ACCOUNTING FOR INDIVIDUAL BEHAVIORS IN A PANDEMIC DISEASE SPREAD MODEL

Dionne M. Aleman
Theodorus G. Wibisono
Department of Mechanical and Industrial Engineering
University of Toronto
5 King’s College Road
Toronto, ON M5S 3G8, Canada

Brian Schwartz
Department of Family and Community Medicine
University of Toronto
263 McCaul Street
Toronto, ON M5T 1W7, Canada

ABSTRACT

Mathematical models to predict the spread of disease during a pandemic largely require overly simplistic assumptions about disease transmission within populations. One significant shortcoming of these models is the inability to account for varying types and amount of contact between individuals, to address individuals’ behaviors or to assess the effectiveness of mitigation strategies. We present a non-homogeneous agent-based simulation of a pandemic in an urban population that accounts for individual behavior and transmission rates in different scenarios. The model is compact and parallelizable, and runs in reasonable computational time for an urban population of nearly five million individuals. Results are presented from modeling the spread of pandemic influenza in the Greater Toronto Area, Ontario, Canada.

1 INTRODUCTION

In order to most effectively prepare for a pandemic disease outbreak, knowledge of how the disease will spread is paramount. This need for disease spread data was highlighted during the global outbreak of Severe Acute Respiratory Syndrome (SARS) in 2002-2003, but is apparent in preparation for and response to all disease outbreaks, from pandemic influenza to avian flu.

The SARS outbreak resulted in 8,096 infected individuals and 774 deaths, highlighting the importance of comprehensive mitigation strategies to protect the population in the event of a pandemic outbreak. Mitigation strategies include decisions regarding the allocation of finite prophylaxis resources, increased load on hospitals and when enact school closures and (voluntary) quarantines. The strategy should be designed to effectively achieve the goals of the pandemic response, whether it is minimize deaths, minimize infections, minimize geographic spread or minimize the duration of the outbreak.

In order to achieve these goals, decision makers must determine how to prioritize antiviral prophylaxis treatment for members of the population. Depending on the specific goals, the prioritization will be different. A goal of minimizing deaths, for example, may indicate that the elderly and very young should receive antiviral treatment. A goal of minimizing the number of people infected may indicate that healthcare workers and elementary school-aged children be treated with antivirals. A detailed model of the disease spread can assess the ability of various antiviral distribution strategies to meet the pandemic response goals. A realistic disease spread model will also indicate when and if school closures and quarantines are most effective.

Models to predict the spread of a disease have largely relied simplifying measures to make the mathematical computations of disease transmission tractable. The most commonly used model, the homogeneous mixing model, makes three main assumptions: 1) the population is fully mixed, meaning that individuals with whom a susceptible individual has contacts are chosen at random from the whole population; 2) all individuals have approximately the same number of contacts in the same period of time; and 3) all contacts transmit the disease with the same probability. In other words, all infected individuals are assumed to transmit the disease to the same number of people (that number is called the basic reproduction number, $R_0$). Those people becoming infected are randomly chosen from the entire population. The resulting predictions are “grossly inaccurate in at least some cases” (Newman 2002), 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). Additionally, as evidenced by recent experiences with swine flu, $R_0$ is not easy to calculate. Moreover, if the swine
flu pandemic had started in New York City instead of small village in Mexico (the current belief at the time of this writing), \( R_0 \) would be very different due to environmental, demographic and behavioral factors, despite being the same disease.

A more realistic model gaining traction recently is the non-homogeneous (heterogeneous) mixing model (Larson 2007, Meyers et al. 2005, Valle et al. 2007). This type of model explains the probability of contacts between any two members in a population. These probabilities are represented in a contact network, a simplified version of a social network. Based on the contact of each pair of individuals in a population, contact networks allow for the reproduction number to be individualized, that is, each infected individual can infect a unique number of other individuals based on the degree of his/her node in the contact network.

Although non-homogeneous mixing models provide increased realism, they generally require complicated mathematics in order to obtain useful information regarding the disease spread, e.g., number of infections/deaths per day, total number of infected/dead, etc. These calculations are often difficult to perform, and several metrics that are obtained do not provide conclusive information regarding the spread of the disease. It is also difficult to evaluate how mitigation strategies and individual behaviors affect the disease spread.

An alternative, we propose a simulation model to emulate the behaviors and contacts of individual members of the population. Certain behaviors, for example, using public transportation or staying home when sick, impact an individual’s probability of becoming infected or probability of infecting another person. By using a simulation model, we can address the probabilistic nature of disease transmission. Rather than deterministically specify a number of individuals that an infectious person will infect, we can calculate each person’s probability of becoming infected based on his/her contacts within the population each day. This method is better suited to accept the available probabilistic disease transmission information as input as opposed to relying on the anecdotal reproduction number.

Our non-homogeneous simulation will be agent-based to allow for each member of a population to be simulated individually, thereby providing a means to consider unique individual characteristics that affect transmission and infection probabilities. An agent-based model can also easily account for the possibility of infected individuals recovering and becoming immune or possibly dying. The model is tested on the Greater Toronto Area (GTA), Ontario, Canada. The factors considered are social and family contact, public transportation use and resulting contact and the fact that some individuals will choose to stay home (and not infect anyone else) when they become infected.

\section{AGENT-BASED SIMULATION MODEL}

As previously said, the agents in the model are the individuals in the population. The S-I-R (Susceptible, Infectious, Removed) model is commonly used to describe the population is disease spread models. In the proposed model, each individual is classified as susceptible, infectious or removed, where susceptible indicates healthy and removed indicates either recovered (formerly sick, but now immune) or dead as a result of the disease. If the disease being modeled is rapidly mutating, it is possible for individuals who have have recovered from being infected to become infected again.

A snapshot of the population is taken at each time period. Each time period, each individual has a certain probability of transitioning from one state (susceptible, infectious or removed) to another. In other words, each individual has a unique Markov chain associated with their classification (Figure 1).

\subsection{Transition probabilities}

The probability of transitioning from susceptible to infected is determined by the amount of contact an individual has with other infected individuals that day. Contact can occur in any setting, and the length of contact is dependent on the setting. The rate of disease transmission between two individuals is dependent on the type of contact between them: casual contact,
close contact, healthcare worker/patient contact, etc. Disease may also be transmitted between two persons indirectly, for example, through bacteria left on a seat on a subway car.

Formally, the probability of person \( j \) transitioning from a susceptible state (S) to an infected state (I) in time period \( n \) is given by

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

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 time period \( n \); and \( b_{ij}^n \) is the probability of disease transmission from person \( i \) to person \( j \) per unit time in time period \( n \). Indirect contact between two individuals is controlled by the value \( c_{ij}^n \), which is independent of the amount of time the two individuals are in the same place. In fact, infected individual \( i \) and healthy individual \( j \) may not ever be at the same place at the same time; but if person \( i \) was on a bus route that person \( j \) later takes, person \( i \)'s bacteria could still be transmitted to person \( j \) via contact with grab bars, chairs, etc.

The “quality” or “intensity” of the contact is controlled by \( b_{ij}^n \), which can be smaller for more formal forms contact, or larger for more intimate forms of contact. The specific values of \( b_{ij}^n \) are determined by the type of link between persons \( i \) and \( j \) during time period \( n \) in the contact network. Traditionally, social networks only specify “close” and “casual” levels of contact; however, forms of contact do not need to be limited to only two categories. Different levels of contact can be created for any conceivable relationship between two individuals, thereby allowing significant flexibility in modeling daily interactions.

An individual \( j \) who becomes infected stay infected for \( m_j \) days, randomly generated according to their age. If the observed death rate of a disease is \( d^* \), each day of infection, a person \( j \) has a certain probability of death:

\[
Pr_j^d(I,D) = 1 - (1 - d^*)^{1/m_j}
\]

where \( D \) indicates a state representing death. This value is determined as the necessary daily death probability to yield an overall chance of death equal to \( d^* \). Thus, the probability of an individual who is currently infected staying infected in the next time period is given by \( Pr_j^d(I,I) = 1 - Pr_j^d(I,D) \) if person \( j \) has been infected for less than \( m_j \) days.

For transitions to the recovered state, two possibilities relating to the specific disease can be tested. For many diseases, it is expected that once a person has and recovers from the disease, the person is immune from contracting the disease again. If such is the case for the disease being modeled, then once person \( j \) has been infected for \( m_j \) days, he/she will transition to state R with probability 1.

If the disease rapidly mutates and recovered individuals cannot rely on immunity, then once person \( j \) has been infected for \( m_j \) days, he/she will transition back to state S with probability 1. In this situation, state R will never be reached by living individuals. Individuals who have been previously infected may exhibit stronger immune responses toward the disease in the form of being sick for a shorter period of time (if they are re-infected) or being less susceptible to becoming sick again in the first place. This can be accounted for by multiplying \( Pr_j^d(S,I) \) by, say, 0.5 to make previously infected individuals 50% less likely to become re-infected. A similar adjustment can be made to \( m_j \).

A rapidly mutating disease may also display changing transmission rates as the virus changing. If there is some information, even anecdotal, about how the transmission rates are changing, this can be easily incorporated into the model by changing the transmission rates \( b_{ij}^n \) for different values of \( n \). For example, if during the course of an outbreak, it is observed that the disease is transmitting faster and faster, policy makers could tell the model to multiply \( b_{ij}^n \) values by 1.05 (a 5% increase) every seven days.

Although it is not an exact science, this flexibility can help policy makers get a general idea of how their mitigation strategies should adjust in the face of different outbreak scenarios.

### 2.2 Behavioral and demographic considerations

This formulation of transition probabilities allows for flexibility in determining contact types, times and transmission rates, and allows for those values to be modified each day. Any type of contact deemed significant or common enough can be incorporated. Changes in daily behavior, e.g., traveling to work on weekdays and staying home on weekends, seeking medical treatment when infected, and more, can be addressed and analyzed. Children and the elderly can be given higher probabilities of becoming infected compared to young adults. Healthcare workers, by virtue of the increased duration and quality of contact with infected individuals as the simulation progresses, will also have higher probabilities of becoming infected.
An important demographic factor that our model considers is the effect of disease on different age groups. It has been shown that the probability of transmission increases proportionally to the frequency of conversation between two individuals, with different age group interactions yielding different transmission rates (Wallinga et al. 2006). The length of contact is also a considerable factor of disease transmission, and the transmissibility of the disease differs among the age-group for the same length of time (Haber et al. 2007). Infected persons will also remain infected for a period of time related to their age. Their probability of death can also be correlated to their age.

Another demographic factor relates to the type of contact between individuals. Transmission probability is significantly higher when a susceptible individual is in close contact with infectious individuals for a cumulative period of time (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. This was confirmed during the SARS outbreak in 2003 where a large portion of the family members of the initial cases were infected (Poutanen et al. 2003). Therefore, the number of individuals in each household plays an important factor in estimating the size of an outbreak. A distribution of household sizes can be obtained from census reports and statistical data and incorporated into the model by assigning contacts times \( t_{ij} \) and transmission rates \( b_{ij} \) accordingly.

Behavioral aspects include daily commutes, attendance at schools/workplaces/hospitals, etc., and intuitively present significant environments wherein transmission can occur. Most respiratory diseases have an asymptomatic period of time during which an individual is infected but not yet aware of the illness, and therefore continues about his/her daily activities. During this time, the disease can be transmitted in a number of environments, including 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).

People’s behavior during the course of the simulation can change due to a realization of the outbreak. For example, infected individuals may voluntarily quarantine themselves or admit themselves to hospitals in time period \( n \) if they are aware of the dangers of the spreading disease. To model this, their corresponding \( t_{ij} \) and \( b_{ij} \) can be modified accordingly. If more social, demographic or behavioral data become available, they too can be easily incorporated into the model.

Another significant behavior aspect is the use of public transportation. Disease transmission is believed to only occur within a two-meter radius (Brankston et al. 2007). Thus, an infected individual will only have a large probability of transmitting the disease to persons traveling the same transit route who are probabilistically located within a two-meter radius of the infected individual. Detailed data regarding bus, streetcar and subway passenger density and ridership can be obtained from the public transportation provider. Individuals can be randomly assigned public transportation based on labor statistics.

In addition, labor statistics as well as hospital polling indicate the number of doctors, nurses and other healthcare workers in the population. Individuals who are healthcare workers can be randomly assigned based on those numbers, and individuals who are in-patients hospitalized for non-outbreak reasons can also be randomly assigned based on hospitals’ average in-patient numbers. The ability to model healthcare workers in a pandemic is important due to their increased exposure and increased, close contact with susceptible patients who may already have compromised immune systems.

Infected individuals can arrive at hospitals during the simulation, and thereby increase the probability of the outbreak propagation through the hospital. If it is assumed that the city is aware of the outbreak, patients presenting with outbreak symptoms may be treated in isolation or with extra protective measures, thereby reducing contamination of the entire facility. This scenario can also be easily modeled by assuming that infected individuals in hospitals after day \( n \) (when it is realized that an outbreak is occurring) have lower quality contact with healthcare workers in the form of extra precautions taken to prevent transmission.

3 RESULTS

The goal of this work is to illustrate that non-homogeneous agent-based simulation is a viable tool for modeling the spread of a pandemic in an urban population. The simulation model was tested on a pandemic influenza outbreak in the Greater Toronto Area. Preliminary results of accounting for certain demographic and behavior patterns are presented, but they are not exhaustive results. Rather, they are a demonstration of the power and ease of agent-based simulation for this application. More thorough results are planned for the future.

The GTA consists of 4.99 million individuals living in 1.8 million households, according to a 2006 census. The census data was used to determine the probability of a household being located in a particular census tract (areas comprising roughly 4,000-6,000 residents). The census data was also used to determine household size distributions and age distributions.
Two types of contact were considered between individuals: close and casual. At this point, only contact between household members and random members of the population, as well as contact on public transportation routes, is considered. More contact scenarios will be added in future versions.

Close contact exists between members of the same household and commuters on the same public transportation routes (since the subway lines are densely packed during rush hour). The number of individuals in a household is determined by census data, while the number of individuals per public transportation route is determined as described below. Casual contact occurs with random members of the population living in the same census tract. The number of random casual contacts comes from estimates provided by (Haber et al. 2007), who specified both inter- and intra-age group average daily contact numbers.

In order to address public transportation, the simplifying assumption that every individual of working age (18-65) traveled the nearest subway line to the main downtown hub station (Union Station) twice a day (to and from work). Public transportation data was obtained from the Toronto Transit Commission (TTC), and only subway routes during rush hour were considered in the model. Specific TTC data used in the model includes average time (in minutes) of each subway route, average passenger density per subway car, average number of passengers per subway car and number of cars per subway route. Using this information, we can estimate the probability that a healthy individual is within a two-meter radius of an infected individual traveling along the same subway route. We consider direct transmission to only be possible in a two-meter radius based on the same conclusion reached by (Low et al. 2007).

Indirect contact on public transportation, e.g., an infected individual sneezing on his/her hand and then touching a grab bar which is later communicated disease to a healthy individual, is assessed by setting \( c_{ij} \) to an appropriate value for all healthy individuals traveling on the same route as an infected individual. As per the recommendations of our epidemiologist collaborators, this value is currently obtained by manually tuning until the resulting disease spread via TTC is within their expectations.

Pandemic influenza was chosen as the disease due to the availability of disease transmission data. Our model uses the pandemic influenza transmission data presented in (Haber et al. 2007); the reader is referred to this publication for specific numbers regarding transmission rates per unit time of contact. This transmission data was chosen because it specifically addresses inter- and intra-age group transmission rates, which are accounted for in the model. The transmission data in (Haber et al. 2007) also provides transmission rates per unit time for close and casual contact scenarios, as well as average lengths of contact time in close and casual settings. In our model, we consider intra-household contact and public transportation contact to be close, while random social contact is casual.

A major concern for agent-based simulation is the computational resources and time required to complete the simulation, especially when simulating a population of almost 5 million individuals. In order to create a model that is as efficient as reasonably possible, the model is written entirely in C++. The output of the model is sent to geographic information systems (GIS) software to display maps of outbreak severity to provide visual tools to assist in policy-making decisions.

On a well-equipped desktop computer, a single day can be simulated in 1-2s, and 60 days can be simulated in \( \approx 1.46\text{min} \). This demonstrates that computer resources are not a limitation of the model. Despite the speed of the simulation model, 1,000 simulations would still require about 24 hours of computation time. Therefore, 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 \( \approx 2.98\text{min} \) on the cluster due to overhead communication and slower file input/output time, all 1,000 simulations were completed in less than 13min.

Two versions of the model were run. One version considered that 50 percent of infected individuals would stay home when they became ill. This is not the same as a self-quarantine. In a quarantine, the infected or exposed individual isolates him/herself completely from both household members and the outside population. When a person stays home, which is a common behavior upon experiencing symptoms of illness, they no longer have contact with the outside world, but their time and quality of contact with members of the household is assumed to remain the same. However, this assumption can be easily changed by modifying \( b_{ij} \) and \( t_{ij} \) to test outcomes using different assumptions regarding changes in individuals’ behavior inside the home.

The other version modeled a scenario in which infected persons continued about their usual day without taking any precautions to prevent transmission to the rest of the population. Both versions of the model simulated 60 days and were run for 1,000 simulations.

The results in Figure 2 confirm expectations that the outbreak is less severe when a significant number of infected individuals stay home. For each of the recorded measures—current number of individuals infected, cumulative number of individuals infected and number of deceased individuals—the outbreak results in fewer affected individuals when 50 percent of infected individuals only have contact with their households. There is also less variation in the outcomes of the individual simulations, which is also expected since there are fewer infected contacts in the population.
Figure 2: Output of the two versions of the model. Left: Infected population members continue about their daily business. Right: Fifty percent of infected population members choose to stay home. Top: Number of individuals who are currently infected per day. Center: Cumulative number of all individuals who have been infected per day. Bottom: Cumulative number of individuals who have died per day. All x-axis values are in hundreds of people.
The output of the simulations were imported into ESRI ArcGIS software to obtain shaded maps of the GTA, where more shading represents a higher percentage of population infected in that census tract. Figure 3 shows the GIS maps of trials that obtained the average number of cumulative infected individuals at the conclusion of the 60-day simulation. As expected (and confirmed by the graphs in Figure 2), the severity and geographic spread of the outbreak is more serious when individuals choose to not stay home when they become ill. Even if only 50 percent of the infected population stays home, there is a significant impact on the disease spread. Note that the mitigation of disease spread would be even more pronounced in the event of an actual quarantine, wherein infected individuals would not even be allowed contact with their families, even though they would be confined to their homes.

4 CONCLUSIONS AND FUTURE WORK

This research demonstrates that agent-based simulation is a reasonable tool for predicting disease spread during a pandemic in large populations. The use of agent-based simulation models allows for more information about the outbreak to be tracked than can be with traditional non-homogeneous mixing model calculations. GIS maps were easily generated from the predicted outbreak, which cannot be done reliably with homogeneous and non-homogeneous mixing models.

This work also enhanced the accuracy of traditional outbreak models by accounting for factors including age, household/family interaction, casual interaction, geographic location and public transportation use. Although these are a small number of considerations, they intuitively have significant bearing on the course of the outbreak. More factors can be considered without fundamental changes to the model structure or calculations. Several extensions are planned for future models.

Future work will incorporate specialized parameters for healthcare workers and young children in schools and daycare, as well as for the expected behavior of infected individuals going to the hospital once symptoms are evident. More advanced assignments of public transportation use will also be developed. The model will additionally be expanded to test mitigation strategies, for example, vaccine drives, school closures and quarantines.

Despite the unique capabilities of the presented model, it is important to emphasize that no simulation model regarding disease spread is perfectly accurate. The purpose of these models is to help policy makers prepare for outbreaks, whether by indicating the average number of expected infections (as in homogeneous and non-homogeneous mixing models) or by providing a platform to compare the outcomes of different mitigation strategies (as in the model presented). Both of these models have shortcomings. Even if it is assumed that $R_0$ is an acceptable one-size-fits-all measurement of disease spread, that number is very difficult to calculate as demonstrated by the fact that $R_0$ for the current swine flu pandemic is still unknown nearly two months after original detection. The presented model, on the other hand, requires the also difficult to obtain transmission rate per unit time. Although the presented model can do much more with its data, both types of models are heavily reliant on their own types of information. It is therefore incumbent on public officials and epidemiologists to create means to collect and share information and to perform tests on disease samples so that the necessary information is available quickly in the early stages of a pandemic.

REFERENCES


Figure 3: GIS maps of every sixth day of representative simulations. Left: Infected population members continue about their daily business. Right: Fifty percent of infected population members choose to stay home.


AUTHOR BIOGRAPHIES

DIONNE M. ALEMAN is an Assistant Professor in the Mechanical and Industrial Engineering at the University of Toronto. She received her Ph.D. in Industrial and Systems Engineering from the University of Florida in Gainesville, FL. Her research interests are medical applications of operations research, specifically, pandemic outbreak planning and radiotherapy treatment optimization. She serves as the Vice-Chair and Chair-Elect for the INFORMS Health Applications Section and is the director of the Medical Operations Research Lab (morLAB) at the University of Toronto. Professor Aleman’s email address is <aleman@mie.utoronto.ca>.

THEODORUS G. WIBISONO received his B.Eng. in mechanical engineering from University of Indonesia. He is currently pursuing his M.Eng. in the mechanical and industrial department, University of Toronto. His current research interests include disease transmission modeling, emergency preparedness, geospatial analysis and engineering application in public health. He has co-authored two publications in the field of disease transmission modeling. Mr. Wibisono’s email address is <theo.wibisono@utoronto.ca>.

BRIAN SCHWARTZ is the Director of the Emergency Management Support of the Ontario Agency for Health Protection and Promotion (OAHPP), and Scientific Advisor for the Emergency Management Unit, Ontario Ministry of Health and Long Term Care. He has collaborated on health services and systems research in emergency medical dispatch, emergency department overcrowding and mitigation strategies for pandemic influenza, and has supervised Masters and Doctorate level students in faculties of Medicine and Engineering. Dr. Schwartz is also Associate Professor in the Department of Family and Community Medicine at the University of Toronto, and currently is Chair of three working groups for the Ontario Health Plan for an Influenza Pandemic, including the Rapid Access to Antivirals working group. Dr. Schwartz’s email address is <brian.schwartz@oahpp.ca>.