

Comparison between Individual-based and Aggregate Models in the context of Tuberculosis Transmission

Yuan Tian¹, Nathaniel Osgood²

Department of Computer Science, University of Saskatchewan

176 Thorvaldson Bldg., University of Saskatchewan

110 Science Place, Saskatoon, SK S7N 5C9, Canada

Phone: (306) 966-4886

Email: yut473@mail.usask.ca¹, osgood@cs.usask.ca²

ABSTRACT

The desire to better understand the transmission of infectious disease in the real world has motivated the representation of epidemic diffusion in the context of quantitative simulation. In recent decades, both individual-based models and aggregate models (such as System Dynamics) are widely used in epidemiological modeling. This paper compares the difference between aggregate models and individual-based models in the context of Tuberculosis (TB) transmission, considering smoking as a risk factor. The merits and impact of capturing individual heterogeneity is examined via representing Bacillus Calmette-Gurin vaccination and reactivation in both models. The simulation results of the two models exhibit distinct discrepancies in TB incidence rate and prevalence. Results also suggest that, at the level of practical application, individual-based models offer significantly greater accuracy and easier extension, especially when representing a decreasing reactivation rate, waning of immunity and heterogeneous individual attributes. Another experiment sought to evaluate the impact of network structure on TB diffusion. Simulations are conducted under three widely used network topologies, namely random, scale-free and small world. The results reveal large differences between results of individual-based models and aggregate models, which further give insights into the difference between these two model types in the context of practical decision-making.

Keywords: individual-based modeling, aggregate modeling, infectious disease, network topology, tuberculosis transmission, epidemiology

Introduction

Every year, infectious diseases cause more than 13 million deaths worldwide, with two-thirds of them occurring among children under 5 years old (GAO, 2001). The top infectious disease killers include Human Immunodeficiency Virus (HIV), Tuberculosis (TB) and malaria (GAO, 2001). Tuberculosis (TB), as an airborne bacterial infection caused by bacillus *Mycobacterium tuberculosis*, is a major cause of global mortality and morbidity, especially in poor and developing countries with limited health care resources and weak health care systems. Although it is a curable and preventable disease, it is reported that two million people die annually from TB (WHO, 2011). In Canada, despite the adoption of guidelines and prevention programs, the incidence of TB remains high in certain geographic and demographic zones. In addition, a variety of epidemiological studies have found that smoking is a risk factor for lung cancer, chronic pulmonary and cardiovascular disease. The association between smoking and Tuberculosis is evaluated in many studies, and some evidence suggests that smoking is strongly associated with development of Tuberculosis, mortality of TB as well as development of severe (and particularly infectious) forms of active TB (Hassmiller, 2006).

In the past decades, a group of distinguished infectious disease specialists have contributed remarkable knowledge to mechanisms of infectious disease pathogenesis and diagnosis. Despite such gains, we are still facing great challenges in early detection and in the development of effective control programs and policies to avoid global outbreaks.

With growing computational power, modeling techniques have increasingly attracted attention as ways of enriching understanding of the causal pathways of infectious disease and for aiding policymakers to implement effective control strategies to prevent the spread of diseases. Computational modeling offers the ability to analyze various possibilities of disease containment and to answer “what-if” questions. In current computational simulation studies, two popular approaches to epidemiological modeling are System Dynamics modeling and agent-based modeling. System Dynamics models of infectious disease spread commonly implement structural principles drawn from the most traditional mathematical epidemiology models, which are aggregate in character. However, there has been a limited amount of System Dynamics modeling performed at the individual level (Vickers and Osgood, 2007). With respect to dynamics of disease, the classic System Dynamics model for propagation of infectious disease is the susceptible-infectious-recovered (SIR) model, firstly developed in 1927 and which has provided fundamental insights into the disease diffusion (Anderson and May, 1991). In such aggregate models, individuals are aggregated into larger groups with same abstracted properties.

Although aggregate modeling can offer powerful insights and has allowed the derivation of the foundational concepts of mathematical epidemiology, there are distinct limitations associated with aggregate modeling when the focus is upon the specifics of the interactions or social contacts through which the infection is spreading. Spurred by increasing computer resources and the needs for realistic scenario evaluation, agent-

based modeling has become increasingly popular. This reflects the fact that it lends extra flexibility in terms of representing population as a system of interacting agents with heterogeneous features and abilities. Social network modeling and analysis, as a complement to agent-based modeling, takes into account the importance of contact structure, pathways of infection spread across the associated transmission and social networks.

Both of these two modeling approaches offer some important insights into the mechanisms of infection dynamics, but the underlying assumptions of these two simulation approaches are quite different. In the context of infectious disease, people groups within same category (stocks in System Dynamics models) are assumed to be homogeneous and well-mixed, which indicates that each individual has an equal chance to spread the disease to every other (Rahmandad and Sterman, 2008). As the disease rests purely upon contacts with infectious individuals, assuming homogeneity and perfect mixing can reduce accuracy in assessing intervention trade-offs and undermines the validity of the model. While the random mixing assumption within aggregate models can be relaxed to allow for representation of distinct groups that exhibit preferential mixing, the representation of such mixing can be cumbersome and complicated. By contrast, agent-based models (as a particularly attractive class of individual-based models) not only can capture feedback effects but also are quite flexible and handy in implementing heterogeneity of individual characteristics (including history information) and for evaluating the interaction of individuals at certain points in a network. However, agent-based models carry their own trade-offs, as they suffer from high computational cost – a substantial concern in light of our limited time and resources, particularly when we are conducting sensitivity analysis and other forms of model analysis. Which modeling approach is more efficient or faithful? To what degree does the added flexibility and finer granularity of agent-based modeling really yield practical benefits when representing realistic models? When should aggregate modeling approach be used, and when are agent-based models more suitable? In this paper, we carry out controlled simulations to compare the difference between agent-based models and aggregate models in the context of M. Tuberculosis transmission. In addition to facilitating an understanding of modeling trade-offs, this approach also aids our understanding of Tuberculosis transmission by using different methodologies of computational modeling.

The paper is organized as follows: epidemiology of Tuberculosis is introduced in the next section, then structure of both the aggregate model and the baseline individual-based model are represented in details. Given the structure of the models, verification of the baseline models and detailed experimental design are provided afterwards. We have designed three experiments with respect to BCG vaccination, non-memoryless reactivation and network topologies. In the last section, the results of the simulations are illustrated in conjunction with discussion about trade-offs between aggregate modeling and individual-based modeling.

Epidemiology of Tuberculosis

Tuberculosis (TB), as an airborne bacterial infection caused by bacillus *Mycobacterium tuberculosis*, is a major cause of global mortality and morbidity, especially in poor and developing countries with limited health care resources and weak health care systems. TB has infected approximately 2 billion people worldwide, and around 10 percent of these infected people will develop active TB in their rest of lives (WHO, 2011). Although it is a curable and preventable disease, it is reported that roughly two million people die annually from TB (WHO, 2011). In Canada, despite the adoption of guidelines and prevention programs, the incidence of TB remains high in certain geographic and demographic zones. Saskatchewan is one of the provinces in Canada possessing a higher incidence rate of TB; however, this statistic masks the tremendous variability in TB risk. Most notably, the large majority of cases in Saskatchewan occur in Aboriginal peoples, including First Nations people.

Before proceeding, we present a brief overview of the terminologies used in the epidemiological context and throughout the models in this study. It is worth noting that TB bacteria transmit from person to person through the air. Infection by TB bacteria does not automatically bring on TB disease. Usually there is an incubation period before an infected individual physically develop the current disease, and there are individual differences in latency.

- *Active TB Disease.* The term “Active TB” typically refers to current disease; people with Active TB typically feel sick and may have some known pathologies in parts of their body (such as enclosed granulomas in various organs) where the TB bacteria cluster. Tuberculosis most commonly affects the lungs, yielding what is known as pulmonary Active TB, and it also can spread to other organs such as bone and brain. Pulmonary TB can be infectious. People with pulmonary infectious TB can breathe out tiny droplets containing TB bacteria when they are coughing, sneezing, singing, and even when just talking (PHAC, 2008). These TB droplets remain in the air for a couple of hours, and people who breath in these TB droplets are exposed to TB bacteria.
- *Primary Progression.* After acquiring TB infection through contact, a small fraction of those infected people will develop active TB in a relatively short period of time due to ineffective control of infection by their immune systems or some other reasons. This mechanism, in which the bacteria evade effective control, is termed as primary progression. The mean time for primary progression varies in different studies. Since TB is an slowly growing infectious disease, commonly used time limits for primary progression is 2 years or 5 years (Hassmiller, 2007; Vynnycky, 1996).
- *Latent TB Infection.* Those infected who are able to effectively control their TB infection without developing active TB are referred as being in a state of latent TB infection. Within the latent stage of infection, people are infected but

they don't have symptoms and the bacteria remains dormant in their organs. Most of them will remain in the latent TB infection stage for their rest of lives; only a small percentage of them will go on to eventually develop active TB. Such reactivation can be brought on by a weak immune system, poor health care, or a combination of other complex risk factors. Theoretically speaking, in a latently infected person, either the TB bacteria is still alive but inactive in his cells or his immune system might completely kill the bacteria. However, it is currently impossible to differentiate between them with readily available diagnostic technologies. Latent TB cases with inactive bacteria can develop the disease later on in life through re-activation; on the other hand, for those with killed bacteria, they can develop the disease through re-infection.

- *Reactivation.* Reactivation refers to the progression to Active TB disease resulting from a latent infection gained a relatively long time ago. Reactivation can be triggered by many of complex risk factors such as HIV, use of immunosuppressant drugs, and weak immune systems.
- *Reinfection.* When an individual remains in the latent stage, he or she can get reinfected by another strain of TB bacteria via a mechanism often referred as reinfection. There is significant controversy regarding the level of reinfection that occurs within the population.

Because of the complexity of Tuberculosis pathology and heterogeneity of human immune systems, some of these terminologies or mechanisms are under debate, and different variations on the above may be used by different researchers and health care practitioners.

Structure of System Dynamics Model

Following the model structure in the literature (Hassmiller, 2007), Mahamoud et al. constructed an aggregate model of TB transmission, including smoking as a risk factor. The model reflects the characteristic stages of TB development as well as a stratification by smoking status.

A simplified version of the structure of Mahamoud et al.'s model is illustrated in Figure 1. Within this model, all the population are categorized into 6 stocks distinguishing people according to both TB and smoking status.: Uninfected Non-smokers (U_n), Latently Infected Non-smokers (L_n), Active TB Non-smokers (T_n), Uninfected Smokers (U_s), Latently Infected Smokers (L_s) and Active TB Smokers (T_s). As it is typical in System Dynamics models, the changes in the stocks over time are caused by inflows and outflows, including recruitment, death from TB or other disease, latent TB infection, primary progression, reinfection, reactivation, natural recovery and treatment. The time unit for the model is one year. The parameters used in the model are displayed in Table 1 and Table 2 (Mahamoud et al., 2009).

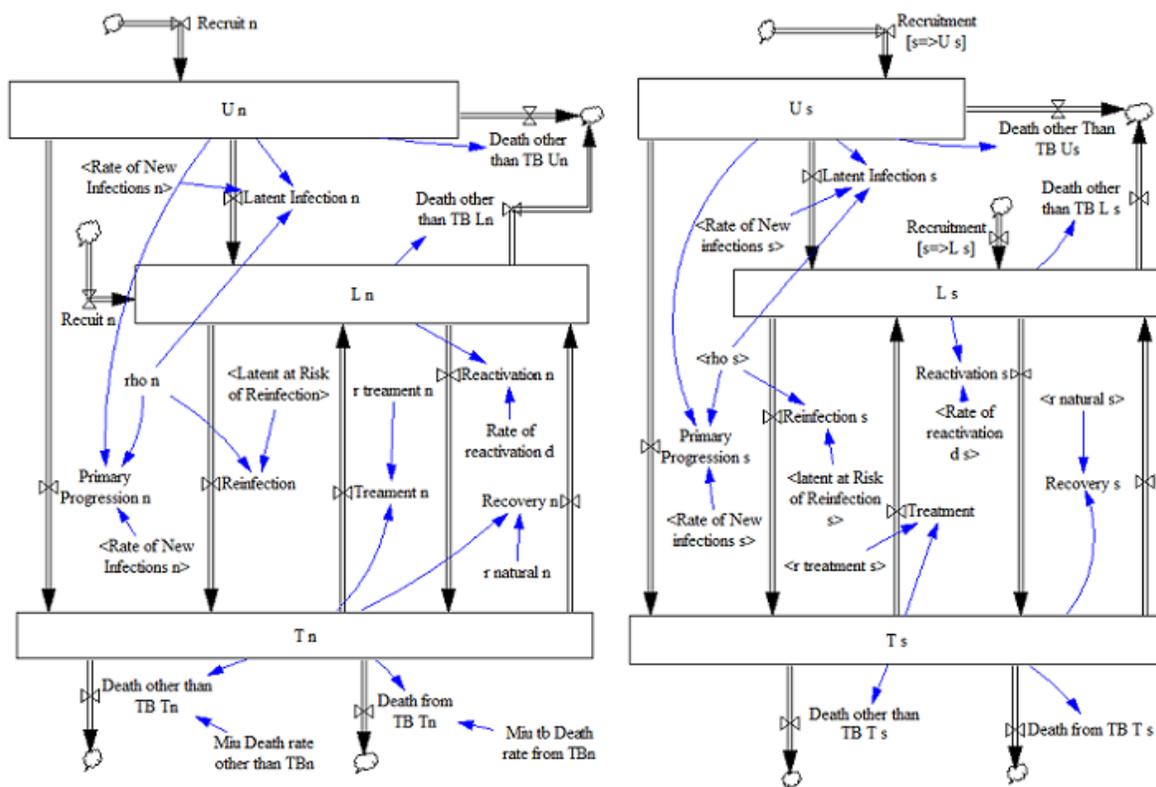


Figure 1: A schematic representation of Mahamoud et al.'s aggregate model of TB diffusion with smoking impact

Tobacco use has been an issue of concern for years, and in Northern Saskatchewan, the overall prevalence of smoking in 2004 was 41%, compared with 28% across the province (Nor, 2004). The relative risk for smokers progressing to active TB is much higher than that for non-smokers (Hassmiller, 2007). Besides those variables used in measuring the dynamics of TB spread, smoking impacts on TB transmission are captured using a list of parameters shown in Table 2. Because this TB model has not been calibrated with empirical data and some parameters are roughly estimated, it is used more for testing and exploring differences in methodology when measuring the dynamics of TB spread in North Saskatchewan, considering smoking as a risk factor.

The equations for non-smoker related stocks and flows illustrated in Figure 1 are as follows:

Parameter	Description	Value	Unit
βc	The average number of infections an individual with active TB (T) causes per year	7.788	person per year
ρ	Proportion of newly infected individuals progressing to primary TB	0.05	1
α	Proportion of the new entrants into the model who were infected prior to their entry time	0.054	1
γ	Treatment rate of TB	1	per person per year
τ	Rate of Natural Recovery	0.25	per person per year
d	Rate of Reactivation (Progression from latent TB to active TB due to endogenous changes)	0.003125	per person per year
e	Proportion of latently infected people with risk of exogenous reinfection	0.25	1
π	Number of new 15 years old entrants to the model per year	720	person per year
μ_{tbn}	Mortality rate from TB for non-smokers	0.037	per person per year
μ_n	Mortality rate from other disease among non-smokers	0.0274	per person per year
p	Proportion of the population over 15 years of age	0.66	1

Table 1: Description of the symbols and parameter settings in Mahamoud et al.'s model

$$\lambda_n = p\beta c \frac{T_n}{N_n} [\sigma_8 + (1 - \sigma_8) \frac{N_n}{N}] + p\beta c \sigma_6 (1 - \sigma_8) \frac{T_s}{N} \quad (1)$$

$$\frac{dU_n}{dt} = (1 - \sigma_0)(1 - \alpha)\pi - \lambda_n U_n - \mu_n U_n \quad (2)$$

$$\frac{dL_n}{dt} = (1 - \sigma_0)\alpha\pi + (1 - \rho)\lambda_n U_n + \gamma T_n + \tau T_n - e\rho\lambda_n L_n - dL_n - \mu_n L_n \quad (3)$$

$$\frac{dT_n}{dt} = \rho\lambda_n U_n + e\rho\lambda_n L_n + dL_n - \tau T_n - \gamma T_n - \mu_n T_n - \mu_{tbn} T_n. \quad (4)$$

Here, N_n denotes the sum of non-smokers in the population where $N_n = U_n + L_n + T_n$. Similarly, N_s denotes the sum of smokers where $N_s = U_s + L_s + T_s$. N denotes the total number of individuals in the population, where $N = N_n + N_s$. λ_n is the rate of new infection for non-smokers. The initial values for the stocks are $U_n(0) = 11429$, $L_n(0) = 2211$, and $T_n(0) = 24$.

Parameter	Description	Value
σ_0	Percentage of the entering clients who are initially smoking	0.412
σ_1	Relative risk imposed by smoking on the rate of new infection	1.93
σ_2	Relative risk imposed by smoking on reactivation	1.53
σ_3	Relative risk of primary progression given smoking	1.53
σ_4	Rate ratio for smoking on the natural recovery from Active TB	0.65
σ_5	Relative risk of TB death rate given smoking exposure	1
σ_6	Relative risk of becoming infected when contacting a smoker with Active TB (compared to contacts with a non smoker)	2
σ_7	Factor by which smoking affects the treatment rate	0.8
σ_8	Mixing parameter denoting the degree of assortativity between smokers and non-smokers	0.3
σ_9	Relative risk of non-TB death rate given smoking exposure	1.14

Table 2: Smoking related parameters in Mahamoud et al.'s model

The equations for smokers demonstrated in Figure 1 are:

$$\lambda_s = \sigma_1 p \sigma_6 \beta c \frac{T_s}{N_s} \left[\sigma_8 + (1 - \sigma_8) \frac{N_s}{N} \right] + \sigma_1 p \beta c (1 - \sigma_8) \frac{T_n}{N} \quad (5)$$

$$\frac{dU_s}{dt} = \sigma_0 (1 - \alpha) \pi - \lambda_s U_s - \sigma_9 \mu_n U_s \quad (6)$$

$$\frac{dL_s}{dt} = \sigma_0 \alpha \pi + (1 - \sigma_3 \rho) \lambda_s U_s + \sigma_7 \gamma T_s + \sigma_4 \tau T_s - e \sigma_3 \rho \lambda_s L_s - \sigma_2 d L_s - \sigma_9 \mu_n L_s \quad (7)$$

$$\frac{dT_s}{dt} = \sigma_3 \rho \lambda_s U_s + e \sigma_3 \rho \lambda_s L_s + \sigma_2 d L_s - \sigma_4 \tau T_s - \sigma_7 \gamma T_s - \sigma_9 \mu_n T_s - \sigma_5 \mu_{tbn} T_s. \quad (8)$$

Here, $U_s(0) = 8008$, $L_s(0) = 2618$, and $T_s(0) = 118$; λ_s denotes the rate of new infection for smokers, and incorporates many factors by which smoking impacts TB progression. In this model, smokers are more susceptible to TB infection (meaning that the chance that they get infected given exposure is higher than for non-smokers). Moreover, smokers are more likely to transmit the disease so that (as given by σ_6) the average number of infections caused by an smoker with active TB (when surrounded by a given group of people) is twice of that by a corresponding non-smoker when surrounded by those same people. Besides, smokers also have a relatively high risk of developing primary progression, reactivation and reinfection. In addition, the death

rate for smokers is also higher than that for non-smokers. In Equation (1) and (5), the assortivity coefficient σ_8 is implemented to represent the interaction pattern between smokers and non-smokers. When $\sigma_8 = 0$, it indicates that all of the population is randomly mixed with each, with no distinction made according to smoking status; however, when $\sigma_8 = 1$, smokers only interact or contact with smokers and non-smokers only mix with non-smokers. By default, $\sigma_8 = 0.3$ means individuals with same smoking status prefer mixing with those sharing their smoking behavior, but also mix with those of different smoking status.

Structure of Individual-based Model

In this work, we recreated Mahamoud et al.’s model in an individual-based fashion in the AnyLogic software package, retaining the same parameters, values, transitions rates and interactions among the agents.

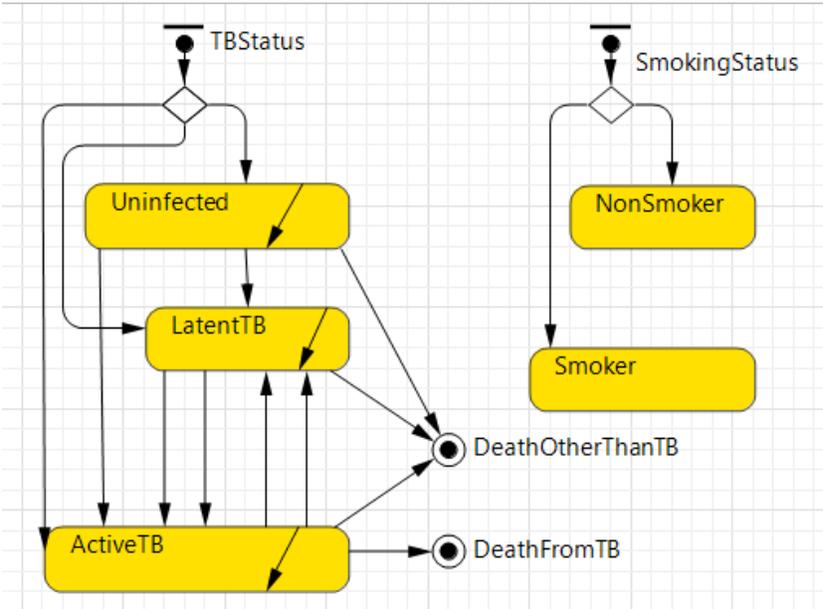


Figure 2: Structure of individual-based Mahamoud et al.’s model of TB diffusion with smoking impact

Two state charts were created to manage the structure of the model shown in Figure 2. One state chart represents individual progression of TB infection; the other is used to represent the smoking status of each individual. The structure clearly shows two dimensions of each individual: TB infection status and smoking status. All the transitions are implemented using rates the same as corresponding rates in the aggregate model. It can be observed that there are some differences in the implementation of these two types of models. In the aggregate SD model, six stocks accumulate and maintain the population in different categories and the inflows and outflows are used

to control the level of the stock directly. So, for each stock, the change across a period of time equals the total inflows minus the total outflows over that period of time.

Each agent is associated with exactly one state of the TB progression state chart and one state of the smoking status state chart. States in an individual-based model don't accumulate a population; they are only used to represent each individual's state. Furthermore, transitions in individual-based models are quite different from the flows in an aggregate SD model. All the transitions in an individual-based model can be triggered at a certain rate, by a timeout, condition or message. Those transition parameters can be defined differently for individual with different attributes or state, or change over time. In the individual-based model, more attributes or status of the individual can be easily represented just by adding additional state charts or variables. Multidimensional status of individual can be captured without creating combinatorial combinations of compartments or stocks for each group of individuals with same attributes. Moreover, maintaining the distinct state charts (one for each transition) permits a "separation of concerns" (Dijkstra, 1982) that allows a modeler to more transparently understand the structure of individual progression along a particular dimension. Finally, given such a representation, it is quite visually clear which aspects of heterogeneity are static in character (requiring only a parameter), versus which are variable (requiring a state chart or variable). However, in an aggregate model, multidimensional representation is required for both static properties and for states (changing dynamically). Adding one more attribute for the population need to subdivide the existing compartments stratified for the new state. Within an aggregate model associated with multiple dimensions of heterogeneity, the need to distinguish individuals according to both static and dynamic attributes requires a separation of the stocks along the dimensions of these attributes. Because the logic associated with progression of individuals along each successive dimension of heterogeneity are all combined in a stock, it is not immediately clear which visual transitions are associated with a certain type of condition. This is particularly significant in light of the disaggregation required by both static and dynamic attributes, as it means that a user is unable to visually distinguish static attributes of heterogeneity from dynamic ones – thereby obscuring the scope of the model. While the rapid visual growth of the model can be somewhat ameliorated through the use of subscripting, the use of subscripting comes with its own drawbacks. Most notably, the equations for progression along different types of subscripts can interact to yield a large number of equations for each stock.

Methods

The individual-based model was firstly verified, and then controlled experiments were designed and simulated. The first group of experiments varied the implementation of individual heterogeneity, and compared the outcome with that of the aggregate SD model. The second group of experiments focused on different topologies, with each conducting 10 simulations of the individual-based model to study the degree of

difference obtaining with the aggregate model.

Verification of Individual-based Model

When first considering this individual-based version of Mahamoud et al.'s model, it is likely that it will contain bugs. Before proceeding, we sought to verify our model via making the individual-based model comparable with the aggregate model. The results of the individual-based baseline model should be analogous to that of the aggregate baseline model since they share identical values for parameters and transitions rates.

Individual Heterogeneity with Respect to BCG Vaccination

BCG, as a vaccine against TB, provides protection to people. Based on the available data, Mahamoud et al. estimated a duration of efficacy of 31 years in Saskatchewan, and the rate of people receiving BCG was assumed to be 20% per year in this simulation. When we try to integrate BCG as an intervention for TB in the aggregate model, since individuals are assumed to be homogeneous under the context of aggregate SD model, everyone administered BCG was assumed to be fully protected for a mean time of 31 years. While in the protected state, they are assumed to experience no risk of developing TB Infection.

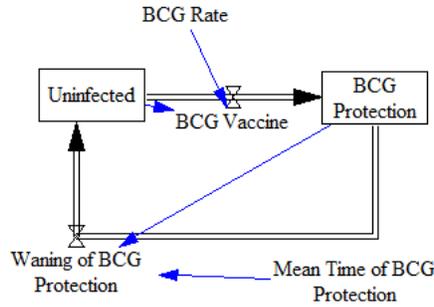


Figure 3: BCG implementation in the aggregate TB model

Figure 3 shows the simplified implementation of BCG in the aggregate model. The state equations for implementation of BCG in the aggregate model shown in Figure 3 are

$$\frac{dU}{dt} = \frac{B}{m} - kU \quad (9)$$

$$\frac{dB}{dt} = kU - \frac{B}{m}. \quad (10)$$

To this structure, U denotes uninfected individuals, B represents the people under BCG protection. k is the BCG rate per year, and m is the mean time of protection

conferred by BCG. In the aggregate SD model, BCG vaccination is implemented separately for smokers and non-smokers by adding two additional stocks to represent those vaccinated who are either smokers or non-smokers.

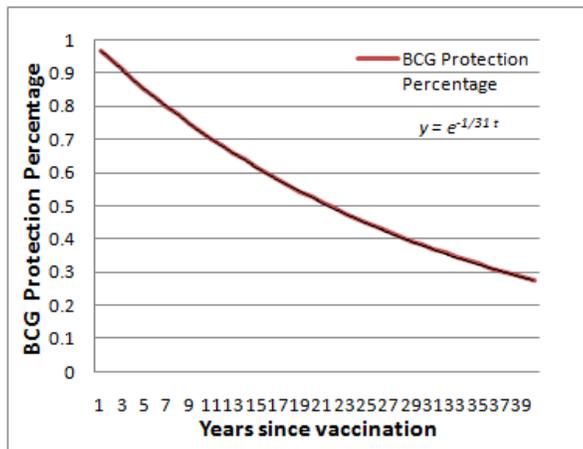


Figure 4: BCG Protection Percentage Over Time

However, the assumption that individuals are fully protected for 31 years is not completely reasonable. Some experts believe that the efficacy of BCG wanes like many other vaccines. This phenomenon is a reflection of biological understanding of the mechanisms of immune system memory. For example, antibodies and Cytotoxic T Lymphocytes decrease over time since last exposure (including vaccination). Moreover, in this view, individuals following vaccination are not fully protected, although they do have a lower chance of becoming infected. Given a mean time of protection of 31 years, we can derive a decreasing protection level from BCG. It is assumed that the BCG protection of an individual depends on the time since he or she receives immunization. The longer the time since that individual received the vaccine, the lower the degree of protection received, and the higher chance that individual will be infected given exposure (although this rate is still lower than that obtaining among those who do not receive the vaccine). In Figure 4, a set of equations describing the decreasing protection of BCG is demonstrated. It shows the fractional degree of BCG protection (y) as a function of the time since he or she was vaccinated, among those who remain uninfected.

Equation (11), coming from a first-order delay, is the mathematical solution of this declining protection level. Moreover, since the chance of developing disease among those who have BCG developing disease is not zero, Equation (12) and (13) are used to represent their risk of getting infected.

$$y = e^{-\frac{1}{31}t} \quad (11)$$

$$\lambda_{n,b} = (1 - y)\lambda_n \quad (12)$$

$$\lambda_{s,b} = (1 - y)\lambda_s. \quad (13)$$

Here y is the (fractional) protection conferred by BCG, $\lambda_{n,b}$ denotes the rate of new infection for the BCG-administered non-smokers remaining uninfected; $\lambda_{s,b}$ denotes the corresponding rate for smokers. Using this characterization, we extended the individual-based baseline model to implement BCG protection based on the period of time since each individual was vaccinated. This implementation is to evaluate the impact of heterogeneity of individuals on results. The “fully vaccinated – fully susceptible” dichotomy is the widely-used and traditional representation of vaccination effects. This experiment can lend some insights into the implications of different representations of BCG protection, and this declining protection of BCG for each individual can’t be easily captured in an aggregate model given a population that is vaccinated at different points in time. So the results of these two models are designed to provide some knowledge about the impact of heterogeneity of individuals on BCG intervention of TB.

Memoryless vs. Non-Memoryless Reactivation

An additional experiment sought to evaluate the merits and impact of capturing the heterogeneous individuals in their progression to Active TB via reactivation. In this experiment, I separately extend the previously created individual-based baseline model. The new structure of the model is similar to that described by Vynnycky (Vynnycky, 1996), and is shown in Figure 5.

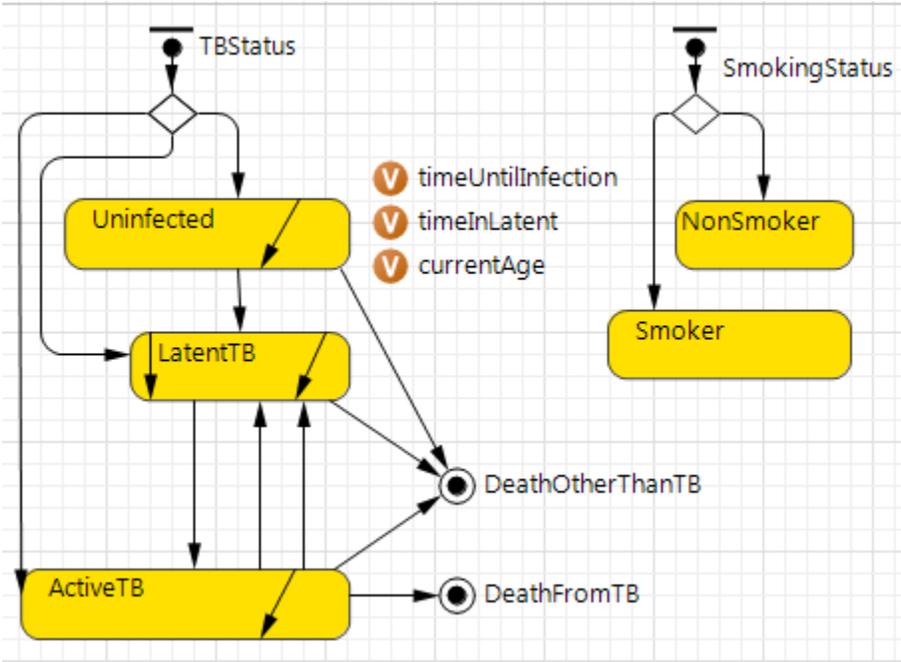


Figure 5: Revised Individual-based Model Structure with Respect to Reactivation

In this new structure, primary progression is no longer represented with a direct

transition from the uninfected state to the Active TB state. Instead, every individual is assumed to go to the latent state following infection. After infection, the chance he or she will develop disease depends on the rate of reinfection and reactivation. Reactivation in this model represents the progression to active TB, which is different from that in the model of Mahamoud et al.'s. In contrast to that model, the reactivation rate here depends on the time since he or she got infected. This reflects the fact that empirical observations suggests that the per-year chance for an individual to develop TB disease is relatively high for the first few years after he or she got infected, and then the chance will decrease over time (Vynnycky, 1996). We note that the “reactivation” transition in this model conceptually represents both primary progression (for those cases in which the progression to Active TB takes place in the first years following infection) and what is classically thought of as reactivation.

This model implements a reactivation rate that varies with the amount of time that has elapsed since infection. In the previously created baseline model, the reactivation rate (d) for non-smokers is 0.003125 per year, while that for smokers ($\sigma_2 d$) is 0.0047 per year. Since the model time line is 50 years, we can derive that the chance that a non-smoking or smoking individual develops active TB via reactivation over the course of those 50 years as follows. We note that this calculation ignores the effects of re-infection and the competing risk of non-TB induced mortality.

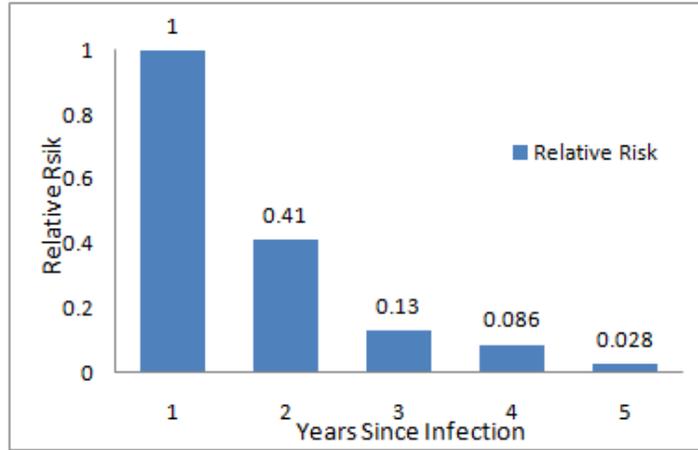


Figure 6: Relative risk of developing Active TB reproduced from Vynnycky (1996)

$$\begin{aligned}
 d'_n &= 1 - e^{-d \times 50} \\
 &= 0.1447 \text{ per 50 years} \\
 d'_s &= 1 - e^{-\sigma_2 d \times 50} \\
 &= 0.2094 \text{ per 50 years.}
 \end{aligned}$$

Figure 6 shows the relative risk of developing active TB since infection. According to the relative size of the rates of years 0 to 5 after infection in Figure 6, we assume the

reactivation rate maintains an exponential decline throughout these years. Following the approach of Vynnycky, we assume that the relative risk will keep constant beyond 5 years since infection, remaining at a rate equal to that of 5 years after infection.

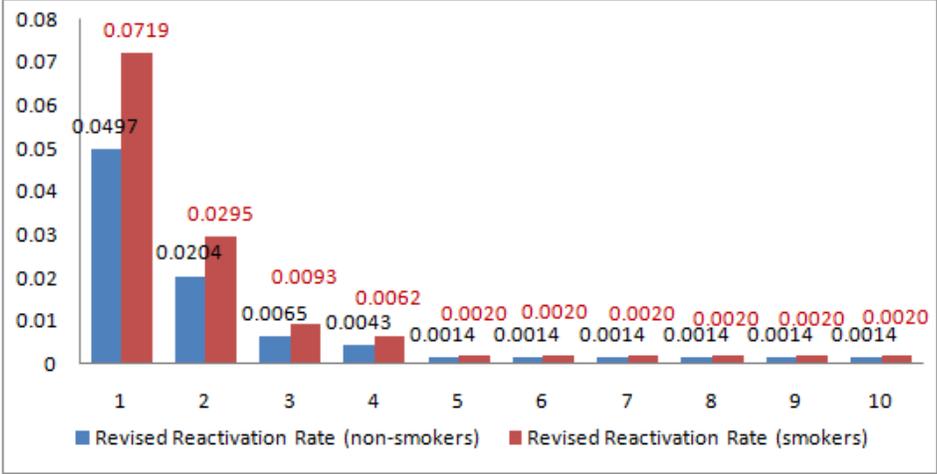


Figure 7: Revised reactivation rate since infection

To compare the results of two models on an equitable basis, it was important that the overall risk of reactivation is preserved. On the basis of this representation, we re-normalized the reactivation rate for each year since infection by maintaining the 50-year risk identical to that in the aggregate model. The revised reactivation rates since infection for our model are shown in Figure 7. Figure 7 only shows reactivation rates for the first 10 years since infection; as noted above, the reactivation rate beyond 10 years since infection is equal to that of 5 years after infection. Besides the decreasing reactivation rate since infection, we further added two attributes to each individual: their current age and the time since he or she developed TB after infection. This require adding a few variables and functions in the Person class. The age of each individual is initialized randomly with a uniform distribution extending between 0 and 75 years of age.

This experiment seeks to give us some insights into the importance of capturing the heterogeneity of individuals' history within a model. Since the structure of the original baseline model was extended, the results of this model are not fully comparable with that of baseline model. But the experiments will provide us with some valuable and detailed information regarding TB transmission, as well as insights into the trade-offs between the two modeling types.

Experiments with Network Structures

In an aggregate SD model, individuals within a compartment are assumed to be perfectly mixed with each other. This means that everyone in the same compartment experiences an identical chance to progress on to another state (such as Active TB)

and an identical chance to meet another. However, such a representation offers limited consideration of the impact of persistent connections between those in the populations, such as those that are common as a result of family structure, workplaces, and limited geographic mobility. In this section, we created a network representing the whole population. Transmission of the disease is triggered by specific person-to-person interactions among the individuals rather than via a calculation based on the mean rate of exposure of a susceptible individual to infectious individuals. Every individual is living in an environment which is defined by certain types of networks. Network topology refers to the layout of the connected nodes.

In order to make the network structured model and aggregate model comparable, σ_8 was set to be 0 in both the individual-based and aggregate model. σ_8 represents degree of assortive mixing between the smokers and non-smokers. $\sigma_8 = 0$ means smokers and non-smokers intermingle without distinction as to smoking status. By contrast, $\sigma_8 = 1$ means that individuals mix in a perfectly assortive fashion – in other words, smokers have no contact with non-smokers and non-smokers also have no chance to meet smokers. Furthermore, in order to maintain a stable network structure, recruitment and death are disabled within this experiment. Although stopping the recruitment and death might lead to an incorrect estimates of the dynamics of TB diffusion in the real-world population, comparing the two models in the absence of such processes will still provide us with some understanding regarding how network structure influences TB transmission.

In order to create a networked individual-based model comparable with the aggregate model, we needed to establish a common risk of infection. The aggregate model maintains a traditional representation of transmission of infection, which is governed by two key parameters – β and c . Because these two parameters are only used in the aggregate model when multiplied by each other, rather than considering each in isolation, it is most convenient to consider the product of the two, βc . This product represents the number of people that an infective person will infect per unit time (here, per year) when surrounded by otherwise susceptible people. In the baseline model, βc for non-smokers is 7.788 persons per year, and that for smokers ($\sigma_6 \beta c$) is 15.57 persons per year. In this experimental design, we assume that the average contacts per susceptible for smokers are the same as that for non-smokers. Based on previous work, we assume here β roughly equals to 0.45. Then we assume that for non-smokers, $\beta_n = \beta = 0.45$ and for smokers $\beta_s = \sigma_6 \beta = 0.9$. From the value of β , and for βc , we can then calculate that the average contacts per susceptible (c) for non-smokers or smokers are around 17 persons. Under each type of network structure, the value of β and c are set as noted above.

Following the establishment of the experimental design, we integrated network structure by extending and revising the individual-based baseline model. The surrounding network of individuals will be separately set to be random, scale-free and small world. We held the same average connections (17 persons) per agents in the random and small world networks, but we didn't hold the same for scale-free network. The following subsections provides background on each of these types of networks. In the simulated

network, only TB cases can transmit the disease by sending messages.

Random Network

Random networks were first presented by Erdős and P. Rényi. If N nodes are connected with n edges in a random network, these edges are selected with uniform probability among $N(N - 1)/2$ possible edges (Ran, 2002). In a random network, each individual is connected randomly with a given average number of connections, regardless of any consideration of spatial position or other individual attributes.

In most analytically tractable random networks, the edges and links of each individual are fixed, which indicates that pathways of disease transmission are almost stable (Keeling and Eames, 2005). Lack of clustered groups and homogeneity of individual-level network characteristics make random network models analogous to a random-mixing aggregate model. Understanding gained from simulations and analysis of random networks can enhance our understanding of the impact of network topologies on disease spread and may aid in further developing more complex social network structures integrating heterogeneous features of individuals.

Scale-free Network

Scale-free networks exhibit a degree distribution following a power law. In reality, many empirically observed networks appears to be approximately scale-free; examples include citation networks, protein networks and social networks (Sca, 2011). Scale-free networks are far from homogeneous, as some individuals have a lot of connections, while most individuals are associated with relatively few connections. Compared with random networks, scale-free networks exhibit wider ranges of heterogeneous connections. In order to capture some complex features of disease spread, it is necessary to incorporate such super-spreaders with larger number of links into the network (Keeling and Eames, 2005).

Since scale-free network can display heterogeneity in terms of the number of contacts, individuals with many connections not only possess high risk of becoming infected (due to many pathways and links with people), but can also transmit the infection rapidly once they are infected. Such effects can, for example, allow an infection to remain endemic in subgroups of a broader population, even when the population average rates of contact would be insufficient to maintain that network. Capturing this phenomenon is of great interest to both modelers and epidemiologists, as effective disease control policy and prevention programs can be enabled when the dynamics of infection in the network and behaviors of these concentrated high risk individuals are well understood.

Small World Network

A small world network is type of network topology within which each individual is connected with a given number of nearby individuals, but there are some larger-range connections. Some observed social networks are found to be small world networks.

In another words, small world networks integrate both locality of connections among individuals (which add the fact that two connected individuals are likely to share additional connections) and some long-range links through which transmission events can be performed (Keeling and Eames, 2005). Such highly clustered connections can exhibit the spread of infection locally, while the long-range pathways can depict the transmission phenomenon that epidemic spread is rapid and unlikely to be constrained within small regions of the population (Watts, 1998).

Results

In this section, the results of experiments will be demonstrated and analyzed.

Individual-based Baseline Model

Since this individual-based model is a stochastic one, I simulate the AB baseline model for 100 runs to verify that it yields results comparable to those associated with the aggregate model. It turns out that the behaviors of these 2 models are almost the same, and the differences between these two models are small and can be explained by stochastic factors, seen in Table 3. The bigger discrepancy of T_n and T_s are due to stochastic factors because of small size of population in these two categories. Jacquez and Simon have proved that small populations are highly affected by stochasticity (Jacquez, 1993). However, the difference will practically disappear when the population in these stocks is above 100. The relative discrepancy for U_n , U_s , L_n and L_s stocks/states is quite small, and is due to stochastic effects.

Stock/State	SDM Results	ABM Mean Results	ABM Std. Deviation	Minimum	Maximum
U_n	8415	8443.04	213.1	7955	8925
L_n	6536	6504.52	182.4	5998	6904
T_n	23.54	23.83	5.3	11	44
U_s	2664	2693.48	132.8	2334	3062
L_s	5981	5943.39	132.9	5612	6308
T_s	49.35	49.25	7.4	32	78
Total Pop.	23668.9	23657.51	148.5	23301	24043

Table 3: Comparison of baseline models results at 50th year

This comparison reflects the fact that the aggregate SD model is a continuous deterministic model which gives a single outcome, while the individual-based based model is constructed from quantized individuals and yields a distribution of outcomes. The need to perform an “ensemble” including multiple simulations (“realizations”) in order

to gain a sense of the range of model behavior further worsens the heavy computational cost of individual-based models.

Evaluation of Heterogeneity through BCG Vaccination

Now we analyze the difference between individual-based models and aggregate models under the scenario of BCG Vaccination and waning immunity.

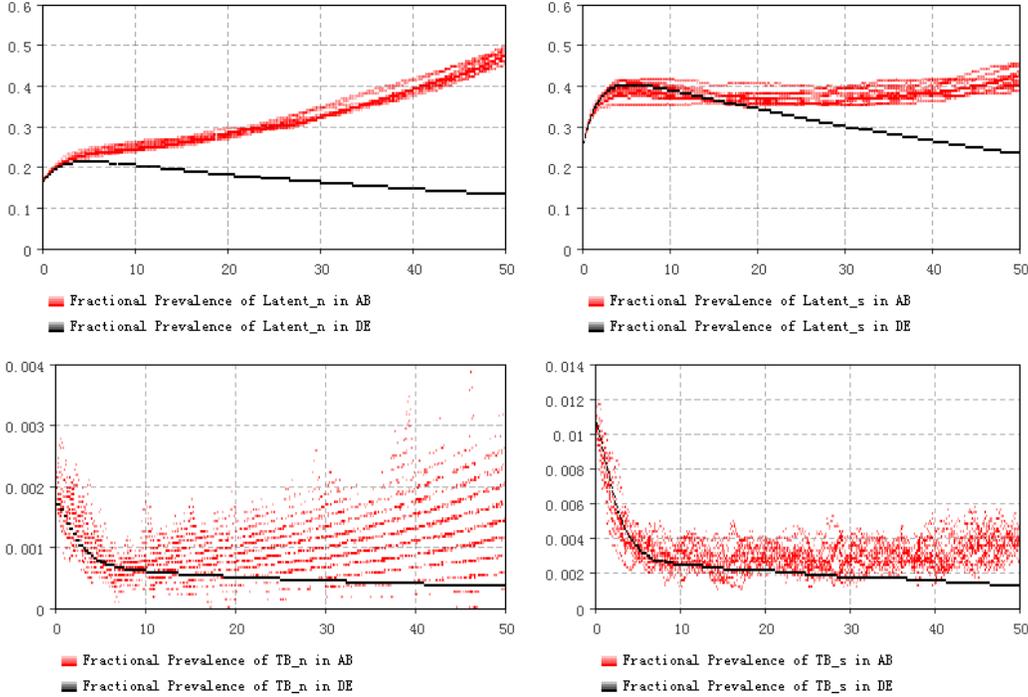


Figure 8: Prevalence of TB Infection and Active TB Given BCG Administration

Figure 8 shows the scenario results coming from both of these two models. The black lines represent the results coming from the aggregate SD model, while the red lines represent the simulation results coming from individual-based models. As noted in the methods section, each individual-based simulation was simulated for 10 realizations. Because of stochastic factors, the results of each simulation in the individual-based model are different from one another.

For non-smokers, we can observe that representing a continuously waning protection from BCG can produce a higher prevalence of latent TB infection compared to that resulting from use of a dichotomous protected/not protected distinction. For example, in the 50th years, the prevalence of latent TB infection in the individual-based model is almost three times higher than that in aggregate SD model. Similarly, the prevalence of active TB is also higher in the individual-based model when it is compared with that against an aggregate model. The situation for smokers displays a similar pattern to

that obtaining among non-smokers. It is worth emphasizing that these differences in rates emerge in spite of the fact that the decay rates in the individual-level model (on the one hand) and aggregate model (on the other) are identical.

In conclusion, the patterns of TB transmission resulting from the assumption of dichotomous waning BCG protection and continuously waning BCG protection over time are quite different. This gives some insights into the difference between these two models. It is possible to capture the heterogeneity of individuals in aggregate models by developing several compartments, each representing a different level of decayed immunity. However, such a representation is awkward and cumbersome, particularly when there are several other dimensions of heterogeneity present. By contrast, individual-based models can easily capture the heterogeneous attributes of each individual. From this point of view, individual-based models can represent the different attributes or status of individuals more gracefully than aggregate models.

The experiment on BCG protection shows that two theories of BCG protection duration produce distinct results, and many studies suggested that vaccines (including BCG) confer a decreasing protection over time. Given the divergence in results, it would appear that the representation of dichotomous susceptibility in the aggregate model represents too extreme a simplification of individual-level dynamics to adequately support investigation of intervention trade-offs.

Evaluation of Non-Memoryless Reactivation

Next, we consider the difference between individual-based models and aggregate models associated with the degree of memory associated with the reactivation process. While many runs of this scenario have been conducted, only one of them is presented here as an example to exhibit the findings.

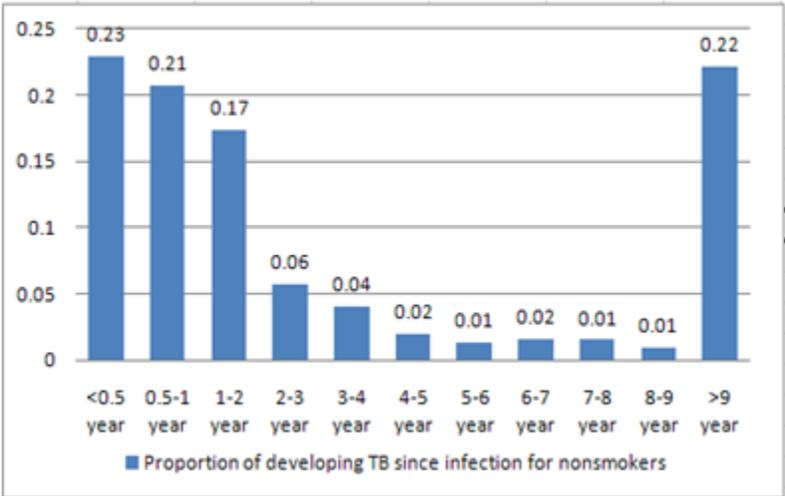


Figure 9: Proportion of non-smokers developing TB within a certain window of time following infection

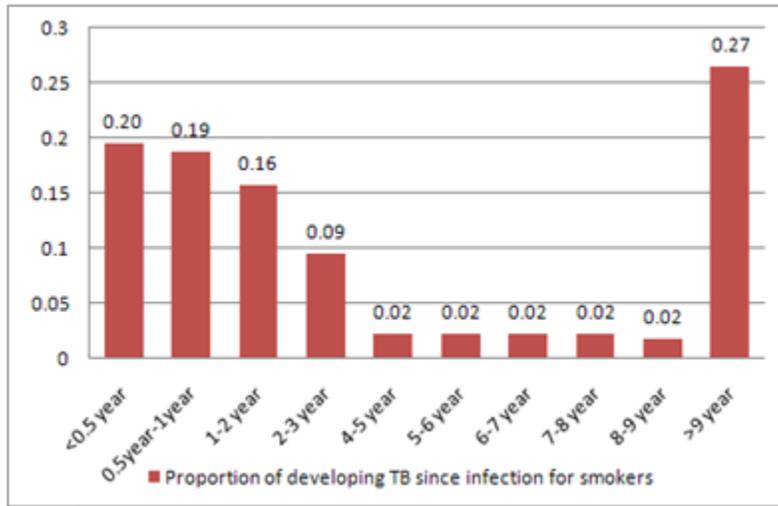


Figure 10: Proportion of smokers developing TB within a certain window of time following infection

In this experiment, the aggregate model assumes a memoryless progression from latent TB infection to Active TB. By contrast, individuals in the agent-based model exhibit a decreasing reactivation rate with rising time since infection. The analysis of time from latent infection to active TB and age structure is important, as it might give some insights into the prevention of TB. Age, often considered as a confounder, can be examined in the individual-based model; among other benefits, such an examination can aid us in finding high risk age groups of individuals who are more susceptible to TB infection. The accessibility of this information within the model could also permit evaluation of policies which explicitly consider the estimated time since an individual's exposure when providing prophylactic treatment.

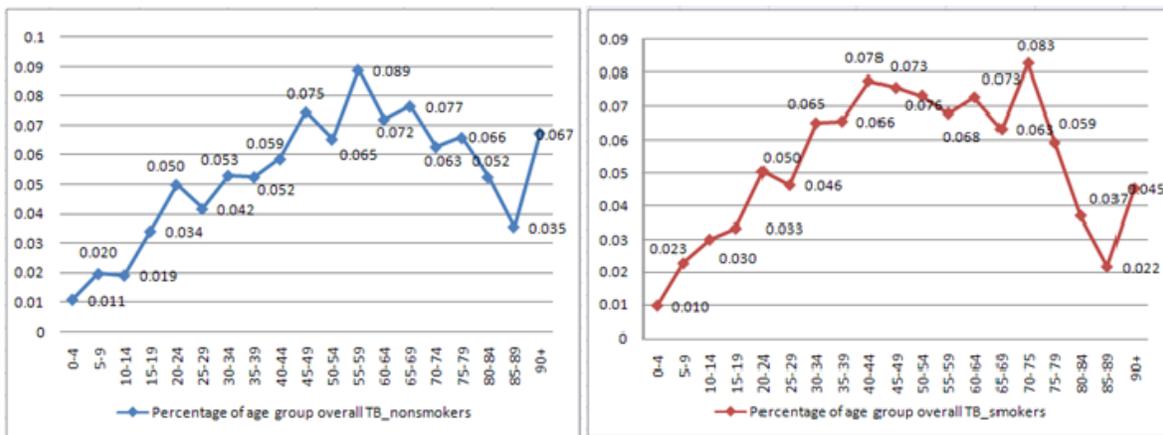


Figure 11: Age structure of non-smoker TB cases and smoker TB cases

Interval From Latent to TB	Cumulative Percentage in AB Model (%)	Cumulative Percentage in 1972 for Non-Indian (%)	Cumulative Percentage in 1972 for Indian (%)
<0.5 year	20.6	31.8	36.4
0.5-1 year	40.0	42.1	50
1-2 years	56.29	52.6	63.6
2-3 years	61.87	78.9	68.2
3-4 year	65.81	84.2	72.2
4-5 year	67.91	89.5	86.4
5-6 year	69.73	-	90.9
6-7 year	71.66	-	95.5
7-8 year	73.55	-	1
8-9 year	74.94	-	1
>9 year	1	1	1

Table 4: Comparison between historical information and estimated results of AB model

Figure 9 and Figure 10 show the time from latent infection to initiation of active TB for both smokers and non-smokers. In these two graphs, it is found that people are more likely develop TB in the first two years after infection. The proportion of TB cases developing TB within the two years following infection is 61% for non-smokers and 55% for smokers.

It is notable that in an aggregate model, it is currently difficult to derive this important individual-level history information in the context of time-varying risks (e.g. associated with reinfection, or due to changes due in delivery of prophylaxis). Moreover, we also have estimated historical data about the interval from latent infection to TB in Saskatchewan. Using the individual-based model provides us the opportunity to use this historical information to calibrate our model and gain confidence that it captures the essentials of TB transmission in Saskatchewan. Table 4 depicts a comparison of agent-based modeling results with historical information from Saskatchewan Anti-TB League Report (Ant, 1972).

In Table 4, although the results of agent-based model are not perfectly consistent with the historic data, parts of results from agent-based model display some consistency. For example, the Interval from 0.5 to 1 year and 1 to 2 year for estimated results in agent-based model are roughly consistent with that for non-Indian in 1972. Furthermore, age is also captured in the agent-based model for this experiment.

Figure 11 shows the age structure of TB cases for both smokers and non-smokers. As exhibited in Figure 11, we can find that non-smokers with age between 55 to 59 account for highest percentage over all the non-smoker TB cases; while smokers within age range (40-44) and (70-74) possess higher percentage over all the smoker TB cases. However, in contrast to the situation for a larger model we have described in the

literature (Osgood et al., 2011), we note that our current model does not capture the higher risk of infection and primary progression for the youngest age categories, so the data shown exhibits significant discrepancies from the historically observed distribution of cases by age.

Through the implementation of age in this experiment is initialized with data on the population age distribution, it yields some useful information about age structure in TB cases, and this experiment also underscores the possibility of integrating the real age structure of the population in Saskatchewan when we conduct the simulation of TB transmission, and calibration to Saskatchewan data on age-specific case rates.

Evaluation of Network Structure on TB Transmission

Within this section, we examine the impact on infection burden of assuming three alternative types of network structure, namely random, scale-free and small world. In order to make them comparable, the average contacts in each network structure are preserved. Here we compare the prevalence of the infection in the individual-based model and the aggregate model.

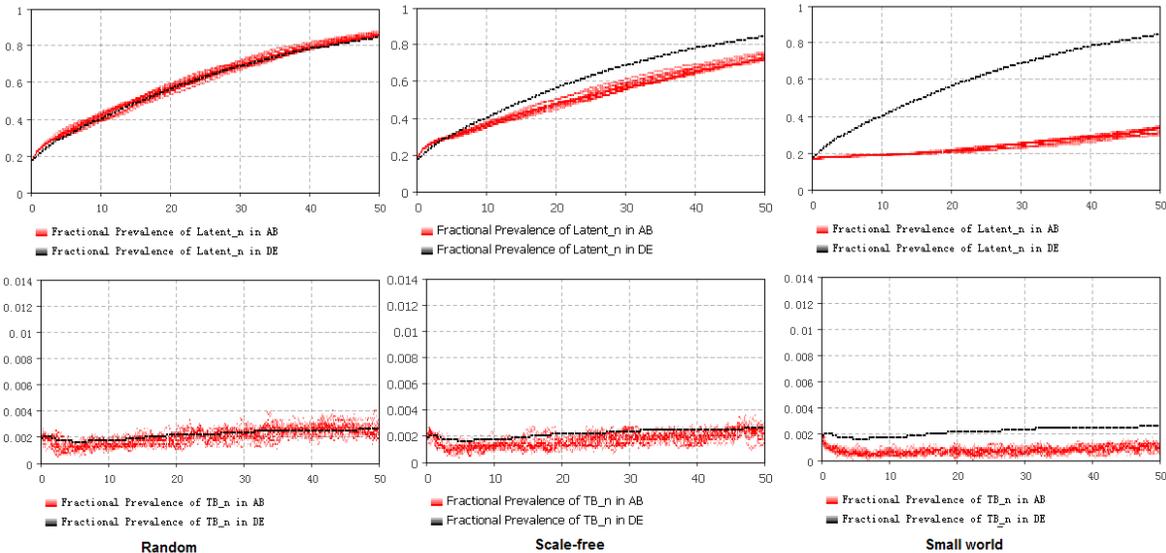


Figure 12: Fractional prevalence of infection of non-smokers under alternative network topology

In Figure 12, we readily find that, among the non-smokers, the random network yields similar results in the aggregate SD model. By contrast, for the scale-free and small world networks, the prevalence of infection is lower than that in the aggregate model. The prevalence of infection in a small world network is even lower than that in the scale-free network. In Figure 13, a random network produces a lower prevalence of latent TB infection among smokers than that emerging from the aggregate model.

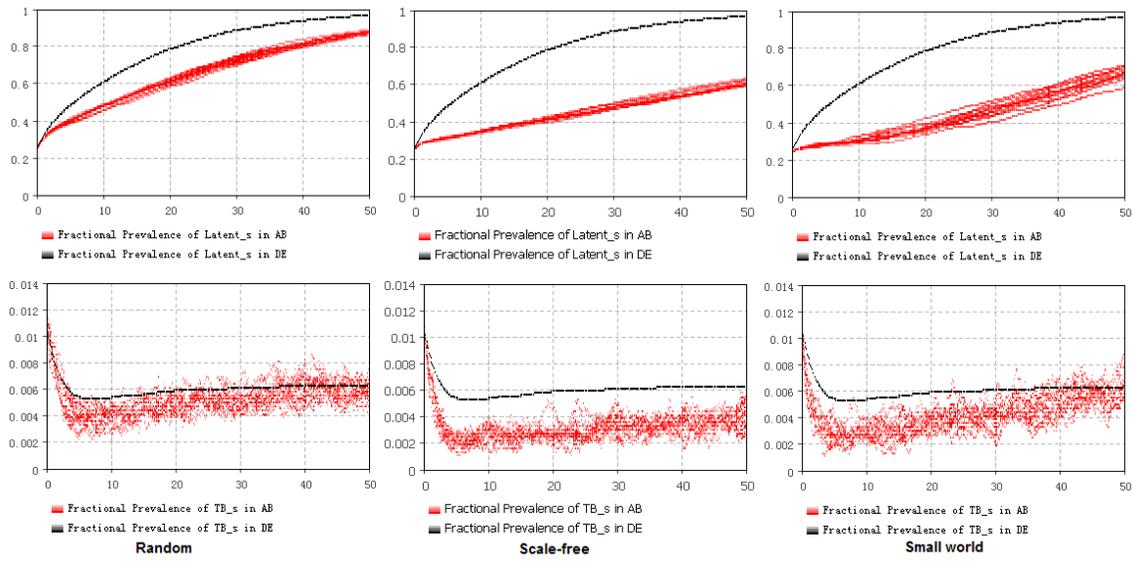


Figure 13: Fractional prevalence of infection of smokers under alternative network topology

The prevalence of TB infection among smokers in scale-free and small world network topologies is much lower than that in random network. The scale-free network produces the lowest prevalence of TB infection among latent TB and active TB for smokers.

From these network experiments, small world network structure and scale free network exhibit the lower level of infection prevalence, while random topology gives highest prevalence of TB infection and Active TB for both smokers and non-smokers. In conclusion, even when maintaining a fixed mean rate of connections, making different assumptions concerning the network types yields a noticeable impact on TB transmission, which should not be overlooked. However the representation in the aggregate model assumes that individuals in the same compartment are perfectly mixed with each other. Based on the results with different types of network, we can conclude that the aggregate model runs the risk of overestimating the burden of infection in the population. From this point of view, the details of elements left implicit in aggregate models (such as network structure) can have a major impact upon model results, and can make a naive parameterized (and uncalibrated) aggregate model diverge in pronounced ways from actual behavior. It is known that social networks are sometimes well approximated by specific network types, such as scale-free or small world. The explicit representation of different network types in an individual-based model can help us produce a more realistic model of the real pattern of network of individuals in TB transmission.

Discussion

After running large volumes of experimental scenarios, network topology and individual heterogeneity are demonstrated to have a significant impact on the dynamics. Based on the conducted comparison between the aggregate and individual-level approaches, which one is better? In reality, we often meet trade-offs between these two modeling approaches. Based on our own experience in modeling the TB transmission in an individual-based level and comparing the difference between these two models, we conclude with a few comments on the trade-offs obtaining between these two methods.

Working at the granularity of individuals, individual-based models can more readily capture diverse attributes of individuals and more flexibly represent more complex processes. In the experiments for BCG and reactivation, individual-based models can easily record and simulate the impact of decreasing BCG protection duration, decreasing reactivation rate over time, the interval between latent and active TB, and age structure.

From the ease of model extension and creation, individual-based models can be extended easier to capture additional components of heterogeneity. When we implement the age structure of each individual, we only need add one more variable associated with a person to represent this status. By contrast, in an aggregate model, you need to modify stocks and flow definitions across the model. Especially when a modeler seeks to implement more attributes, even static attributes (such as gender), the number of compartments in the stocks required rises geometrically; particularly when the attributes (such as age or time since infection) are dynamic, this can lead to very complex, inter-mixed formulas for flows. From this angle, individual-based models are easier to create and more flexible to extend. In addition, the representation of waning of immunity (or other transitions with a similar fashion) can be quite awkward in an aggregate model, since a group of compartments exhibiting different level of declined immunity need to be created. This can be particularly cumbersome when the population represented in the aggregate model already has many attributes (such as age group, ethnicity and gender). For example, suppose we want to implement such waning of immunity with 10 decayed level in an aggregate model with many attributes, 10 compartments under each element of each attribute need to be created, which can end up yielding a huge number of stocks.

Moreover, such a representation exhibits poor separation of concerns (Dijkstra, 1982): the logic needed to achieve progression along this dimension of heterogeneity frequently becomes tangled with the logic associated with progression along other dynamic dimensions of heterogeneity (e.g. age). By contrast, representation of such waning phenomenon in an agent-based model is much easier, it can be accomplished via implementing one more function in the person class instead of adding a large number of compartments.

From the point of view of computational resource demand and speed, individual-based models are typically less effective – and frequently far less effective – than aggregate models. Individual-based models can be time-consuming; for example it takes around 6 hours to run 10 simulations on around 40,000 individuals in our experiment.

But the simulation time for aggregate models are quite short, and can almost be ignored. The simulated population size has a significant impact on the computational trade-offs. When we double the simulated population, the time cost for aggregate models doesn't grow at all; however, the time and memory consumption of the individual-based models grows at least linearly with the population (and potentially non-linearly, depending on memory hierarchy effects, network density, and other considerations). If we want to simulate a larger population, the performance of individual-based models is a big concern. In addition, individual-based models, compared with deterministic aggregate models, require more time to verify its correctness due to its stochasticity and the poor expressiveness of general purpose programming languages exhibit when compared to the domain-specific languages commonly used by System Dynamics packages. Since we have limited resources and time, this can further limit our ability in conducting more sensitivity analysis, interactive model exploration, and additional experiments. Of particular note here is heterogeneity associated with individual history. Looking across pathogens, such history information (such as the duration of time since a contact of a case was exposed, or the history of Active TB in a person) can be of considerable interest when designing interventions. Moreover, such information provides an important source of model-generated data to compare with empirical data during calibration and model validation. While rich history information is readily collected within an individual-based model, it is typically infeasible to maintain more than a modicum of historical information in an aggregate model. This limitation constraints a modeler's options for calibration, as well as the types of interventions that can be investigated.

Networks have an important role in shaping our understanding of infectious disease. The focus on individual-level interactions within a network, rather than the population level dynamics, attempts to address the vitally important processes of the actual infection and disease diffusion. Through the implementation of networks, individual-based models can more accurately simulate and exhibit the association of transmission of infection and the presence of long-term relationships between individuals, and their position within the network. By contrast, aggregate models typically operate under the idealized mixing assumption which might overlook important patterns of TB diffusion. Scenarios with three types of network topologies suggest that small differences in the structure of the network can lead to significant changes in epidemic behaviors which can eventually alter the aggregate spread of infection. In addition, taking network topologies into account allows us to more accurately capture and model several important preventions, including contact tracing, screening program or vaccine; and more sophisticated control policies and different strategies can be tested or simulated in a virtual environment with use of network modeling tools. We particularly note the potential for individual-based models to evaluate policies and protocols which take into account features of the case-contact network collected by contact tracing. However, the need to represent networks – as opposed to mixing matrices – does typically demand that creators of individual-based models offer hypotheses about a range of details that can be conveniently omitted from an aggregate model. A mixing matrix in an aggregate model can readily be created from partial network data, without a need

to reason about the driving factors (in the form of movement patterns or an encompassing network) underlying that contact data. By contrast, reasoning about infection spread across a network on which partial data is available requires that an individual model posit hypotheses regarding the structure of the remaining network. Similarly, in an individual-based model in which contacts are driven by movement patterns, it may be necessary to broaden the model to consider the structure of – and even the driving factors governing – those movement patterns. Such considerations typically need not be considered when building an aggregate model.

The three types of networks discussed here are static – the links between the individuals don't change over time; as a result, the intuitive human relationships elements of breaking and forming new connections are not currently represented. The dynamics of networks are believed to be important in understanding the spread of some pathogens (Morris and Kretzschmar, 1995). Designing networks allowing for changes of connections over time is an ongoing challenge. However some pioneering work in tracking the movement and behavior of individuals in real time using mobile device and GPS to collect contact information between individuals allows approximating more comprehensive network structure and more accurate simulation of the spread of pathogens across a population (Keeling and Eames, 2005; Morris and Kretzschmar, 1995; Hashemian et al., 2010).

References

- “Saskatchewan Anti-TB League Report.”, 1972.
- “Random Networks.”, 2002. http://www-f1.ijs.si/~rudi/sola/Random_Networks.pdf.
- “Northern Saskatchewan Health Indicators Report 2004.”, 2004. <http://www.athabascahealth.ca/images/reports/2004%20Health%20Indicators%20Report%20revision1.pdf>.
- “Scale-free network.”, 2011. http://en.wikipedia.org/wiki/Scale-free_network#cite_note-0.
- “Small world network.”, 2011. http://en.wikipedia.org/wiki/Small-world_network.
- Anderson, R. M., and R. M. May. *Infectious Diseases of Humans: Dynamics and Control*. Oxford University Press, 1991, 1st edition.
- Dijkstra, Edsger W. *Selected Writings on Computing: A Personal Perspective*. Springer, 1982, 1st edition.
- Forrester, J.W. *World Dynamics*. MIT Press, 1971, 2nd edition.

- GAO. “Global Health: Challenges in Improving Infectious Disease Surveillance Systems.”, 2001. <http://www.gao.gov/new.items/d01722.pdf>.
- Hashemian, M., K. Stanley, and N. Osgood. “Flunet: Automated tracking of contacts during flu season.” In *Proceedings of the 6th International workshop on Wireless Network Measurements (WiNMee 2010)*. 2010, 348–353.
- Hassmiller, K.M. “The association between smoking and tuberculosis.” *Salud Publica Mex* 48 (Supple I): (2006) S201–S126.
- Hassmiller, Kristen M. *The Impact of Smoking on Population-Level Tuberculosis Outcomes*. Phd dissertation, University of Michigan, 2007.
- Jacquez, S. “The stochastic SI model with recruitment and deaths - Comparison with the closed SIS model.” *MATHEMATICAL BIOSCIENCES* 117: (1993) 77–125.
- Keeling, Matt J., and K.T.D. Eames. “Networks and epidemic models.” *Journal of the Royal Society Interface* 2, 4: (2005) 295–307.
- Mahamoud, A., Y.Q. Liu, and S. Sarker. “Tuberculosis Modelling Project: Reconstruction of Kristen M. Hassmiller’s Model.”, 2009. CMPT858 report, Department of Computer Science, University of Saskatchewan.
- Meng, A., and N. Osgood. “The System Dynamics Longitudinal Analysis System: Quantifying the Hidden Trajectories of System Dynamics Models.” Austin, Texas: Extended abstract and presentation at System Dynamics Winter Conference, 2011.
- Morris, M., and M. Kretzschmar. “Concurrent partnerships and transmission dynamics in networks.” *Soc. Networks* 17: (1995) 299–318.
- Osgood, N., A. Mahamoud, K. Hassmiller-Lich, Y. Tian, A. Al-Azem, and V. Hoepfner. “Estimating the Relative Impact of Early-Life Infection Exposure on Later-Life Tuberculosis Outcomes in a Canadian Sample.” *Research in Human Development* 8, 1: (2011) 27.
- PHAC. “Tuberculosis FACT SHEETS.”, 2008. <http://www.phac-aspc.gc.ca/tbpc-latb/fa-fi/trans-eng.php>.
- Rahmandad, Hazhir, and John Sterman. “Heterogeneity and Network Structure in the Dynamics of Diffusion: Comparing Agent-based and Differential Equation Models.” *Management Science* 54, 5: (2008) 998–1014.
- Vickers, D.M., and N.D. Osgood. “A unified framework of immunological and epidemiological dynamics for the spread of viral infections in a simple network-based population.” *Theoretical Biology and Medical Modelling* 4: (2007) 49–62.
- Vynnycky, Emilia. *An Investigation of the Transmission Dynamics of M. tuberculosis*. Phd dissertation, London University, 1996.

Watts, D.J. “Collective dynamics of ‘small world’ networks.” *Nature* 393: (1998) 440–442.

WHO. “Tuberculosis (TB).”, 2011. <http://www.who.int/tb/en/>.