model input, which the electronic companion to this paper shows, is contained in tab-delimited text files, which are easily created by exporting Microsoft Excel spreadsheets containing the data. Some frequently edited data (e.g., disease transmission rates and population behavior) are in separate text files, which describe the specific scenario to be modeled.

The model output is simply a series of text files, one for each simulation, containing all the desired measurements. These measurements include, but are not limited to, daily numbers of individuals in the S, I, and R states (as defined in the *Simulation Model* section), number of deceased individuals, and number of infected individuals per census tract. These text files are parsed by a simple *MathWorks MATLAB* function that creates graphical histograms and cumulative probabilities of outcomes, and calculates basic statistical information (e.g., mean, standard deviation, and confidence intervals).

Results

The specific question posed was whether an agent-based simulation modeling an outbreak in a large urban center could be designed to run quickly enough to be applicable to a real-life pandemic planning situation. Major concerns for using agent-based simulation are the computer resources and the time required to complete the simulation, especially when simulating a population of almost five million individuals. To make the model as efficient as reasonably possible, we

wrote it using C++. The output of the model is sent to GIS software to display maps of outbreak severity to provide visual tools to assist in policy-making decisions.

The simulations were run on a 32-node Beowulf cluster with 256 CPUs. The parallelization was done using MPICH libraries. Although each individual simulation required approximately 2.98 minutes on the cluster because of overhead communication and slow file input/output time, all 1,000 simulations were completed in less than 13 minutes.

The histogram of the 1,000 model outcomes in Figure 3 confirms our expectations that the outbreak is less severe when significant numbers of infected individuals stay home. The top graph in Figure 3 illustrates histogram results (i.e., the number of instances in the 1,000 simulations that resulted in specific ranges of infected and deceased individuals). The histograms are shown as line plots rather than as bars because of the large number of implementations tested. To facilitate comparison between the scenarios, the bottom graph in Figure 3 illustrates the cumulative probabilities of outcomes. From these cumulative probabilities, the significant decrease in levels of infected and deceased persons as increasing numbers of infected individuals stay home is evident. This result agrees with the social-distancing outcomes presented by Germann et al. (2006), Ferguson et al. (2005), and Wu et al. (2006). Similar results were seen for deaths.

Table 1 contains statistical results of the total number of infected individuals and deceased individuals, respectively. When 0–70 percent and 100 percent of individuals stay home, the mean, median, and mode are all close, indicating stable outcomes in the simulations. When 80–90 percent of infected individuals stay home, the majority of simulations ended with relatively low numbers of infections; however, in some simulations, the rate of infection in the population was significant. For example, in the 90 percent scenario, 8 percent of simulations (80 of 1,000) resulted in more than 9,000 infections, and 2 percent of simulations (20 of 1,000) resulted in more than 10,000 infections. Similar results were seen for deaths.

We also performed statistical comparisons between each mitigation scenario; we address questions such as, "Are we 90 percent confident that 20 percent of infected individuals staying home is better than 10 percent staying home?" To answer such questions from a statistical significance perspective, we examine the difference in all 1,000 outcomes from scenario i to scenario j:

$$\mathbf{x} = \text{infections}_i - \text{infections}_j.$$

If scenario j is routinely better than scenario i (i.e., fewer infections occur in scenario j), then the difference in outcomes $\mathbf{x}$ will have mostly positive elements. We can then use confidence intervals to determine how confident we are that scenario j is better than scenario i. For a desired confidence level of $(1-\alpha)$, the confidence interval is defined by

$$\text{CI} = \hat{\mu}_x \pm z_{\alpha/2}\hat{\sigma}_x,$$

where $\hat{\mu}_x$ is the mean of $\mathbf{x}$ and $\hat{\sigma}_x$ is the standard deviation of $\mathbf{x}$. The value $z_{\alpha/2}$ follows from the normal distribution function. If we calculate a 90 percent confidence interval for the increase in infections that occur in scenario i compared to scenario j, 90 percent of the differences in infections will be contained in the confidence interval. Therefore, if the lower bound of the confidence interval is positive, then scenario j is better than scenario i 90 percent of the time.

Table 2 shows the 90 percent confidence intervals in performing pairwise comparisons between scenarios i (rows) and scenarios j (columns). The table shows that most of the confidence intervals are positive, indicating that the scenarios in the columns (i.e., more people stay home) are usually better than the scenarios in the rows (i.e., fewer people stay home). The confidence intervals with negative lower bounds occur near the diagonal of the table, where there is only a 10–20 percent difference in the number of individuals staying home. For those comparisons, it is not clear that scenario j is convincingly better than scenario i, although only a small part of those confidence intervals contain negative values.

If the attack rate (i.e., the number of individuals infected) of the disease falls within a confidence interval, this information can be used to help select the best policy.

To assess how confident we are that scenario j is better than scenario i, we test each pairwise comparison to determine the α value that yields a positive

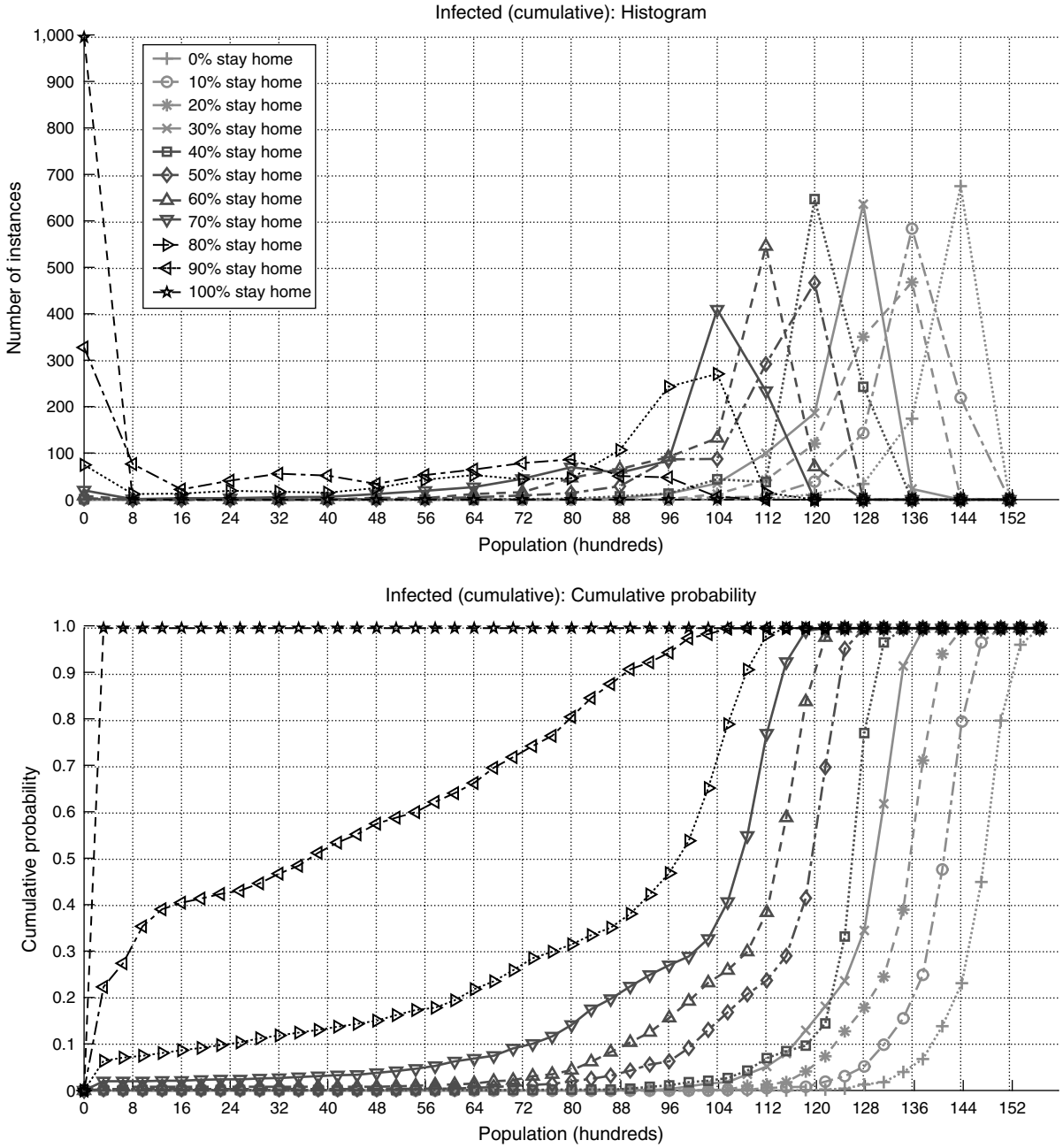

Figure 3: The cumulative infected populations at the conclusion of the modeled period for various probabilities of infected persons staying home generally show statistically significant differences. Top: Histogram of cumulative number of infected persons. Bottom: Cumulative probabilities.

lower bound in the confidence interval. Table 3 shows the resulting confidence levels that scenario j is better than scenario i. The table also shows that almost 70 percent of comparisons indicate that scenarios in which more individuals stay home are better with at least 90 percent confidence, almost 60 percent are better with at least 95 percent confidence, and 34 percent are better with the maximum 99 percent confidence.

	Percent stay home (%)										
	0	10	20	30	40	50	60	70	80	90	100
Average	14,670	13,978	13,375	12,768	12,459	11,521	10,844	9,903	8,235	4,027	21
Median	14,799	14,124	13,575	13,012	12,628	11,962	11,404	10,795	9,779	3,654	21
Mode	14,718	14,135	13,407	12,955	12,572	12,061	11,686	11,111	17	19	18
St. dev.	538	604	675	758	776	1,252	1,576	2,229	3,254	3,515	5
Minimum	11,767	10,562	9,992	8,805	18	12	12	13	13	12	11
Maximum	15,693	15,008	14,570	14,023	13,480	12,766	12,438	12,003	11,425	10,604	41

Table 1: Statistical results indicate a steady decrease in infections as infected individuals choose to stay home.

Along the diagonal, we clearly cannot be very confident that a 10 percent increase in the number of individuals who stay home will improve the outbreak. This appears to contradict the cumulative probability results illustrated in Figure 3; however, the cumulative probabilities between similar scenarios are very close; in some cases, they overlap for low numbers of infections. Thus, instances in which scenario i is better than scenario j occur almost entirely for low numbers of infections in both scenarios; in those instances, both scenarios are desirable, although scenario i is slightly better.

We imported the simulations outputs into ESRI ArcGIS software to obtain shaded maps of the GTA; more shading represents a higher percentage of population infected in that census tract. Figure 4 shows the GIS maps of trials that obtained the average number of cumulative infected individuals at the conclusion of the 60-day simulation. If only 50 percent of the infected population stays home, there is a significant impact on the disease spread.

Although Toronto recently experienced a pandemic SARS outbreak, like most pandemic models, the outcomes for a hypothetical pandemic influenza outbreak cannot be truly validated. The 2003 SARS outbreak occurred prior to the existence of the OAHPP; indeed, this public health agency was established in Ontario largely as a result of the SARS Toronto experience. At the outset of the SARS outbreak, public health authorities enacted measures that included quarantine, infection control, and personal protective equipment. However, the mode of spread of the SARS virus was different from that of an influenza pandemic: transmission occurred primarily in health care and home settings rather than in the community. Moreover, quarantine is not thought to be useful in managing influenza virus spread (Cetron and Ladwiorth 2005, Heymann 2004). Hence, the SARS experience would not serve as an appropriate validation tool for the model described in this paper. Therefore, this model is better used as a "what-if" tool

	0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
0%	—	(45, 1,263)	(580, 1,960)	(884, 2,847)	(1,299, 3,623)	(1,442, 4,792)	(1,538, 6,149)	(1,317, 8,505)	(1,700, 11,326)	(4,957, 15,999)	(13,596, 15,484)
10%	—	—	(−5, 1,238)	(307, 2,116)	(719, 2,895)	(891, 4,035)	(992, 5,387)	(732, 7,783)	(1,119, 10,599)	(4,340, 15,307)	(12,877, 14,895)
20%	—	—	—	(−222, 1,412)	(216, 2,166)	(309, 3,383)	(443, 4,703)	(188, 7,094)	(570, 9,915)	(3,784, 14,630)	(12,164, 14,375)
30%	—	—	—	—	(−227, 1,418)	(−72, 2,575)	(31, 3,925)	(−283, 6,375)	(73, 9,221)	(3,199, 14,025)	(11,279, 14,069)
40%	—	—	—	—	—	(−582, 1,893)	(−501, 3,266)	(−829, 5,729)	(−463, 8,566)	(2,640, 13,393)	(10,458, 13,699)
50%	—	—	—	—	—	—	(−726, 2,179)	(−1,182, 4,771)	(−901, 7,693)	(2,034, 12,687)	(9,275, 13,571)
60%	—	—	—	—	—	—	—	(−1,564, 3,699)	(−1,327, 6,666)	(1,390, 11,878)	(7,891, 13,502)
70%	—	—	—	—	—	—	—	—	(−1,695, 4,898)	(272, 10,860)	(5,542, 13,716)
80%	—	—	—	—	—	—	—	—	—	(−1,247, 9,176)	(2,712, 13,342)
90%	—	—	—	—	—	—	—	—	—	—	(−1,727, 9,852)
100%	—	—	—	—	—	—	—	—	—	—	—

Table 2: The confidence intervals for pairwise comparisons generally indicate a statistical significance at $\alpha = 10$ percent (90 percent confidence level) when there is a 20 percent difference in the percentage of infected individuals staying home.

Figure 4: GIS maps of every 12th day of representative simulations, starting on the fifth day, show the percentage of population in a census tract that is infected. Darker colors indicate higher percentages of infection. Left: Infected population members continue about their daily business. Right: Fifty percent of infected population members choose to stay home.

	0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
0%	—	92	99	99	99	99	99	97	97	99	99
10%	—	—	89	97	99	99	98	95	95	99	99
20%	—	—	—	76	95	95	95	91	93	99	99
30%	—	—	—	—	76	88	90	86	90	99	99
40%	—	—	—	—	—	61	77	78	86	98	99
50%	—	—	—	—	—	—	58	67	80	97	99
60%	—	—	—	—	—	—	—	49	72	96	99
70%	—	—	—	—	—	—	—	—	57	91	99
80%	—	—	—	—	—	—	—	—	—	78	98
90%	—	—	—	—	—	—	—	—	—	—	75
100%	—	—	—	—	—	—	—	—	—	—	—

Table 3: The confidence levels for pairwise comparisons generally show strong confidence for at least a 20 percent difference in the percentage of infected individuals staying home, but only moderate confidence for a 10–20 percent difference.

to compare relative outcomes of different scenarios and mitigation strategies.

Conclusions and Future Directions

Our model provides further evidence that agent-based simulation is both a computationally viable tool for predicting disease spread during a pandemic in very large populations and a method for testing the relative effectiveness of mitigation strategies. The model can also enhance the accuracy of traditional homogeneous outbreak models by considering factors, including age, household and family interaction, casual interaction, geographic location, and public transportation use.

From the perspective of government planning groups, including the OAHPP, our model considers public transportation use in urban cities; therefore, it addresses an important factor for disease transmission not considered in previous models, such as Das et al. (2008) and Stroud et al. (2007). Its development allows planners to consider and test such interventions as social distancing and antiviral prophylaxis and treatment, and contributed to planning and policy decisions.

More recently, our model served as a tool for the OAHPP emergency planners to assess the effects of interventions in real time, in isolation, or in combination to lend plausibility to these strategies. Although, like most predictions of future pandemic disease spread, the model cannot be reliably validated, it can be used as a "what-if" vehicle to make comparisons among different scenarios. These scenarios can include various disease parameters and interventions (most recently, different vaccination strategies) to assist scientists in forecasting the relative effects of recommendations and interventions, enabling further planning even during a pandemic event. It has enabled estimation of relative levels of outcomes based on compliance rates; moreover, it can assist in determining the cost effectiveness of large media campaigns or vaccine delivery systems.

Despite this model's unique capabilities and its real-world use in informing the scientific response to the pandemic in Ontario, it is important to remember that simulation models are not crystal balls. The purpose of these models is to help policy makers prepare for outbreaks. Although traditional homogeneous and nonhomogeneous mixing models can estimate metrics, such as total number infected, they are limited in their ability to test the effectiveness of mitigation strategies, which is ultimately the most important feature of a model from a real-world perspective.

Future directions will address the effects of disease spread on hospitals, health-care workers (including general practitioners), and kindergarten through grade 12 school environments. More advanced assignments of public transportation use will also be developed. We are currently performing extensive sensitivity analysis, with particular emphasis on rates of disease transmission and required numbers of simulations to achieve a stable distribution of outcomes. We also plan to expand the model to test more sophisticated mitigation strategies (e.g., vaccine drives, school closures, and quarantines).

Electronic Companion

An electronic companion to this paper is available as part of the online version that can be found at http://interfaces.pubs.informs.org/ecompanion.html.

References

Anderson, R. M., R. M. May. 1991. *Infectious Diseases of Humans: Dynamics and Control*. Oxford University Press, New York.

Bean, B., B. M. Moore, B. Sterner, L. R. Peterson, D. N. Gerding, H. H. Balfour. 1982. Survival of influenza viruses on environmental surfaces. *J. Infectious Diseases* **146**(1) 47–51.

Brankston, G., L. Gitterman, Z. Hirji, C. Lemieux, M. Gardam. 2007. Transmission of influenza A in human beings. *Lancet Infectious Disease* **7**(4) 257–265.

Burr, T. L., G. Chowell. 2008. Signatures of non-homogeneous mixing in disease outbreaks. *Math. Comput. Model.* **48**(1–2) 122–140.

Cetron, M., J. Ladwiorth. 2005. Public health and ethical considerations in planning for quarantine. *Yale J. Biol. Medicine* **78**(5) 325–330.

Cliff, S., P. Haggett, J. K. Ord. 1986. *Spatial Aspects of Influenza Pandemics*. The Pion Limited, London.

Das, T. K., A. A. Savachkin, Y. Zhu. 2008. A large scale simulation model of pandemic influenza outbreaks for development of dynamic mitigation strategies. *IIE Trans.* **40**(9) 893–905.

Degli Atti, M. L. C., S. Merler, C. Rizzo, M. Ajelli, M. Massari, P. Manfredi, C. Furlanello, G. S. Tomba, M. Iannelli. 2008. Mitigation measures for pandemic influenza in Italy: An individual base model considering different scenarios. *PLoS ONE* **3**(3) e1790.

Del Valle, S. Y., J. M. Hyman, H. W. Hethcote, S. G. Eubank. 2007. Mixing patterns between age groups in social networks. *Soc. Networks* **29**(4) 539–554.

Edmund, W. J., C. J. O'Callaghan, D. J. Nokes. 1997. Who mixes with whom? A method to determine the contact pattern of adults that may lead to the spread of airborne disease. *Proc. Biol. Sci.* **264**(1384) 949–957.

Eubank, S. 2005. Network based models of infectious disease spread. *Japanese J. Infectious Diseases* **58**(6) S9–S13.

Eubank, S., G. Hasan, A. Kumar, M. V. Marathe, A. Srinivasan, Z. Toroszkai, N. Wang. 2004. Modelling disease outbreaks in realistic urban social networks. *Nature* **7**(6988) 180–184.

Ferguson, N. M., D. A. T. Cummings, C. Fraser, J. C. Cajka, P. C. Cooley, D. S. Burke. 2006. Strategies for mitigating an influenza pandemic. *Nature* **442**(7101) 448–452.

Ferguson, N. M., D. A. T. Cummings, S. Cauchemez, C. Fraser, S. Riley, A. Meeyai, S. Iamsirithaworn, D. S. Burke. 2005. Strategies for containing an emerging influenza pandemic in Southeast Asia. *Nature* **437**(7056) 209–214.

Germann, T. C., K. Kadau, I. M. Longini, C. A. Macken. 2006. Mitigation strategies for pandemic influenza in the United States. *Proc. Natl. Acad. Sci. USA* **103**(15) 5935–5940.

Glass, R. J., L. M. Glass, W. E. Beyeler, H. J. Min. 2006. Targeted social distancing design for pandemic influenza. *Emerging Infectious Diseases* **12**(11) 1671–1681.

Haber, M. J., D. K. Shay, X. M. Davis, R. Patel, X. Jin, E. Weintraub, E. Orenstein, W. W. Thompson. 2007. Effectiveness of interventions to reduce contact rates during a simulated influenza pandemic. *Emerging Infectious Disease* **13**(4) 581–589.

Halloran, E. M., N. M. Ferguson, S. Eubank, I. M. Longini, D. A. T. Cummings, B. Lewis, S. Xu, et al. 2008. Modeling targeted layered containment of an influenza pandemic in the United States. *Proc. Natl. Acad. Sci. USA* **105**(12) 4639–4644.

Heymann, D. L., ed. 2004. *Control of Communicable Diseases Manual*, 18th ed. American Public Health Association, Washington, DC.

Lee, B. Y., V. L. Bedford, M. S. Roberts, K. M. Carley. 2008. Virtual epidemic in a virtual city: Simulating the spread of influenza in a US metropolitan area. *Translational Res.* **151**(6) 275–287.

Lipsitch, M., T. Cohen, B. Cooper, J. M. Robins, S. Ma, L. James, G. Gopalakrishna, et al. 2003. Transmission dynamics and control of severe acute respiratory syndrome. *Science* **300**(5627) 1966–1970.

Longini, I. M., M. E. Halloran, A. Nizam, Y. Yang. 2004. Containing pandemic influenza with antiviral agents. *Amer. J. Epidemiology* **159**(7) 623–633.

Longini, I. M., A. Nizam, S. Xu, K. Ungchusak, W. Hanshaoworakul, D. A. T. Cummings, M. E. Halloran. 2005. Containing pandemic influenza at the source. *Science* **309**(5737) 1083–1087.

Matsumoto, M., T. Nishimura. 1998. Mersenne twister: A 623-dimensionally equidistributed uniform pseudo-random number generator. *ACM Trans. Model. Comput. Simulation* **8**(1) 3–30.

Meyers, L. A., B. Pourbohloul, M. E. J. Newman, D. M. Skowronski, R. C. Brunham. 2005. Network theory and SARS: Predicting outbreak diversity. *J. Theoret. Biol.* **232**(1) 71–81.

Newman, M. E. J. 2002. Spread of epidemic disease on networks. *Phys. Rev. E* **66**(1) 016128.

Patel, R., I. M. Longini, M. E. Halloran. 2005. Finding optimal vaccination strategies for pandemic influenza using genetic algorithms. *J. Theoret. Biol.* **234**(2) 201–212.

Potter, C. W. 1998. Chronicle of influenza pandemics. K. G. Nicholson, R. G. Webster, A. J. Hay, eds. *Textbook of Influenza*. Blackwell Science, Oxford, UK.

Poutanen, S. M., D. E. Low, B. Henry, S. Finkelstein, D. Rose, K. Green, R. Tellier, et al. 2003. Identification of severe acute respiratory syndrome in Canada. *New England J. Medicine* **348**(20) 1995–2005.

Public Health Agency of Canada. 2003. Severe acute respiratory syndrome (SARS): Preliminary epidemiological findings as of April 4, 2003. Accessed November 17, 2010, http://www.phac-aspc.gc.ca/sars-sras/pef-dep/sars-pef20030404-eng.php.

Shi, P., P. Keskinocak, J. L. Swann, B. Y. Lee. 2010. Modelling seasonality and viral mutation to predict the course of an influenza pandemic. *Epidemiology Infection* **138**(10) 1472–1481.

Statistics Canada. 2006a. Census 2006. Accessed November 17, 2010, http://www12.statcan.ca/census-recensement/2006/rt-td/index-eng.cfm.

Statistics Canada. 2006b. Profile of marital status, common-law status, families, dwellings and households for census metropolitan areas and census agglomerations, 2006 census. Accessed November 17, 2010, http://www12.statcan.ca/census-recensement/2006/rt-td/index-eng.cfm.

Stroud, P., S. Del Valle, S. Sydoriak, J. Riese, S. Mniszewski. 2007. Spatial dynamics of pandemic influenza in a massive artificial society. *J. Artificial Soc. Social Simulation* **10**, http://ideas.repec.org/a/jas/jasssj/2007-34-2.html.

Toronto City Planning, Research and Information. 2006. Living downtown survey. Accessed November 17, 2010, http://www.toronto.ca/planning/pdf/living_downtown_nov1.pdf.

Toronto Transit Commission. 2007. TTC operating statistics 2007. Accessed November 17, 2010, http://www3.ttc.ca/About_the_TTC/Operating_statistics.jsp.

Wallinga, J., P. Theunis, M. Kretzschmar. 2006. Using data on social contacts to estimate age-specific transmission parameter for respiratory-spread infectious agents. *Amer. J. Epidemiology* **164**(10) 936–944.

Wu, J. T., S. Riley, C. Fraser, G. M. Leung. 2006. Reducing the impact of the next influenza pandemic using household-based public health interventions. *PLoS Medicine* **3**(9) 1532–1540.

Brian Schwartz, MD, Director of Emergency Management Support, Ontario Agency for Health Protection and Promotion, writes: "This is to confirm that the work presented in the [above noted] paper, has and will continue to be utilized in the planning for and management of pandemics in the Province of Ontario.

"As Chair of the Ontario Health Plan for an Influenza Pandemic (OHPIP) Primary Care Working Group and the Rapid Access to Anti-Viral Medication Working Group, I am actively employing these models in our pandemic planning.

"The models used in this paper were requested after discussion at the OHPIP Steering Committee regarding potential effectiveness of public health measures. These discussions led to a collaboration between myself as Ministry of Health lead and Professor Aleman's group. The first step in this process was the development of the disease spread model outlined in the submitted manuscript. This model is more complex and realistic than those developed previously, and will be used in assessment of impact of public health measures such as school closure, antiviral treatment and prophylaxis and social distancing. This process has already resulted in a reassessment of the "treatment only" strategy for antivirals (to model the effect of targeted prophylaxis on the new model) and collaboration with other modellers in the cost effectiveness/risk-benefit of school closures."