AGENT-BASED MODELING AND SYSTEM DYNAMICS MODELING ON TRANSMISSION OF TUBERCULOSIS IN SASKATCHEWAN

A Thesis Submitted to the
College of Graduate Studies and Research
in Partial Fulfillment of the Requirements
for the degree of Master of Science
in the Department of Computer Science
University of Saskatchewan
Saskatoon

By
Yuan Tian

Permission to Use

In presenting this thesis in partial fulfilment of the requirements for a Postgraduate degree from the University of Saskatchewan, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised my thesis work or, in their absence, by the Head of the Department or the Dean of the College in which my thesis work was done. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any material in my thesis.

Requests for permission to copy or to make other use of material in this thesis in whole or part should be addressed to:

Head of the Department of Computer Science
176 Thorvaldson Building
110 Science Place
University of Saskatchewan
Saskatoon, Saskatchewan
Canada
S7N 5C9

Abstract

The desire to better understand the transmission of infectious disease in the real world has motivated the representation of epidemic diffusion processes in the context of qualitative simulation as a computational model on provincial and community levels. In this thesis, we have developed both agent-based models and System Dynamics models within the context of M. Tuberculosis (TB) transmission in Saskatchewan and a community in Saskatchewan to evaluate the efficiency of prevention programs such as contact tracing investigation. New insights about how dynamic models and agent-based models can assist policy development and decision making in disease control will be generated.

Moreover, we sought to compare these two modeling approaches to gain insights in TB diffusion in Saskatchewan as well as guidance in choosing the appropriate modeling approach for particular problems.

Acknowledgements

I would like to express my greatest appreciation to Dr. Nathaniel Osgood for his careful supervision and invaluable support to my research work. It is my great forture having him as my supervisor. During my graduate study, Dr. Osgood suggested me to take 2 additional courses, namely Epidemiology and Biostatistics; and these two courses enriched and reshaped my understanding of my thesis topic a lot and saved me tremendous time. Dr. Osgood's abundant experience in System Dynamics and Agent-based modeling and his truly scientific intuition have inspired me to conduct productive research. I enjoyed discussing with him about various topics, and he gave me many useful suggestions on my research, even my life.

I would like to acknowledge my other committee members, Dr. Vernon Hoeppner, Dr. Kevin G.Stanley and Dr. Christopher Dutchyn, for their helpful suggestions. Dr. Hoeppner gaves me invaluable information, support and suggestions on my research work. I am very grateful for the opportunity to learn from and interact with him. I also would like to thank Dr. Assaad Al-Azem for his help and suggestions in health science and Tuberculosis control.

Finally, I would thank all my family members for their support, their love is the most precious gift in my life.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Motivation

Every year, infectious diseases cause more than 13 million deaths worldwide, with two-thirds of them occuring among children under 5 years old [18]. The top infectious disease killers include Human Immunodeficiency Virus (HIV), Tuberculosis (TB) and malaria [18]. At the same time, we are imperiled by newly emerging and re-emerging infectious disease such as H1N1 (swine flu) and H5N1 (bird flu). 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 modeling, corresponding to nonlinear differential equations (DE), is a type of equation-based modeling which can easily express the cause-effect relationships among variables in the complex systems. Forrester's model of the world, one of the earliest and best known examples, was used to explore outcomes involving different future population sizes. It used variables such as growing pollution and consumption of natural resources, where those variables interacted via causal links and feedback loops [14]. System Dynamics models of infectious disease spread commonly implement structural principles drawn from the most traditional mathematical epidemiology models, which are aggregate in character. 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 [1]. In such aggregate models, individuals are aggregated into larger groups with same abstracted properties. However, there has been a limited amount of System Dynamics modeling used at the individual level [49].

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 need 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 [40]. 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 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 thesis, 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 via using different methodologies of computational modeling.

In spite of the focus on understanding the disease within modeling applications, the need for integrating the prevention strategies and control policies into epidemiological models has long been acknowledged. Both deterministic and stochastic models have been developed to evaluate the impact and efficiency of different control prevention and programs for infectious disease such as

vaccination, screening, targeted treatment and contact tracing [13, 24, 25]. In the course of TB epidemics, although previous work has explored some theoretical aspects of different treatments and prevention [2, 58], they lack detailed representation of the ongoing operational processes of control programs; it is necessary to characterize the dynamics of a combination of several important prevention strategies such as active diagnosis, treatment of Latent TB infection (TLTBI) and contact tracing investigation. In order to provide an illustration of how dynamic models can be used to evaluate and guide operational control strategies, we introduce a System Dynamics model of TB transmission that explores some approaches to integrating infection dynamics with a group of ongoing control policies, most notably contact tracing and TLTBI, across a community in the Canadian province of Saskatchewan. Through sensitivity analysis on contact tracing speed and priorities, our model not only advances the overall theoretical understanding of TB transmission, but also assists in creating operational optimum prevention strategies.

Notable advances in immunology and microbiology have provided fundamental insights into the detailed mechanism of TB infection in the past century. At the same time, epidemiological modeling centered specifically on the dynamics of infections at the population level has profoundly enriched our understanding of the properties of TB and prevention [1, 20]. In spite of the remarkable insights gained from biological research and dynamic modeling, the TB incident rate is still high in certain geographic and demographic zones. In the Canadian province of Saskatchewan, the Saskatchewan Anti-Tuberculosis League and the Provincial TB Control program made historical inroads against TB, but the TB incidence rate in the province remains among the highest across Canada, with the situation remaining especially severe among Aboriginal peoples [37]. It is still challenging to investigate the implications of various risk factors associated with dynamics of TB diffusion. To gain insight into development of optimum intervention policies and how the risk factors and contact structure, at an individual level, and the protocols used for contact tracing will eventually affect the dynamics of TB, we are motivated to create an agent-based model of TB diffusion in Saskatchewan integrated with various control strategies. Contact structure and contact tracing investigation will be further examined in this model to capture the impact of the contact pattern and heterogeneity on infection and how to establish practical and efficient prevention policies.

1.2 Epidemiology of Tuberculosis

In order to facilitate computational model building and enhance the understanding of the aggregate and agent-based models examined in this thesis, we review the basics of epidemiology of Tuberculosis as well as the prevention and control programs in Saskatchewan.

1.2.1 Historical Background and Epidemiological Terminology

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 [45, 54]. Although it is a curable and preventable disease, it is reported that roughly two million people die annually from TB [54]. In Canada, despite the adoption of guidelines and prevention programs, the incidence of TB remains high in certain areas. Saskatchewan is one of the Canadian provinces 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 which include First Nations, Inuit and Métis.

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 most mycobacteria TB transmit via airborne mechanisms. Infection by mycobacteria TB does not automatically bring on TB disease. Usually there is an incubation period before an infected individual physically develops the current disease, and there are individual differences in latency.

- *Active TB Disease.* The term "Active TB" typically refers to current disease; people with advanced disease typically feel sick and may have some known pathologies in parts of their body (such enclosed granulomas in various organs) where the mycobacteria TB cluster.

- *Pulmonary TB.* M. 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 mycobacteria TB when they are coughing, sneezing, singing, and even when just talking [39]. These TB droplets remain in the air for a couple of hours, and people who breath in these TB droplets are exposed to mycobacteria TB.

- *Non-infectious active TB.* Not all active TB cases are infectious, some of them are non-infectious. These non-infectious active TB cases can be pulmonary TB or other forms of active TB (such as TB in bone or brain). Non-infectious active TB cases can not breath out droplets with mycobacteria TB, so they can't transmit the disease via the air.

- *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. This mechanism, in which the mycobacteria evade effective control, is termed as primary progression. The mean time for primary progression varies in different studies. Since TB is a slowly growing infectious disease, commonly

used time limits for primary progression are 2 years or 5 years [20, 50].

- *Latent TB Infection.* A TB infection does not instantly bring on active TB disease; typically there is an incubation period before disease emerges. 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 mycobacteria persists in an immunologically-controlled state. 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 HIV, a combination of complex risk factors (such as age, ethnicity and smoking), interval from infection, a weak immune system or poor health care. Theoretically speaking, in a latently infected person, either the mycobacteria is still alive but inactive in his cells or his immune system might completely kill the mycobacteria. However, it is currently impossible to differentiate between them with readily available diagnostic technologies. Latent TB cases with inactive mycobacteria can develop the disease later on in life through re-activation; on the other hand, for those with killed mycobacteria, 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 age, ethnicity, 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 mycobacteria via a mechanism often referred as reinfection. There is significant controversy regarding the level of reinfection that occurs within the population. Some literature estimate that 25% of the latently infected people is at risk of exogenous re-infection [20].

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.

1.2.2 Tuberculosis Prevention and Control

TB, as a serious contagious disease, is the second largest cause of death from contagious disease all over the world [20]. It is estimated that 30% of the world's population is infected with TB [56], and the incidence rate of TB in Saskatchewan is relatively higher than that in other regions in Canada [35]. Following the recommendation provided by World Health Organization in 1997,

the Tuberculosis Prevention and Control Unit of Health Canada have implemented guidelines in an attempt to eliminate Tuberculosis [3]. Here, we present a list of adopted tests, prevention and control programs which are known as effective strategies and investigation in the battle against TB.

- *Bacillus Calmette-Gurin (BCG).* BCG is a vaccine against Tuberculosis (TB). However, the duration and efficiency of the protection given by BCG are not clearly known, and the estimates of its protection are controversial. One study shows an efficiency of 84% for 5 years after immunization, and the protective efficacy declines over time [19]. However, another study shows that BCG contributes only an efficacy of 14% in reducing TB, and it appears to be less effective among those most in need [9].

- *Active TB Treatment.* Normally patients with active TB are treated for 6 to 9 months unless they are resistant to the medication. An another therapy regimen named directly-observed treatment short-course (DOTS) has been widely adopted in Canada and worldwide due to its efficiency in elevating compliance rates and lowering the risk of drug resistance [55].

- *Treatment for Latent TB Infection (TLTBI).* The target group of this therapy is recently infected TB individuals with no current disease. If eligible for treatment, they are treated for 6 months. This treatment seeks to give protection to those infected people in the latent stage who might possess a higher risk of developing active TB.

- *Screening.* The primary purpose of Screening programs is designed to detect potential infected individuals as early as possible. In Saskatchewan, screening targets at pre-school and school children on reserve.

- *Contact Tracing Investigation.* Contact tracing, as a form of target-oriented control towards potential next-generation cases, has been used in preventing infectious disease, including sexually transmitted disease, TB and measles. Under the context of TB control, the main purpose of contact tracing is the early detection of infectious TB cases and recently infected persons who may have a higher risk of developing the disease. Contact tracing focuses on two types of persons, namely infectious TB cases and primary TB cases [43]. Contacts are identified and located on the basis of information provided by the active TB cases via interview.

- *Mantoux Test.* As part of TB investigation, a Mantoux test is used to check whether or not a client is infected. Positive result suggests that the client is infected by mycobacteria TB, otherwise the risk of active TB disease is ruled out. In contact tracing and screening, the Mantoux test is used to check the status of an individual's TB infection.

- *Chest X-Ray.* A chest X-Ray is one of the tests used in active TB diagnosis.

As listed above, remarkable prevention efforts in TB control have been made to ensure prompt detection, effective treatment and airborne precautions. Given the relatively high rate of TB within Saskatchewan, policies and procedures for TB control should be continuously developed and evaluated for effectiveness to yield further insight into actions effective in minimizing the risk for TB transmission.

1.3 An Overview of Agent-based Modeling

An agent-based model (ABM), a form of computational model, is a multi-agent system composed of a number of interacting agents in order to satisfy their goals within an environment [17]. It often simulates in a simplified fashion of the processes that are thought to exist and operate in the real world. The distinct feature of agent-based model is that agents are represented explicitly to interact with one another through passing of messages or transferring of data in a network. Agents, as discrete social actors with their own behaviors, possess the capability to react to the computational environment. Agents are also frequently capable of moving and have the ability to record their historic states. Given the status, history or environment of agents, agents can achieve sophisticated behavior following a set of incorporated rules, events or strategies [17]. Conventionally, agents possess the following properties [57]

- *Autonomy*. There are some built-in states within the agents which allow them to make decisions according their current state. They are capable to learn and adapting their actions based on experience. State charts are one of the forms used in representing their previous or current states as well as the transitions from state to state.

- *Social Ability*. This term refers to the interaction among agents via some kind of agent-communication language. In programming terms, it means agents can send messages to and receive messages from other agents.

- *Reactivity*. Agents are placed in an environment, they can perceive their environment and some other agents in their neighborhood.

- *Pro-activeness*. Goal-directed behaviors is another feature of agents. They frequently have goals to pursue on their own initiative, and undertake actions so as to attain those goals.

Agent-based modeling, also known as individual-based modeling, has been used in many application and fields, including but not limited to air traffic control, business process management, consumer behavior investigation and diffusion of epidemics. In this thesis, we have applied agent-based modeling techniques in representing the transmission of infectious disease. The agent-based models in this study use agents to represent people with specific social status and different levels of

risk in developing Tuberculosis. Different scenarios are associated with different model designs and different representation of epidemics, resulting in different implementations of the environments in the models, contact patterns of individuals and characteristics of agents. For example, in one scenario, individuals are allocated in a specific type of network structure and infections are performed through direct contacts and messages passing from one infectious individual to another individual within the network. In another scenario, individuals have their own states and internal likelihood of getting infected or developing the disease. Infection are not transmitted by direct contacts in this design, however the risk of developing disease depends on individuals' own status and several global variables which capture the overall situation of the epidemics of the disease.

1.4 An Overview of System Dynamics Modeling

System Dynamics (SD) modeling is an approach to understand dynamic behaviors and mechanism of complex systems over time, so as to better manage those systems. System Dynamics models often incorporate elements including causal forces, time delays, feedbacks, interactions and non-linear relationships. The heart of System Dynamics approach are feedback and causal loops which provide a platform to capture the circular causal effects and conceptualize the structure of a complex system. The complex behaviors usually arise from interactions of two types of feedback loops, namely reinforcing (or positive feedback) loops and balancing (or negative feedback) loops [44]. Reinforcing loops often amplify or accelerate divergence in behavior, while balancing loops counteract changes and tend towards balance and equilibrium. Stocks and flows are employed in illustrating the structure of a dynamic system (containing such feedback loops within their structure). And mathematically speaking, a formal System Dynamics model is a combination of non-linear differential equations represented via stocks and flows.

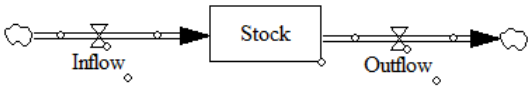


Figure 1.1: General structure of stock and flow diagramming notation [44]

Under the context of System Dynamics modeling, real world systems can be described in terms of continuous and connected quantities (stocks and flows) within loops of causal feedback, and applicable insights and model-based understanding can be examined and derived from such stock-and-flow and causal feedback structure of the system [53]. Figure 1.1 gives a diagrammatic representation of stocks and flows; while in terms of mathematics, the differential equation in Figure 1.1 can be represented as equation (1.1) [44]:

$$\frac{\mathrm{d}\, stock(t)}{\mathrm{d}t} = Inflows(t) - outflows(t) \tag{1.1}$$

In the past decades, System Dynamics modeling approach has widespread applications on many domains such as management, economics, public policy, ecology, epidemiolgy, and so on. For the purpose of understanding the nature of complex systems and making reasonable and sound decision, it has been applied in solving dynamic problems arising from corporate strategy to the dynamics of ecological system, from the supply chain management to the battle against infectious disease [44].

1.5 Thesis Statement

In this thesis, we seeks to explore the impact of heterogeneity and model architecture on TB outcomes via comparing both aggregate and individual-based TB transmission models. In addition, given the high TB incidence rate among First Nation people in Saskatchewan, we use both aggregate and agent-based models to evaluate current contact tracing investigation, and to suggest more effective contact tracing practice in TB control in Saskatchewan.

1.6 Thesis Organization

The remainder of the thesis is arranged as follows. Chapter 2 reviews related research on mathematical models of infectious disease, agent-based modeling, networks and epidemic models, and effective active TB control. Chapter 3 compares an aggregate model with an agent-based model in the context of TB diffusion with heterogeneous individuals and network structures evaluation, and some new insights in methodological aspects are proposed. Chapter 4 presents a System Dynamics TB model integrating operational TB control programs in a community of Saskatchewan. An network-based model of TB transmission for Canadian province of Saskatchewan is developed and simulated in Chapter 5, and a number of scenarios regarding different protocols of contact tracing are investigated. Chapter 6 concludes summary, contribution and future work of our studies.

Chapter 2

Literature Review

Mathematical models of disease transmission within human populations have been acknowledged in helping policy makers and epidemiologists interpret epidemiological trends, understand the dynamics of disease spread and measure the efficiency of disease prevention and control, such as measles, HIV and other emerging infections [1].

Starting in the 1920s, susceptible-infectious-recovered (SIR) models and variants like susceptible-exposed-infectious-recovery (SEIR) models were introduced and helped establish the foundations of much of mathematical epidemiology. This basic SIR model consists of three compartments of individuals who 1) are susceptible to, 2) have been infected by and are infectious with or 3) get recovered from a particular contagious disease. An important derived quantity in the SIR model is the force of infection which denotes the dynamic rate at which susceptible individuals are infected [1]. Force of infection (λ), representing the "infection pressure" resulting from the interaction of people within a population, is a function of the number of infectious people (I), the population size (N), transmission rate (β) and mean contacts per person per unit time (c). When the population is randomly mixed, the simplest framework treats the force of infection as $\lambda = \beta c \frac{I}{N}$. Although differences in susceptibility to infection and variation in contact patterns between heterogeneous individuals are neglected within this simplified context, the SIR model has exerted a profound and persistent influence on modeling of infectious diseases. In contrast to the SIR model (which offers a description better suited infectious diseases conferring lifelong immunity, such as smallpox and many childhood infections), other models such as the Susceptible-Infected-Susceptible (SIS) and Susceptible-Infected-Recovered-Susceptible (SIRS) are better characterizations of infectious diseases where repeated infections are commonly observed and long-standing immunity is not conferred, such as several sexually transmitted diseases (e.g. Chlamydia, Gonorrhea), influenza, as well as some infections with relatively rapid waning of immunity (e.g. pertussis) [15].

Some modeling studies enrich the basic model framework with heterogeneities by subdividing subgroups, so as to generate greater and more realistic structure [16, 20, 37]. Such disaggregation (which uses the technique of attribute-based disaggregation [36]) can be used to stratify a model to reflect more complex hierarchy of population or to integrate personal characteristics to generate rich dynamical behaviors. To highlight the variations in mixing patterns (e.g. people who speak the

same birth tongue are inclined to mix more readily among themselves), a "mixing matrix" is used to adapt the simplified formulation of the force of infection introduced above. This mixing matrix describes contact coefficients among different groups. The Tuberculosis (TB) model in [37] has integrated such transmission preference to express the mixing preferences of different age groups and ethnic categories. In addition, Hassmiller's TB model [20] with smoking impact evaluation also stratify people into subgroups regarding smoking status and apply specific mixing patterns.

The methodology of agent-based modeling is now widely used in a variety of social science disciplines, including psychology, public health [16], and political science. As a powerful simulation modeling tool, a number of real-world applications (e.g. market simulation and diffusion simulation) are developed in reproducing the natural emergent phenomena of a system (e.g. traffic jams, or a stock market) and generating explanation for social or other observed phenomena [6]. We discuss each of these below.

There is an upsurge of interest in using agent-based modeling to simulate markets in recent years; one of the pioneering commercial application is the agent-based NASDAQ stock market model [10]. In the agent-based NASDAQ model, market participants, market institutions, market rules and their interactions are simulated in a way approximating the processes operating in the real-world market: investor agents are capable of buying or selling shares, following a variety of strategies from simple to complicated ones involving learning. Due to the capability of agent-based modeling to represent the system from the perspective of heterogeneous individual behaviors and real-world activities other than the abstract processes or averages, the NASDAQ model can simulate the impact of changes on the financial market under different circumstances, provide warning of unexpected outcome of new strategies in advance, and monitor behavior of agents in response to different implemented regulations [6, 10].

Diffusion often represents the cases where people are influenced by others around them, and it is a fundamental process observed in diverse psychological, social, and economic circumstances. Agent-based modeling can be used to describe many diffusion phenomena in human systems, such as product adoption and disease diffusion models. Use of agent-based modeling in health science has gained momentum in recent years [29, 34, 38]. For the purpose of exploring the propagation of communicable disease through a defined population, an agent-based modeling approach is proposed and used to simulate the spread of disease in an urban environment with geographic information system (GIS) and spacial network integration [38]. In this work, a measles outbreak, as a case study, is implemented in the agent-based model within a closed population where interactions among individuals are associated with places and mobility of agents are encapsulated in a transportation network. A variety of scenarios of an outbreak are carried out in this study to answer "what-if" questions in a simulated geographical environment.

In addition, constant network structure, through which the disease diffuses, can be well cap-

tured in agent-based modeling techniques. In a world where persistent patterns of contact are an important feature of many human institutions, the ability to reason about such representations enriches our understanding of determinants of specific epidemic patterns and better captures the intervention trade-offs. In contrast to aggregate models in which it is assumed to be safe to assume an approximation of a "fully-mixed population", agent-based models are flexible in simulating disease transmission on networks. Given a network structure of concern, integrating such network representations into a model is relatively straightforward in addressing social connections and characteristics [28]. A large volume of research has studied network topologies and their impact on various processes in the spread of disease [34]. Usually the simulated network is defined in terms of an individual's distribution in space or the formation under which their connections are established. Reflecting the fact that different infections are passed via different pathways, the mixing network should be pathway specific to satisfy the context of a particular communicable disease [28]. Within a network structure used for disease diffusion, levels of susceptibility or relative risk can be introduced for individuals according to different characteristics such as age, ethnicity or gender; variations around some average level of rates of infection can also be applied to individuals even in the same group. Given such incorporation at individual level, history information is typically far easier to capture; in addition, such information can be further used in model calibration and accessed to build adaptive policies (which change based on the characteristics of the person involved) [31].

Beyond merely stably formed network structures, some studies seek to represent networks as emergent properties of agent movement. One study captures human movement patterns and data (including agents, their routine movements and locations) to derive the transmission networks [47].

Stimulated and supported by increasing computational power, agent-based modeling offers particular attraction due to its ability to represent more complex behavior patterns (e.g. those depending on individual history and learning) and in capturing the dynamics of real world systems in a natural way. However, in light of the fact that traditional aggregate compartmental modeling (such as aggregate System Dynamics modeling or other models based around differential equations) have more established history and a larger volume of past applications, much questioning about the trade-offs involved in agent-based modeling has taken place in recent years. Sterman et al.[40] compared agent-based models with differential equation models in the context of a SEIR disease model. In their work, experiments involving disease diffusion are carried out within the stochastic agent-based SEIR models and the classic SEIR model, and the implications for choice of model types are discussed in detail. The assumption of thorough mixing and homogeneity in classical SEIR models are relaxed in the agent-based version by integrating network topologies and heterogeneous individuals, and the outcome of simulations demonstrated that certain dynamics or behavior patterns that emerged in agent-based model are distinctly different from those observed

in differential equation models. Interactions of individuals can generate network effects which eventually lead to significant deviations from the predicted trends in a naively parameterized aggregate level. The model architecture also impacts how model assumptions affect our evaluation of reality and the perceived optimality of interventions or policies. In addition, the results suggest that the granularity and heterogeneity of individuals' characteristics is poorly captured or analyzed in aggregate models, and some dynamics involving heterogeneity are infeasible to generate or reproduce in the naively parameterized deterministic differential equation model.

Another difference between the two model types is that, for a given scenario involving a specific model parameterization, the stochastic agent-based model gives a distribution of outcomes while the deterministic differential equation model generates only one trajectory to represent the epidemiological pattern under the "mean-field approximation" [40]. For a nonlinear system, applying the system to the mean of a distribution can generate very different results than what is obtained when taking the mean of the distribution resulting from applying the system to elements drawn from the distribution. For policymakers working with non-linear systems, the averages obtained in an aggregate deterministic model do not always provide a close approximation to the mean of the ensemble of realizations associated with an individual-based model which eventually might lead to a biased estimation of the real situation [6]. Capturing outcome variability in an agent-based model offers extra flexibility in analyzing extreme cases, in the application of explicit risk preferences, and in understanding the degree of variability between cases (e.g. the degree of variability in system results that are associated with each of a set of interventions). Another study conducted by S. Wagar et al. compared these two modeling methodologies in a SIR model, it is found that the population size can trigger different simulation outcomes in these two models [27]. It bears emphasis that for small populations, continuous approximations to counts of individuals sharing certain characteristics can lead to very different results than obtain when considering discrete individuals. (For example, if the discrete nature of infectious individuals are captured, an infection can much more readily go extinct than if the count of infectious individuals is treated as a continuous quantity).

After the worldwide emergence of the SARS epidemic, some epidemiologists have been aware of the public-health importance of understanding the source of infection and transmission networks [41]. And the process of infection tracing (contact tracing) is recognized as an integral part of disease control programs. Most notably, contact tracing gives the possibility reaching into those potential high risk contacts, identifying active cases among them before they spread the infection widely, and providing treatments to those traced cases – eventually leading to lower incidence and prevalence of disease.

Many epidemiological models are developed in investigating programs related to disease control (such as contact tracing) to generate effective preventive measures in order to overcome a variety

of communicable disease including STDs, HIV and some airborne infections such as SARS and Tuberculosis (TB). One disease model integrating a representation of contact tracing was built for a randomly interacting population in [33]. In their work, critical thresholds principles and likelihood of an outbreak are derived and analyzed. This work also points out that a deterministic model approximating the stochastic tracing process can be developed in investigating the dynamics of disease diffusion from the perspective of contact tracing [33]. In contrast to depicting the tracing procedure in terms of aggregate models, other studies emphasize the use of contact network (containing the transmission pathways), in representing the nature of human interaction in a more accurate way [13]. Pairwise-approximation methods and fully stochastic simulations are proposed in [13] to estimate the utility of contact tracing in SIR and SIS models in terms of sexually transmitted disease (STDs), and relationship between efficiency of contact tracing and basic reproductive ratio (R_0) of disease is observed and analyzed in the simulations. Combining computer-generated clustered networks within the model, it is found that clustering accounts for the destruction of possible relationships and a lower requirement for contact tracing efficiency in contrasted with that predicted [13]. In another study, two HIV transmission models with control and prevention measures are formulated based on a differential infectivity (DI) model and a staged-progression (SP) model to evaluate the effectiveness of contact tracing and random screening [25]. Reproductive number and endemic equilibrium are derived in estimating the impact of different levels of intervention programs. The effectiveness of random screening and contact tracing varies between these two models which reminds us that the underlying etiology of the disease transmission cannot be neglected when measuring the efficiency of prevention programs. The contribution also points out the possibility of integrating cost as another measure in such evaluation, as well as conducting comparison with agent-based models within the same context but using different assumptions.

In addition to the objective of identifying new active cases, contact tracing investigation can also work as a complementary tool targeting persons with recent infections but no current disease. To reduce the risk of developing active TB especially among those within their first 2 years of infection (who are at particular risk of progression to Active TB), a mathematical model of the TB epidemic is used to quantify the effectiveness of treatment for early latent TB infection (TLTBI) [58]. Positive effects of TLTBI are observed in lowering the incidence of TB and eliminating the disease, which suggests that targeted preventive therapy for newly infected contacts through contact tracing investigation may ultimately offer great contribution in TB control. Another TB model including preventive treatment for Latent TB infection produces similar outcome and confirms the effectiveness of contact tracing in decreasing the incidence rate of TB [2].

Besides the usage of traditional modeling tools (such as differential equations models) in understanding the disease dynamics and contact tracing program, a variety of efforts have been made in investigating disease transmission in conjunction with contact tracing in social networks. Never-

theless; there are many critical issues about simulating contact tracing in networks such as purely relying on self-reported manner of personal relationships, expensive process of data collection, idealized assumptions of transmission network topology, difficulty in representing the dynamics of human interactions including breaking and forming and challenges in depicting the strength of relationships in a network (beyond merely dichotomous categories of present or absent) [28].

For the past century, much effort has been adapted in controlling and eliminating contagious diseases including Tuberculosis (TB), measles, HIV and so on. For TB in particular, a variety of mathematical models have been extended and simulated to capture epidemiological trends of TB, understand TB dynamics as well as explore various "what-if" questions to optimize the ongoing policies and prevention strategies [20, 50, 37]. To capture risk factors associated with TB, Hassmiller has adapted one TB model in conjunction with the impact of smoking to capture the Indian TB epidemiological context; a model with TB transmission was extended by adding smoking status stratification via disaggregating the stages of TB progression [20]. In another study, an age-stratified deterministic model describing TB in England and Wales since 1900 is well established and calibrated to estimate the relative risk of primary progression, reinfection and reactivation for people in many age groups [50]. It is found that the age and calendar year at infection have a distinct impact on lifelong risks of developing TB. Recently, Osgood and Mohamoud et al. [37] have extended their aggregate TB models by incorporating age as well as ethnic stratification to fit TB data from the Canadian province of Saskatchewan and to investigate targeted intervention strategies for high risk subgroups and their impact on lifetime TB outcome. It is observed that a temporary elevation in incidence rate can bring notable influence on individuals' life long risk of TB, and it indicates the presence of system memory in the form of latently infected population.

Chapter 3

Comparison Between Individual-based and Aggregate Models under context of Tuberculosis Transmission

Both individual-based models and aggregate models (such as the classical focus on methodology of System Dynamics) are widely used in epidemiological modeling. In recent years, many researchers have been interested in the pros and cons of these two modeling approaches. In this chapter, an aggregate System Dynamics model and an individual-based model involving TB transmission with smoking as a risk factor are compared using controlled experimental simulations, and evaluation of these two models will be presented. This chapter is submitted and published in the Proceedings of the 29th International Conference of the System Dynamics Society [46].

3.1 Background

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 [22]. Hassmiller [20] evaluates the situation of TB transmission in India, and provides some important insights into smoking effects on TB diffusion using a compartmental model. Mahamoud et al. recreated Hassmiller's model using System Dynamics modeling and evaluated the impact of smoking on TB diffusion in Canadian province of Saskatchewan [30]. At a population level, Mahamoud et al.'s model gives some important insights into the impact of smoking on TB transmission in Saskatchewan. However, some important characteristics, such as contact patterns and the diversity of individual properties, can't be easily investigated and simulated at an aggregate level.

Seeking to shed light on the impact of aggregation on TB model results, we sought to recreate their aggregate TB model at an individual level. We then designed simulations to compare the results of these two modeling approaches. Through this process, we aimed not only to imple-

ment important characteristics of TB diffusion at individual level, but also to analyze the network structure of TB diffusion and individual heterogeneity to gain more insights into dynamics of TB transmission in Saskatchewan.

3.2 Stucture of Tuberculosis Transmission Models with Smoking Impact

3.2.1 Structure of System Dynamics Model

Based on the scheme in [20], 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.

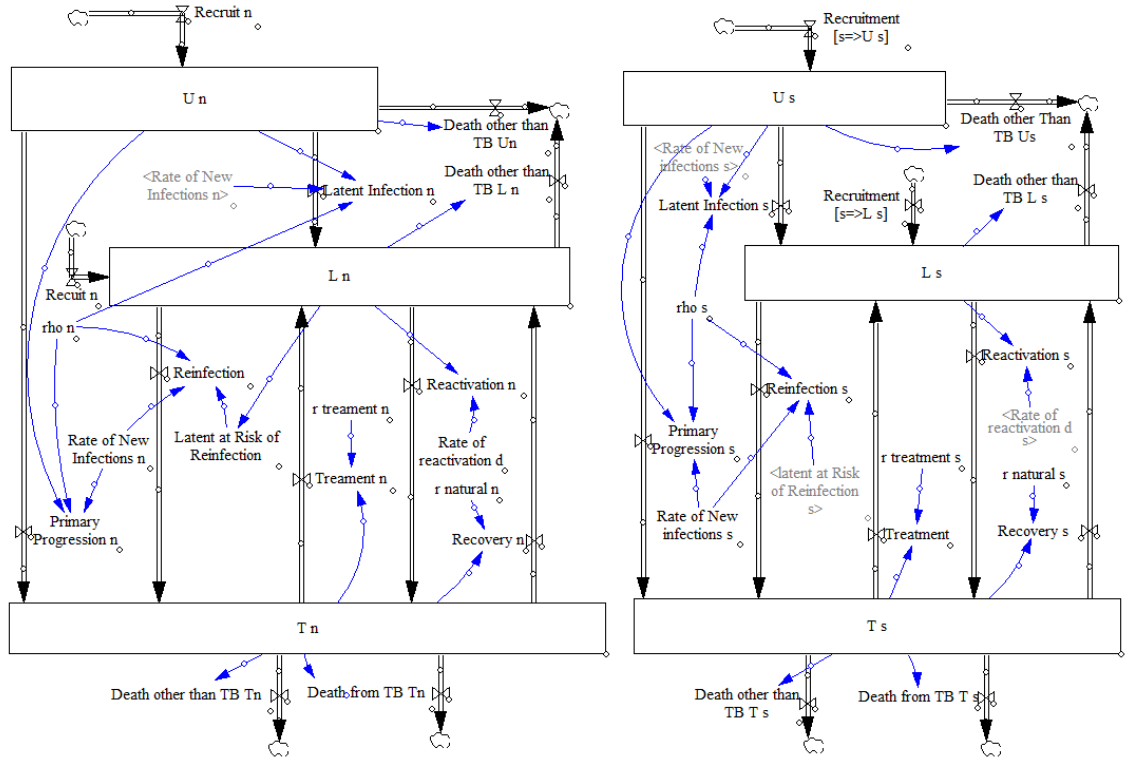


Figure 3.1: A Schematic Representation of Mahamoud et al.'s Aggregate Model of TB Diffusion with Smoking Impact

A simplified version of the structure of Mahamoud et al.'s model is illustrated in Figure 3.1. Stocks, depicting states or accumulations, are represented by the rectangles. Flows, which cause changes to the stocks over time, are shown as solid arrows. Within this model, all the population are categorized into 6 stocks distinguishing people according to both TB and smoking status (seen in Figure 3.1): 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 3.1 and Table 3.2 [30].

Table 3.1: Description of the symbols and parameter settings in Mahamoud et al.'s model [20, 30]

Parameter	**Description**	**Value**	**Unit**
βc	β is defined as the probability of being infected given the exposure, and c is the average number of contacts per TB case per year. βc gives the average number of infections an individual with active TB (T) causes per year.	7.788	persons per year
ρ	Proportion of newly infected individuals progressing to primary TB	0.05	1
λ	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)	3.125×10^{-3}	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

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 [35]. The relative risk for smokers progressing to active TB is much higher than that for non-smokers [20]. 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 3.2. 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.

Table 3.2: Smoking related parameters in Mahamoud et al.'s model [30]

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 disassortivity between smokers and non-smokers	0.3
σ_9	Relative risk of non-TB death rate given smoking exposure	1.14

The equations for non-smoker related stocks and flows illustrated in Figure 3.1 are as follows:

$$\lambda_n = p\beta c\frac{T_n}{N_n}[\sigma_8 + (1-\sigma_8)\frac{N_n}{N}] + p\beta c\sigma_6(1-\sigma_8)\frac{T_s}{N} \tag{3.1}$$

$$\frac{\mathrm{d}\,U_n}{\mathrm{d}t} = (1-\sigma_0)(1-\alpha)\pi - \lambda_n U_n - \mu_n U_n \tag{3.2}$$

$$\frac{\mathrm{d}\,L_n}{\mathrm{d}t} = (1-\sigma_0)\alpha\pi + (1-\rho)\lambda_n U_n + \gamma T_n + \tau T_n - e\rho\lambda_n L_n - dL_n - \mu_n L_n \tag{3.3}$$

$$\frac{\mathrm{d}\,T_n}{\mathrm{d}t} = \rho\lambda_n U_n + e\rho\lambda_n L_n + dL_n - \tau T_n - \gamma T_n - \mu_n T_n - \mu_{tbn} T_n. \tag{3.4}$$

Here N_n denotes the sum of non-smokers in the population where $N_n = U_n + L_n + T_n$, while 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 and $N = N_n + N_s$, and R_n is the rate of new infection for non-smokers. The initial values for the stocks [30] are $U_n(0) = 11429$, $L_n(0) = 2211$, and $T_n(0) = 24$.

The equations for smokers demonstrated in Figure 3.1 are:

$$\lambda_s = \sigma_1 p \sigma_6 \beta c \frac{T_s}{N_s}[\sigma_8 + (1-\sigma_8)\frac{N_s}{N}] + \sigma_1 p \beta c (1-\sigma_8)\frac{T_n}{N} \quad (3.5)$$

$$\frac{\mathrm{d}\, U_s}{\mathrm{d}t} = \sigma_0(1-\alpha)\pi - \lambda_s U_s - \sigma_9 \mu_n U_s \quad (3.6)$$

$$\frac{\mathrm{d}\, L_s}{\mathrm{d}t} = \sigma_0 \alpha \pi + (1-\sigma_3\rho)\lambda_s U_s + \sigma_7 \gamma T_s + \sigma_4 \tau T_s - e\sigma_3 \rho \lambda_s L_s - \sigma_2 d L_s - \sigma_9 \mu_n L_s \quad (3.7)$$

$$\frac{\mathrm{d}\, T_s}{\mathrm{d}t} = \sigma_3 \rho \lambda_s U_s + e\sigma_3 \rho \lambda_s L_s + \sigma_2 d L_s - \sigma_4 \tau T_s - \sigma_7 \gamma T_s - \sigma_9 \mu_n T_s - \sigma_5 \mu_{tbn} T_s. \quad (3.8)$$

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 (3.1) and (3.5), 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.

3.2.2 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. Object-oriented design is applied when recreating Mahamoud et al.'s aggregate model in AnyLogic, attributes and behaviors of each individual are identified and implemented.

Each individual's characters and behaviors are controlled by two state charts shown in Figure 3.2. One state chart represents individual progression of TB infection; the other is used to represent the smoking status of each individual. In the TBStatus state chart, a self-transition from "Uninfected" to "Uninfected" state (the same for "LatentTB" and "ActiveTB" states) represents the process of re-entry the same state with a certain timeout. Such self-transitions help update the dynamic rates of an particular 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

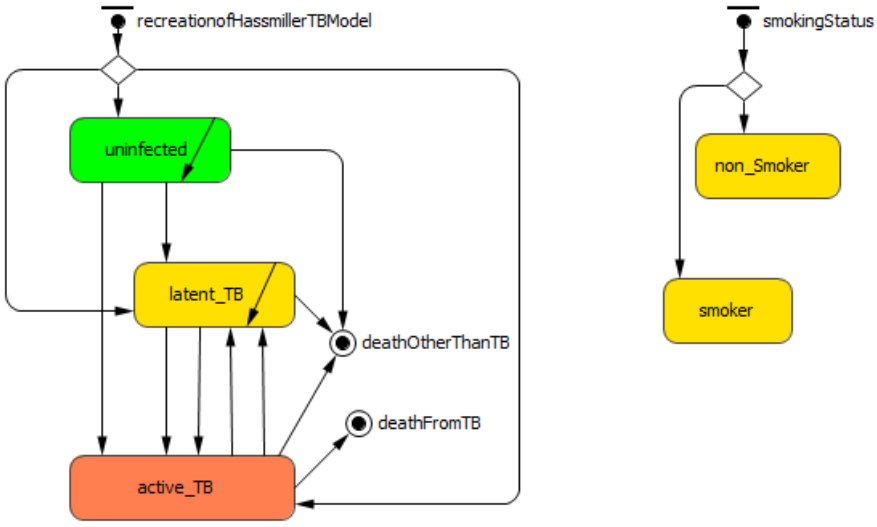


Figure 3.2: Structure of individual-based Mahamoud et al.'s Model of TB Diffusion with Smoking Impact

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" [11] 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 which 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.

3.3 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.

3.3.1 Verification of Individual-based Baseline 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. Given the stochastic features of the individual-based model, we need to conduct many simulations in the individual-based model with the same parameters value and transitions rates as those in the aggregate model to investigate whether the results of these 2 models are the analogous. This version of the model, referred to as the "Individual-based Baseline Model", reflects an individual-based model which is comparable with the aggregate model since identical values of parameters and rates, same logics and dynamic behaviors as well as analogous results are shared.

3.3.2 Individual Heterogeneity with Respect to BCG Vaccination

BCG, as a vaccine against TB, provides protection to people. In our model, we assume that it gives 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.

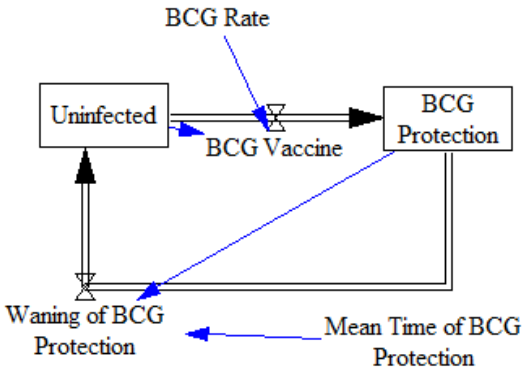


Figure 3.3: BCG Implementation in the aggregate TB model

Figure 3.3 shows the simplified implementation of BCG in the aggregate model. The state equations of implementation for BCG in the aggregate model shown in Figure 3.3 are

$$\frac{\mathrm{d}\,U}{\mathrm{d}t} = \frac{B}{m} - kU \tag{3.9}$$

$$\frac{\mathrm{d}\,B}{\mathrm{d}t} = kU - \frac{B}{m}. \tag{3.10}$$

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.

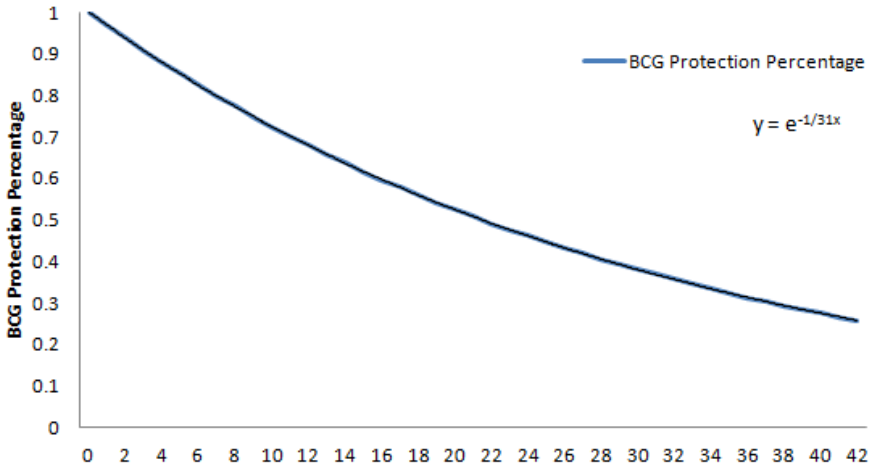


Figure 3.4: BCG Protection Percentage Over Time

However, the assumption that individuals are fully protected for 31 years is not completely

reasonable, some reviews shows that the efficacy of BCG wanes like many other vaccines [19]. 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 3.4, 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 (3.11), 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 (3.12) and (3.13) are used to represent their risk of getting infected.

$$y = e^{-\frac{1}{31}t} \tag{3.11}$$

$$\lambda_{n,b} = (1-y)\lambda_n \tag{3.12}$$

$$\lambda_{s,b} = (1-y)\lambda_s. \tag{3.13}$$

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 sought to investigation the impact of different representations of BCG protection on the TB outcome, 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.

3.3.3 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, we separately

extend the individual-based baseline model. The new structure of the model is similar to that described by Vynnycky [50], and is shown in Figure 3.5.

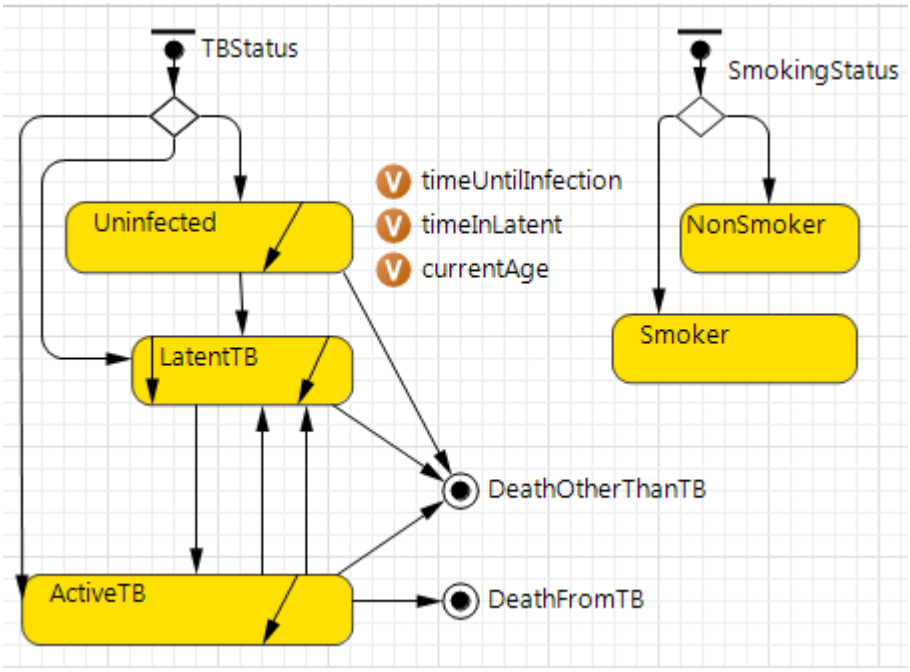


Figure 3.5: Revised Individual-based Model Structure with Respect to Reactivation

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 [50]. 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 3.125×10^{-3} per year, while that for smokers ($\sigma_2 d$) is 4.7×10^{-3} 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.

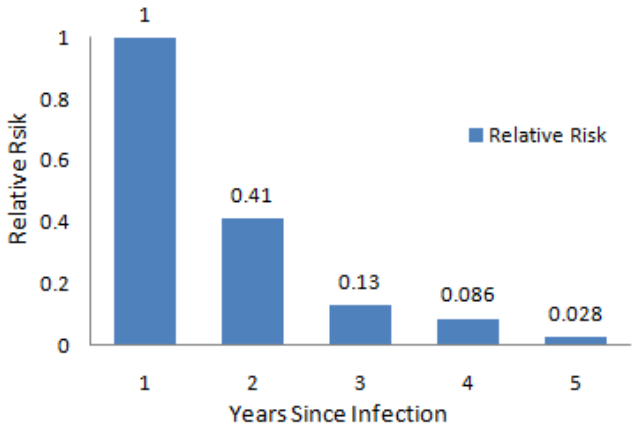


Figure 3.6: Relative risk of developing Active TB reproduced from [50]

$$d'_n = 1 - e^{-d\times 50} \tag{3.14}$$

$$= 0.1447 \ \ per\ 50\ years \tag{3.15}$$

$$d'_s = 1 - e^{-\sigma_2 d\times 50} \tag{3.16}$$

$$= 0.2094 \ \ per\ 50\ years. \tag{3.17}$$

Figure 3.6 shows the relative risk of developing active TB since infection. According the relative size of the rates of years 0 to 5 after infection in Figure 3.6, 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.

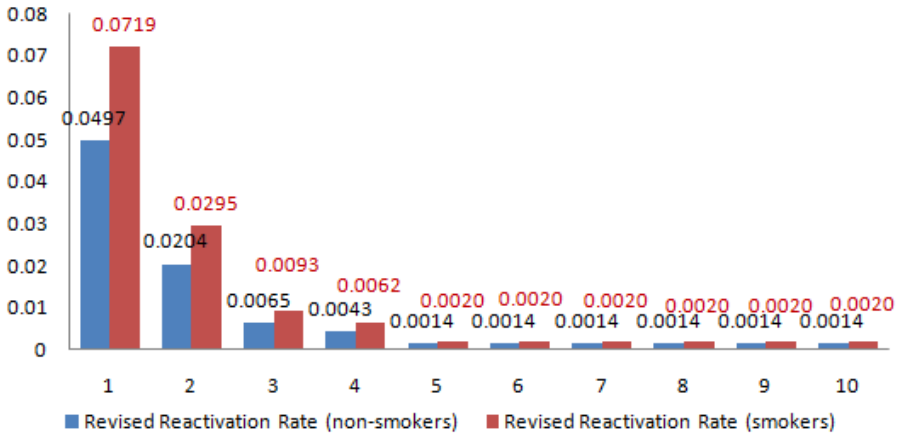


Figure 3.7: Revised Reactivation Rate since Infection

To compare the results of two models on an equitable basis, it is 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 3.7. That figure 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.

3.3.4 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, the degree of assortive mixing between smokers and non-smokers, was set to be 0 in both the individual-based and aggregate model. $\sigma_8 = 0$ 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 (defined in Table 3.1). 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 [30]. 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 Erdös and P. Rényi. In a random network, the probability that two nodes are connected is assumed uniform, and each individual is connected randomly with a given average number of connections, regardless of any consideration of spatial position or other individual attributes [5].

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 [28]. Lack of clustered groups and homogeneity of individual-level network characteristics make random network models analogous to a random-mixing aggregate model, such as the aggregate SD model presented in [28]. 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 e-mail networks,Internet networks and the structure of software modules [23]. 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 [28].

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 broadly once they are infectious TB cases. 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. 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 [28]. 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 [52].

3.4 Results

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

3.4.1 Individual-based Baseline Model

Since this individual-based (or agent-based) model is a stochastic one, we simulate this baseline model for 100 runs to verify that it yields results comparable to those associated with the aggregate

(or System Dynamics) model. The results suggest that the differences between these two models are small and can be explained by stochastic factors, seen in Table 3.3 and Figure 3.8. 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 [26]. 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 likely due to stochastic effects.

Table 3.3: Stocks/States of System Dynamics Model(SDM) and of Agent-based Baseline Model(ABM) at 50th Year

Stock/State	**SDM Results**	**ABM Mean**	**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

Figure 3.8 shows the baseline trajactories for all the stocks/states of both System Dynamics model and agent-based baseline models. The black lines show the population size in each stock over time, while the red ones are the results from agent-based baseline model. The behavior of the agent-based baseline model over time is quite consistent with that of aggregate System Dynamics (SD) model.

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.

3.4.2 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 3.9 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

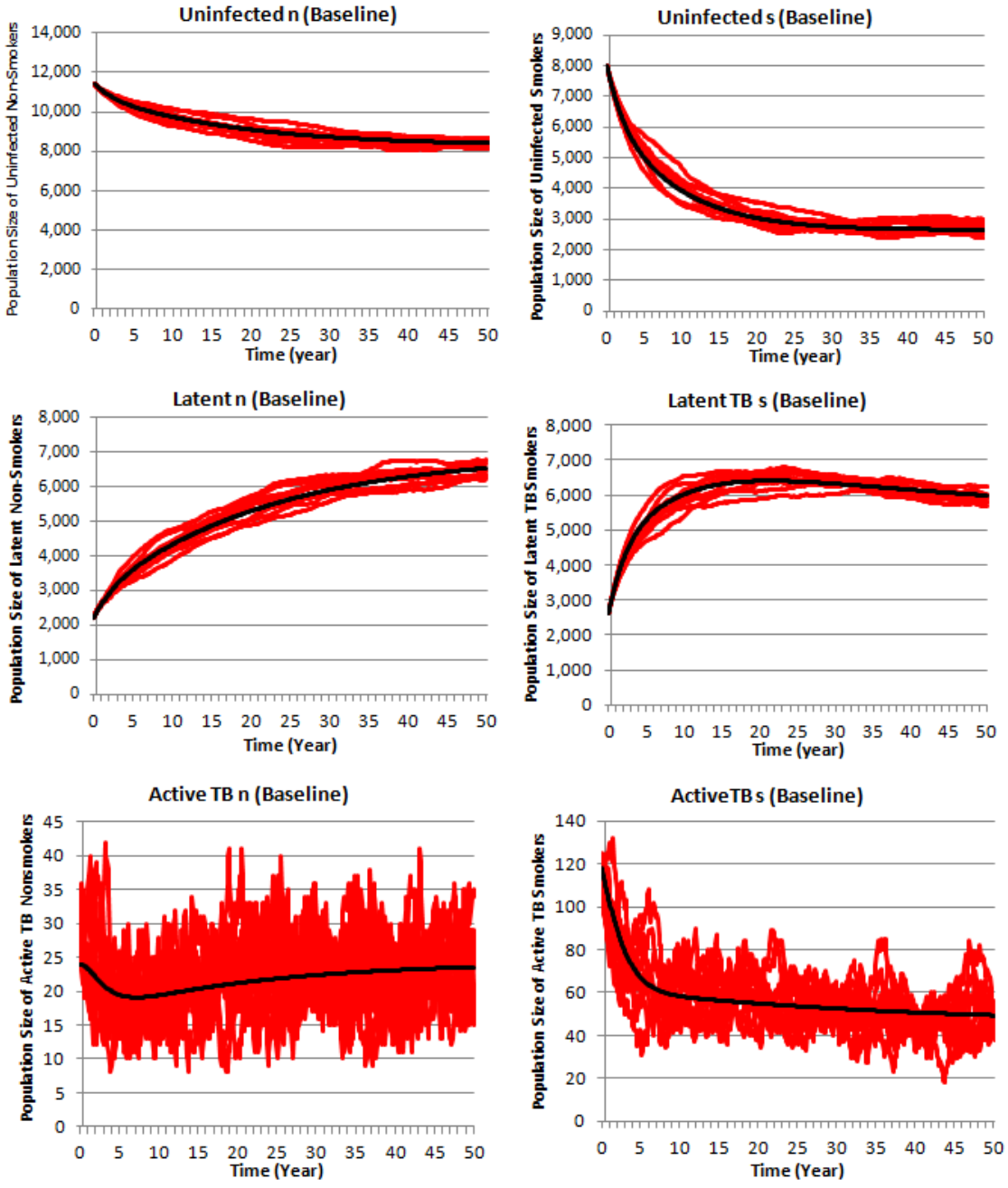


Figure 3.8: Comparison of System Dynamics Model and Agent-based Baseline Model

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.

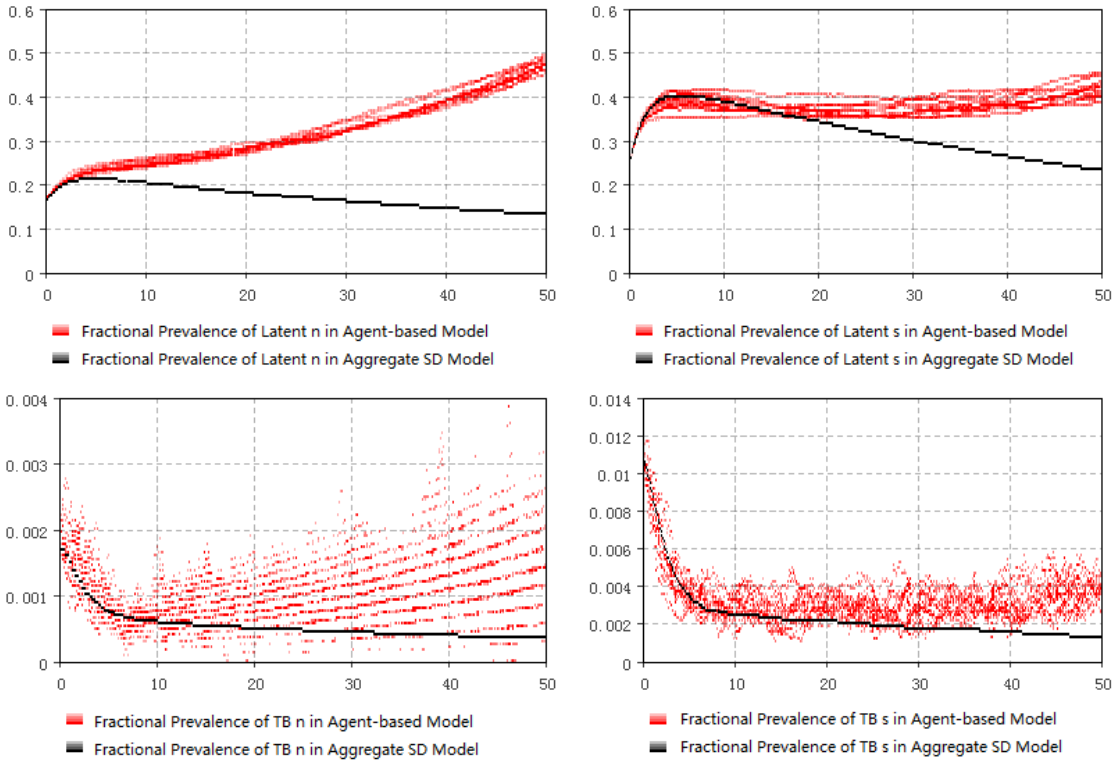


Figure 3.9: Prevalence of TB Infection and Active TB Given BCG Administration in Agent-based Model and Aggregate Model

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 straightforward 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.

3.4.3 Evaluation of Memoryful 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.

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 3.10 and Figure 3.11 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 3.4 depicts a comparison of agent-based modeling results with historical information from Saskatchewan

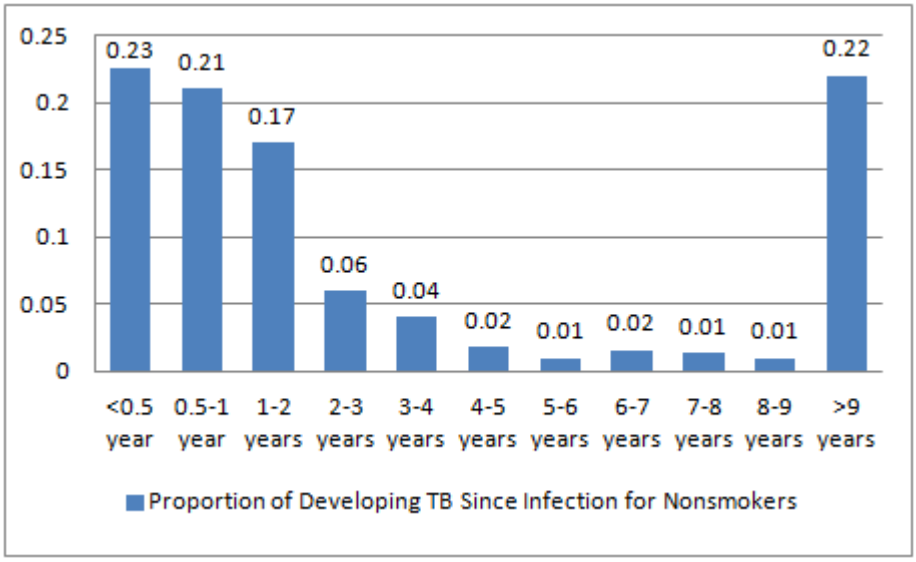


Figure 3.10: Interval from Latent to Active TB for non-smokers

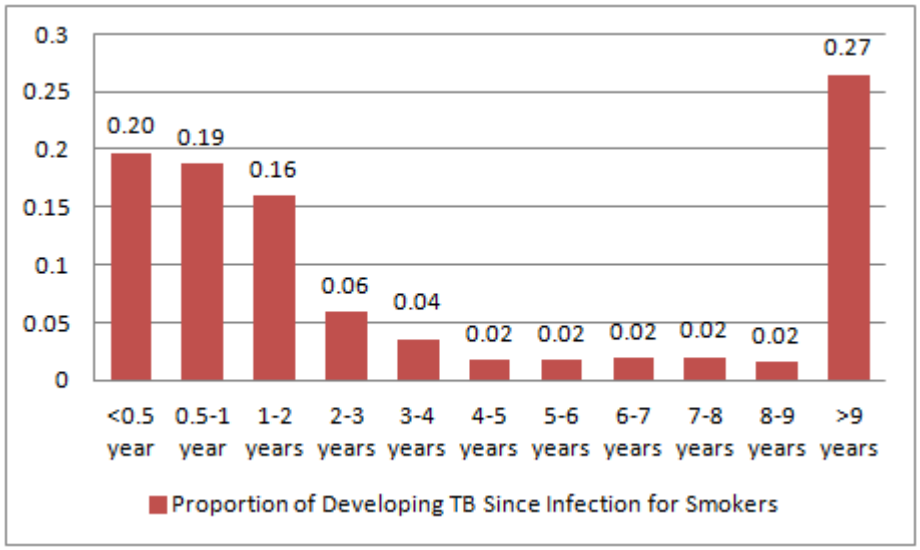


Figure 3.11: Interval from Latent to Active TB for smokers

Anti-TB League Report [42].

In Table 3.4, 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.

Table 3.4: Comparison between historical information and estimated results of AB model

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	-	100
8-9 year	74.94	-	100
>9 year	100	100	100

Figure 3.12 shows the age structure of TB cases for both smokers and non-smokers. As exhibited in Figure 3.12, 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 [37], 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.

The implementation of age in this experiment underscores the possibility of capturing age distribution among TB cases. By integrating the real age structure of the population in Saskatchewan, such information can be valuable especially when we calibrate our model with Saskatchewan data on age-specific case rates.

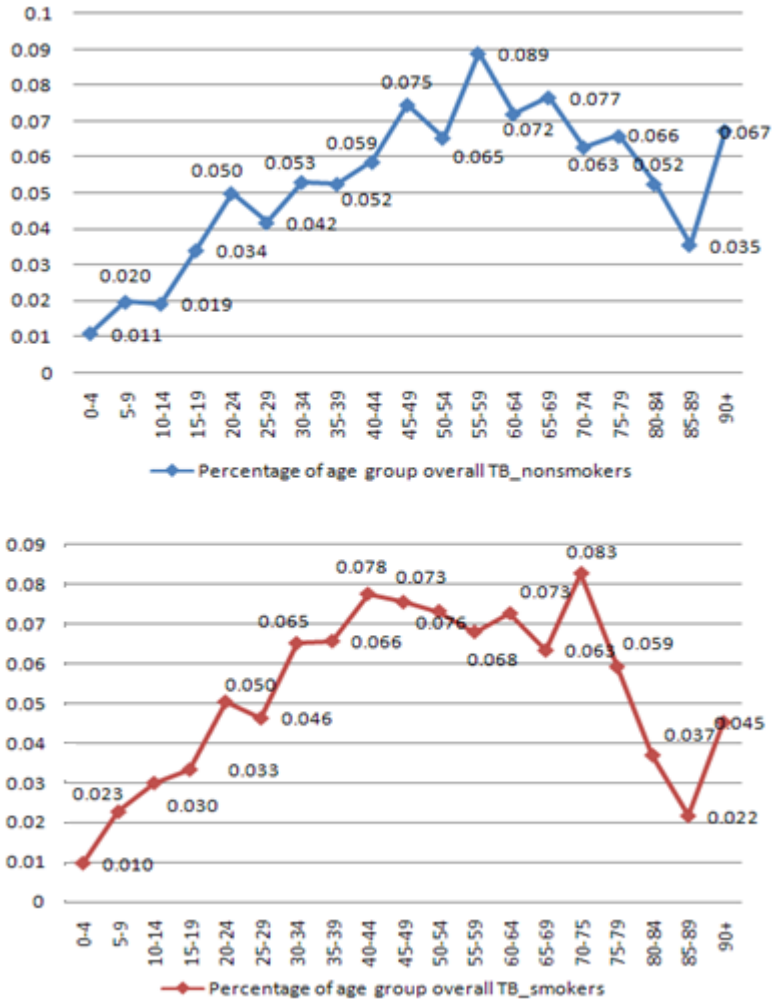


Figure 3.12: Age Structure of non-smoker TB Cases and smoker TB Cases

3.4.4 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 after simulating the individual-based model for 10 realizations.

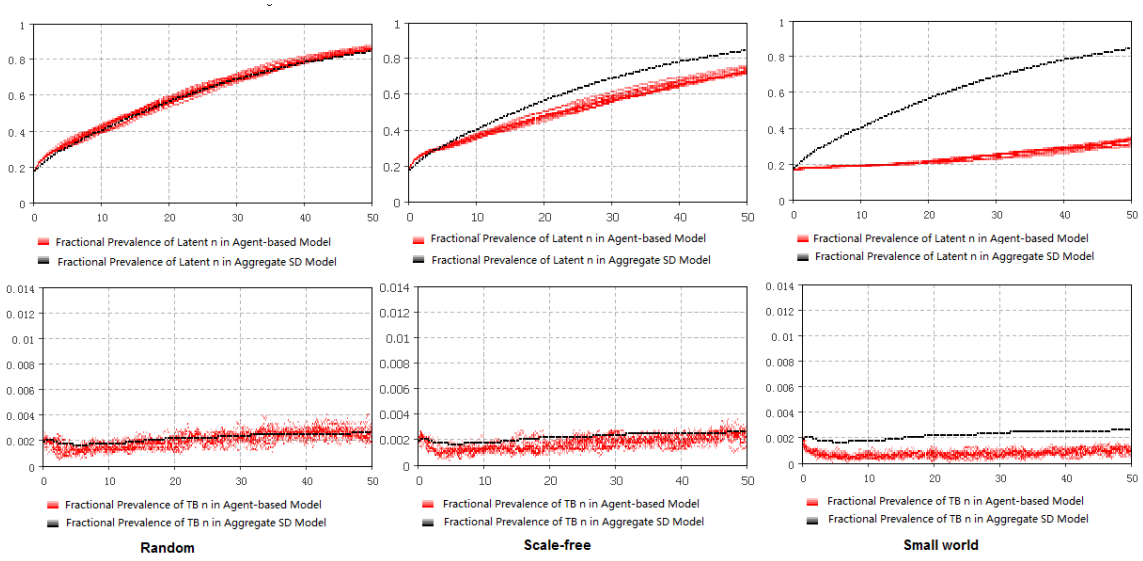


Figure 3.13: Fractional Prevalence of Infection of Non-smokers Under Alternative Network Topology in Agent-based Model and Aggregate SD Model

In Figure 3.13, 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 3.14, a random network produces a lower prevalence of latent TB infection among smokers than that emerging from the aggregate model. 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 networks exhibit a lower level of infection prevalence, while the 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

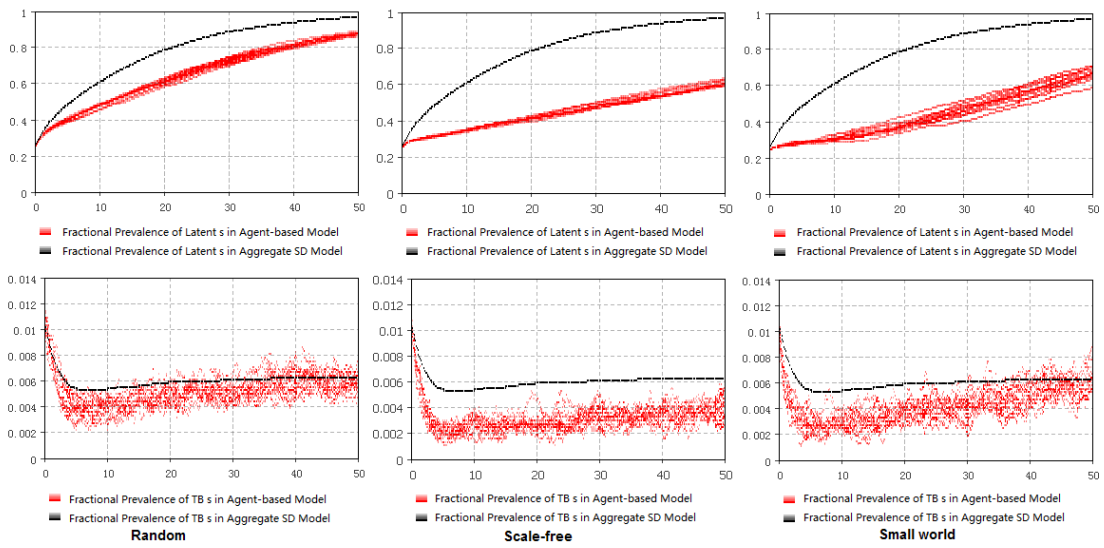


Figure 3.14: Fractional Prevalence of Infection of Smokers Under Alternative Network Topology in Agent-based Model and Aggregrate SD Model

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.

3.5 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 comments on the trade-offs 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, the modeler 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, intermixed 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 levels 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 [11]: 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; however, sometimes it is worth while to simulate individual-based models to gain richer understanding of the system behaviors given the detailed level of information obtained in such 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 simulate and exhibit the association of transmission of infection and the presence of long-term relationships between individuals more realistic, 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 [32]. 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 [12, 21, 28, 32].

Chapter 4

A System Dynamics Model of Tuberculosis Diffusion With Respect to Contact Tracing on Community 1

This chapter describes a preliminary dynamic model to evaluate the role of current contact tracing policies in managing TB transmission. Through a novel representation of contact tracing dynamics, the model supports investigation of how TB outcomes are affected by changes to the breadth and timeliness of contact investigation. This chapter is submitted and accepted for publication as a full paper in Proceedings of the 2011 Winter Simulation Conference December 2011.

4.1 Background and TB Control Activities in Saskatchewan

The difficulty of studying the application-oriented complex systems with use of traditional epidemiological tools was recognized for a long time. In recent decades, System Dynamics modeling and related compartment models have been used as a complement to resolve some challenges in infectious disease control. Dynamic modeling has shaped our understanding of the internal linkages between causal factors and system feedback, interpreted epidemiological patterns and trends, and informed us in policy design and evaluation at a practical level. The complexity and variability of policy related issues accompanying the dynamics of infectious disease make applications of system dynamics modeling a powerful and promising tool in investigating some important transmission patterns and analyzing the trade-offs of diverse intervention programs.

Tuberculosis (TB) is a communicable disease which drains public health systems. Health representatives in many countries plan and organize diverse control and prevention programs to fight against the disease. The process of contact tracing investigation (CTI) serves a key function in many TB control programs in developed countries.

In the Canadian province of Saskatchewan, most of the TB cases occur among First Nations people (one of Canada's Aboriginal people). They are found to suffer from a high prevalence of TB infections. Saskatchewan health organizations have conducted many prevention programs

to control its diffusion; such TB control activities include screening, lab investigation, contact tracing and treatment. The contact tracing program, in particular, works as an epidemiological tool in finding potential cases, sources of infection and latently infected contacts with a high risk of developing TB in their lifetime.

In this chapter, we will adapt a System Dynamics model to investigate the effectiveness of the contact tracing policies in community 1 which locates in Saskatchewan. Our model extends existing mathematical models of TB dynamics (including the classic SIR model[1] and Osgood et al.[37]) to capture the dynamics of TB in the context of a community as well as to estimate the efficiency of the current contact tracing program.

To improve the quality and efficiency of disease control, sensitivity analysis can be conducted to sort through alternatives and to assist the decision making process in identifying the best choice with which to serve both financial constraints as well as objectives of intervention programs. We present and discuss two modeling questions to illustrate how System Dynamics models could be used to support and improve the development of control policies, contact tracing investigation strategy in particular. With respect to efficiency of contact tracing, our first question aims at exploring the efficiency in terms of investigating fraction of the contacts. This paper also seeks to explore the benefits conferred by speeding up the contact tracing investigation procedure. More detailed explanation regarding these two questions is demonstrated in Section 4.4. The experimental results are presented in Section 4.5. Section 4.6 discusses model findings and summarizes our work.

4.2 Contact Tracing Objectives and Procedure in Community 1

The main objective of contact tracing is to find latently infected people, and active TB cases – especially infectious ones. Contact tracing investigation in Saskatchewan targets at two types of person, namely those with infectious TB and primary TB [43]. Infectious TB cases are transmitters of the infection; people who come in contact with infectious active TB cases can be infected. Tracing infectious TB cases can eventually help find more infected persons with high risk of developing TB in their lifetime. Providing treatment for latently infected people can lower their risk of developing active TB and stop the potential pathways of TB transmission. The term "reverse contact tracing" describes the investigation of primary TB cases who are recently infected [43]. Since contacts of primary TB cases can be sources of TB infection, infectious TB cases are the main targeted group in this case [43].

Within Saskatchewan, contact tracing focuses on all the contacts within 30 days of diagnosis of the active TB case. A contact list is created and ordered from the greatest to the least amount of exposure time [43]. It includes the following:

- Household members: it involves family members and babysitters who live with the client in the same house. Children have the highest priority for contact tracing, and then any family member who has breathed the same indoor air with the client for at least 10 hours within the past month prior to the diagnosis.
- Immediate family members who do not live with the client in the same house.
- First generation biological relatives, including aunts, uncles and cousins, etc.
- Family or friends who are current or past TB cases.
- Current or past active TB cases in the community, and those who have contacts with them.
- People from public places including schools, restaurants, bars, work and so on.
- Relatives by marriage: aunts, uncles, cousins, nieces and nephews.
- Hospitals and special care homes.
- People who ride in the same vehicle repeatedly (taxi, car, bus, school bus).
- Visitors.

4.3 System Dynamics Model of Tuberculosis Transmission in Community 1

We have developed System Dynamics model of TB transmission to reflect the community 1 context and contact tracing investigation. In this model, we assume that the population changes only via birth or death, since the data regarding the mobility of the population in Community 1 is limited.

4.3.1 Structure of System Dynamics Model

The structure of our model is illustrated by stocks and flows in Figure 4.1. Our model consists of eleven stocks. Of these eleven stocks, ten represent the population, distinguishing individuals according to TB-specific health status and status with respect to contact tracing. The remaining stock named "Named By TB Cases for Investigation (Cv) represents a name list of contacts to be investigated via contact tracing. It is worth emphasizing that this stock does not represent on segment of the population collectively represented the other stocks. There are no individuals in this stock - rather, the stock holds a list of names of the contacts which are obtained from those traced active TB cases. The named individuals are assumed to reside in one of the other 10 stocks. Transitions depicted by flows in the upper part of the diagram reflect the pathways through which people's health and contact tracing status can progress over time.

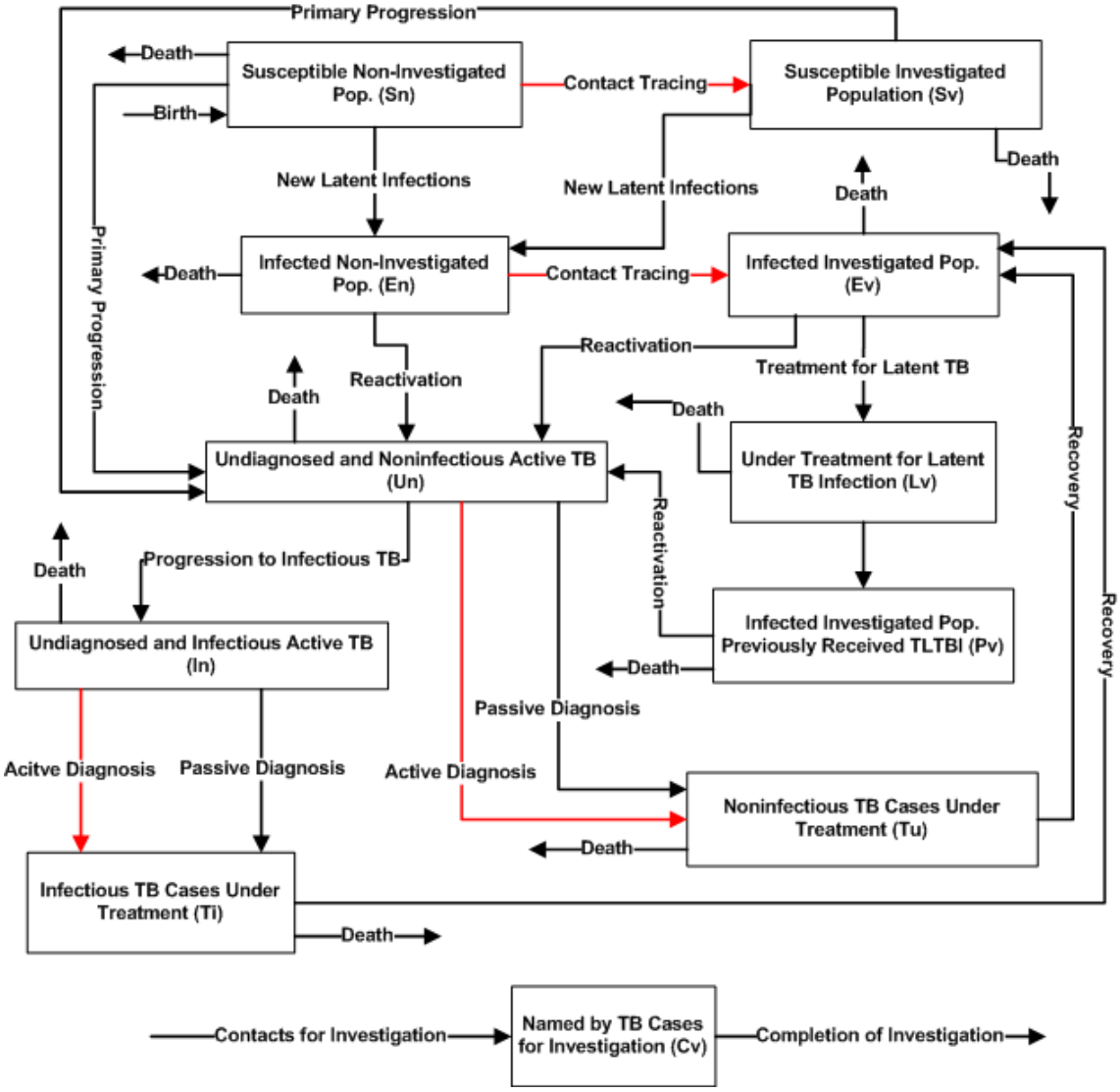


Figure 4.1: Structure of TB Model With Respect to Contact Tracing

According to Figure 4.1, Susceptible people are divided into "Susceptible Non-Investigated Pop.(S_n)" and "Susceptible Investigated Pop.(S_v)". Similar categorizing patterns have also been applied on infected people with no current disease. Here, the "investigated" designation denotes those for whom we have a record of their TB infection status, while "Non-Investigated" represent people who might be uninfected, infected or active TB cases, but where those authorities do not know their TB infection status. The investigation process is accomplished via a contact tracing program and is supported use of the Mantoux skin test to test whether an individual is infected by TB. Individuals who are susceptible whether investigated or not - can acquire TB infection by contact with an infectious cases; once infected, they will move into the "Infected Non-Investigated Pop.(E_n)" stock via latent infection or "Undiagnosed and Noninfectious Active TB (U_n)" via primary progression. After acquiring TB infection, those previously investigated susceptible people are no longer considered investigated, since their TB infection status has been changed and we cannot recognize these newly updated TB status unless they are investigated another time. Suppose a person is traced and found to be Mantoux negative; he will then stay in the "Susceptible Investigated Pop.(S_v)" stock. If this person is subsequently infected and becomes a latent TB case, he will go to the "Infected Non-Investigated Pop.(E_n)" stock. The reason why his investigation status goes to Non-Investigated is that, his current Mantoux test result – which following infection should be Mantoux positive – is as yet unknown to those performing contact tracing. His investigation status will not be updated to investigated unless he is traced and tested for a second time.

People in "Infected Non-Investigated Pop.(E_n)" can either move into "Infected Investigated Pop.(E_v)" via contact tracing, or progress to the active TB disease state. Those infected and investigated people have a chance to undergo a prophylactic treatment (treatment for latent TB infection) for 6 months to lower their risk of progression to active TB disease. Not every infected and investigated individual can get TLTBI, because the eligibility of TLTBI is restricted by certain criteria, including those considering age, health status, history of Active TB and so on. The "Under Treatment for Latent TB Infection (L_v)" and "Infected Investigated Pop Previously Received TLTBI (P_v)" stocks characterize those traced infected people who are currently under the treatment for latent TB infection (TLTBI) and those who have finished their prophylactic treatment and are currently under certain protection from developing disease. The proportion of protected population with TLTBI at risk of developping active TB is 0.32 [37].

People can progress to active TB via either primary progression or reactivation. In our model, we distinguish people with active TB disease into 4 stocks according to their diagnosis status and TB infectiousness, namely "Undiagnosed and Noninfectious Active TB (U_n)", "Undiagnosed and Infectious Active TB (I_n)", "Noninfectious TB Cases Under Treatment (T_u)" and "Infectious TB Cases Under Treatment (T_i)". Those undiagnosed noninfectious TB cases might develop infectiousness within a period of time if they are not diagnosed. It is worth emphasizing that both

passive and active diagnosis methods are implemented in our model to highlight the role of contact tracing in identifying active TB cases, particularly those with the infectious form of active TB. After recovery from either the infectious TB state or the noninfectious TB state, they will return to the "Infected Investigated population stock because authorities will know that their Mantoux test results remain positive.

The stock "Named by TB Cases for Investigation (Cv)" represents the number of contacts that are queued for tracing and testing. Contacts who are named within this stock will be found to be infected investigated people (Mantoux Positive with no current disease), Susceptible investigated people (Mantoux Negative) or active TB cases through active diagnosis.

Every year, a number of infectious or noninfectious TB cases will be diagnosed. Those diagnosed as infectious TB cases will be traced. However, for those diagnosed as noninfectious TB cases, only a fraction of them (typically the primary TB cases) will be traced. Using average contacts per active TB cases per year, we can derive the average number of contacts traced per year. An important motivation for tracing primary TB cases is the fact that most such cases occur in children. Compared with adults, children are highly susceptible to TB, and have generally acquired the infection relatively recently. Tracing primary TB cases (which are by definition those perceived as recent) is undertaking with the goal of identifying the source case. Tracing those infectious ones can help stop the spread of the TB infection, and tracing the noninfectious ones can help find the sources of the infection.

4.3.2 Parameterization

Parameters used in the model are either derived from the Saskatchewan TB Control Program database and provided supplemental information, or from a variety of literature [20, 30, 37]. Key parameters used in the model are depicted in Table 4.1.

Table 4.1: Description of the Symbols and Parameter Settings in TB Model with Contact Tracing

Parameter	**Description**	**Value**	**Unit**
β	Likelihood of infection per contact between a susceptible and an infectious active TB case	0.46	1
C	Average number of contacts per TB case per unit time (year)	40.96	persons per year
P_r	The proportion of protected people with latent TB infection at risk for active TB	0.32	1
π_t	The mean time of treatment for active TB cases	0.75	year

π_n	The mean time spent under TLTBI treatment for latently infected people	0.5	year
π_{in}	Mean Time for a noninfectious TB case developing infectiousness	1.29	years
π_{dn}	The mean time to trace and investigate a given contact of that case (averaging over all contacts)	0.25	year
μ_{tb}	Mortality rate from TB	0.037	per year
μ_n	Mortality rate from other disease	0.0274	per year
d	Rate of Reactivation (Progression from latent TB to active TB due to endogenous changes)	3.125×10^{-3}	per year
ρ	Proportion of newly infected individuals progressing to primary TB	0.05	1
τ	Fraction of infected TB cases eligible for TLTBI	0.3	1
C_{tn}	Average contacts per traced noninfectious TB cases per year	13.44	persons
C_{ti}	Average contacts per traced infectious TB case per year	54.625	persons
p	Chance of an noninfectious TB case (particularly primary TB case) been traced	0.18	1

4.3.3 Calibration of System Dynamics Model

Although the model includes many parameters values estimated from the relevant datasets and literature, there are still some parameters which can't be retrieved or easily estimated. Most notably, there are a large number of people with unknown/non-investigated TB status in our model. Several studies provided us with some rough estimates of the real world situation, such as data from mass screening for TB in Saskatchewan during the 1970s and earlier. But it is only a rough estimation, and it is likely to exhibit pronounced variation.

In order to improve the validity and reliability of our model, it is highly valuable to calibrate our model to best match historically observed values between 2001 and 2007. These historical data across the time series include incidence rate, number of traced people with Mantoux positive results, and the population size of community 1. Calibration was performed using a Powell Optimization algorithm (used within Vensim software) [48] to adjust several parameter values so as to find the best observed match to the historic data. The parameters to be calibrated in our model are the birth rate, initial fraction of TB infection, initial undiagnosed noninfectious TB cases and initial undiagnosed infectious TB cases. Firstly, birth rate was calibrated via setting its initial value as

well as its upper and lower boundaries. Then, with use of the calibrated birth rate, the remainder of the three parameters are calibrated by matching incidence rate and traced people with Mantoux positivity status. Each calibration is simulated for around 100,000 times to optimize a weighted sum of square about the discrepancies between the historical data and simulated results. The best calibrated parameter values we can get are shown in Table 4.2.

Table 4.2: Calibrated and Estimated Symbols and Parameters

Parameter	**Description**	**Value**	**Unit**
a	Annual birth rate of community 1	0.0469	per year
λ	Initial Fractional Prevalence of TB Infection. Infected people with no current disease people over the total susceptible and infected people in community 1 at 2001	0.286	1
t_{dn}	Mean Time until diagnosis for non-infectious active TB	0.6413	year
t_{di}	Mean Time until diagnosis for infectious active TB	0.5	year
Initial In	Initial value of undiagnosed and Infectious Active TB stock	1	person
Initial Un	Initial value of undiagnosed and noninfectious Active TB stock	1	person

4.4 Scenario Definitions

The main focus of our model is to investigate the effect of contact tracing on capturing the source of infection as early as possible and in controlling the spread of the disease. In order to assist the process of policy formulation and decision making in disease control using the model, we have designed a variety of scenarios to answer two what-if questions, and each scenario is simulated from 2001 to 2030 to investigate the potential outcomes under different parameter settings.

- Question 1: *How would the benefits of contact tracing vary if contact tracing were to be carried out on various fraction of the contacts of infectious active TB cases (e.g. 10%, 20%, 50% or all contacts)*? To assess the cost-effectiveness of tracing all contacts, compared to tracing only a subset such as those thought to be at high risk (e.g. by reasons of prolonged or intimate exposure), we frame the question in the context of a community whose identity is restricted for reasons of confidentiality (community 1) by asking what level of incidence of active TB would result from tracing an successively larger fractions of contacts of an infectious active

TB. We firstly abstract away from the issue of differing levels of risk for different subsets of the contact population, and instead consider the impact of investigating different fractions of the contact population, assuming an equal distribution of risk across that population. What might the incidence rate be if contact tracing is not performed? How much improvement in disease control can be attributed to a contact tracing investigation? In addition, considering the risk in terms of the incidence rate of active TB, we also explore the impact of partial contact tracing on prevalence of TB infection across the overall population.

- Question 2: *If we improve the speed of contact tracing, what would the impact be on TB incidence rate and prevalence over the time*? Tracing contacts more promptly provides an opportunity to reduce exposure time and new TB infection cases. If we go a step further, lowering the number of newly infected people can eventually reduce the prevalence of TB infection as well as the incidence rate of active TB. Considering the highest level of population health, it is natural to seek to operate contact tracing investigation as fast as we can, nevertheless we are bounded by a limitation of human resources and budgets. A step towards understanding the cost-benefit trade-off of contact tracing speed is to evaluate the benefits of speedier contact tracing. If public health authorities can deploy sufficient resources to reduce the mean time of contact tracing by a factor of 10 or even of 20, what attributable level of benefit would be secured? How many new TB infection cases can be eliminated? Our TB model seeks to explore these potential gains in the context of TB transmission.

4.4.1 Sensitivity Analysis on Coefficient of Tracing Contacts

This experiment is designed to answer question 1 about the impact of tracing partial contacts of active TB cases. As a first step, we set our calibrated baseline model to represent ongoing contact tracing. For the calibrated baseline model, the coefficient of investigated contacts is 1 (or 100%). Then based on this calibrated baseline model, we adjusted the coefficient of contacts coming into investigation to 0%, 10%, 50% and 150% of that baseline value. Coefficient with value of 0% indicates that there are no contact tracing ongoing within community 1, while 150% means we spend extra effort in the interview session to identify additional contacts, with an additional 50% margin of those reported contacts coming into investigation. This analysis is based on the assumption that people might not report their full social networks and contacts due to difficulties in recollection or personal reasons, and the response rate can hardly be 100%, since contacts might have problems coming into investigation due to personal mobility or other reasons.

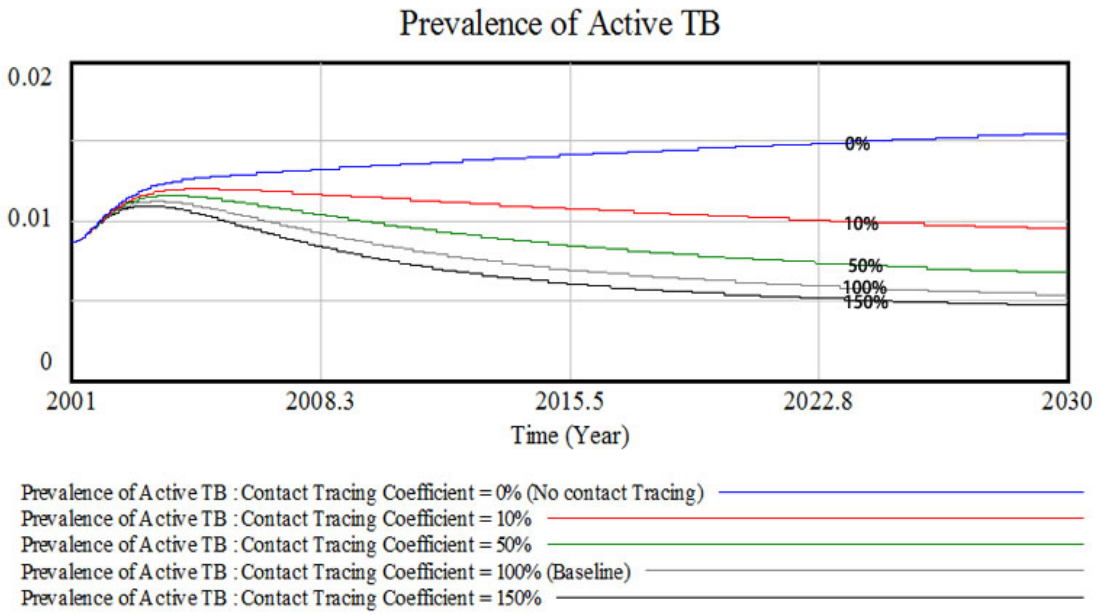


Figure 4.2: Prevalence of Active TB Corresponded with Different Coefficients of Investigated Contacts

4.4.2 Sensitivity Analysis on Mean Time of Contact Tracing

Within this scenario, we sought to investigate the effect of accelerating or decelerating the contact tracing process on TB control as a whole. For this experiment, we conducted simulations with different mean time for contact tracing including both speeding up and slowing down by a factor of 10 or even of 20. For this experiment, the coefficient of investigated contacts remained 1 for all experiments.

4.5 Experimental Results

4.5.1 Results of Coefficient of Tracing Contacts

Figure 4.2 presents the results for scenarios examining changes to the assumptions regarding the breadth of contact tracing. Absent contact tracing, the model suggests that the prevalence of active TB will be increasing over time, and it reaches 1.54% in 2030. However, for scenarios with contact tracing, the prevalence of active TB declines over time, following a peak at roughly 2003. Tracing contacts at 10% of the baseline level shifts the increasing trend of prevalence into a decreasing one after a short period of increasing between 2001 and 2003. When tracing 50% of the baseline level, the prevalence of active TB decreases dramatically. When tracing at 100% of the baseline level, although the prevalence of active TB (0.53% in 2030) is lower than that resulting from tracing at 50% of that level (0.67% in 2030), the incremental benefits are very limited. In short, the reduction

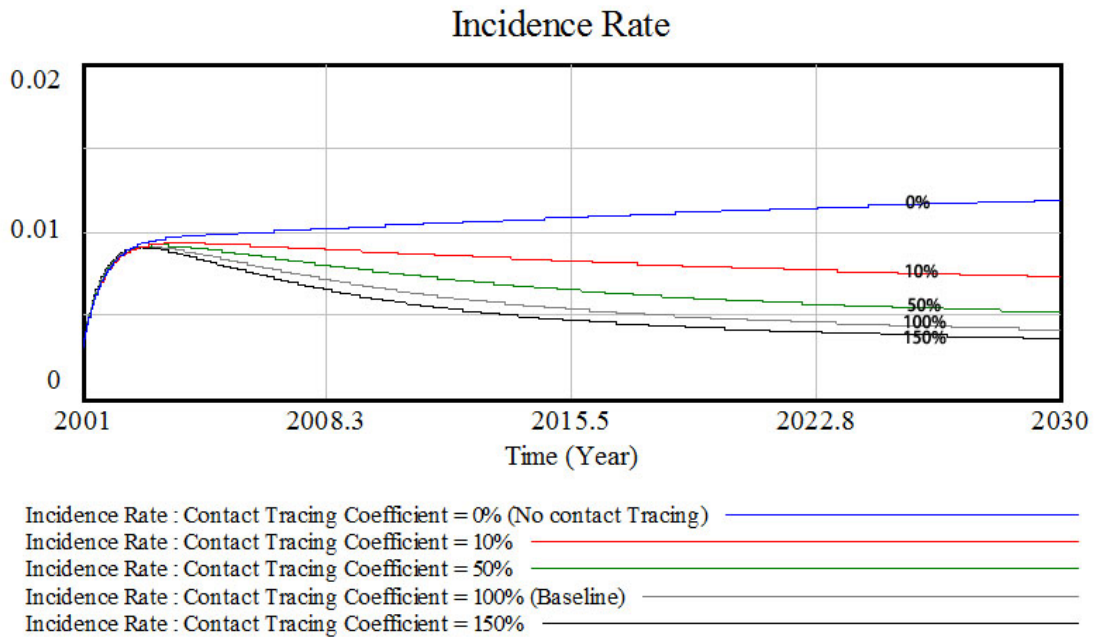


Figure 4.3: Active TB Incidence Rate for Different Coefficients of Investigated Contacts

in prevalence secured by tracing the first 50% of the baseline level of contact tracing(0.87%) is much larger than that obtained by tracing at 50% of the baseline level (0.14%).

Figure 4.3 shows the impact of additional levels of contact tracing on the incidence rate for Active TB. All scenarios initiate with a transient consisting of a sharp rise in the incidence rate. This transient suggests an inconsistency between factors within the model, such as between of the fraction of infection assumed within the initial population on the one hand, and the rates of mixing (c) and transmissibility of the bacterium within this epidemiological context (β). Absent contact tracing, the active TB incidence rate increases slowly following the initial transient. When tracing at 10% of the baseline level, the incidence rate lies below that resulting from no contact tracing with a declining trend. When tracing at 50% of the baseline level, the incidence rate declines slowly over time after reaching its peak around in 2003. The incidence rate decreases faster after 2002 when tracing more contacts, but once again the extra benefits we obtain when tracing incrementally larger fractions of contacts (compared to the baseline) yield markedly decreasing returns. For example, tracing an extra 50% increase in the rate relative to the baseline scenario contributes even smaller benefit.

4.5.2 Results of Mean Time of Tracing

This experiment tries to answer question 2 regarding the impact of reducing the mean time of contact tracing on TB control, as judged by the prevalence of active TB. According the trend lines displayed in Figure 4.4, as a whole all the trajectories have the declining tendency over the time,

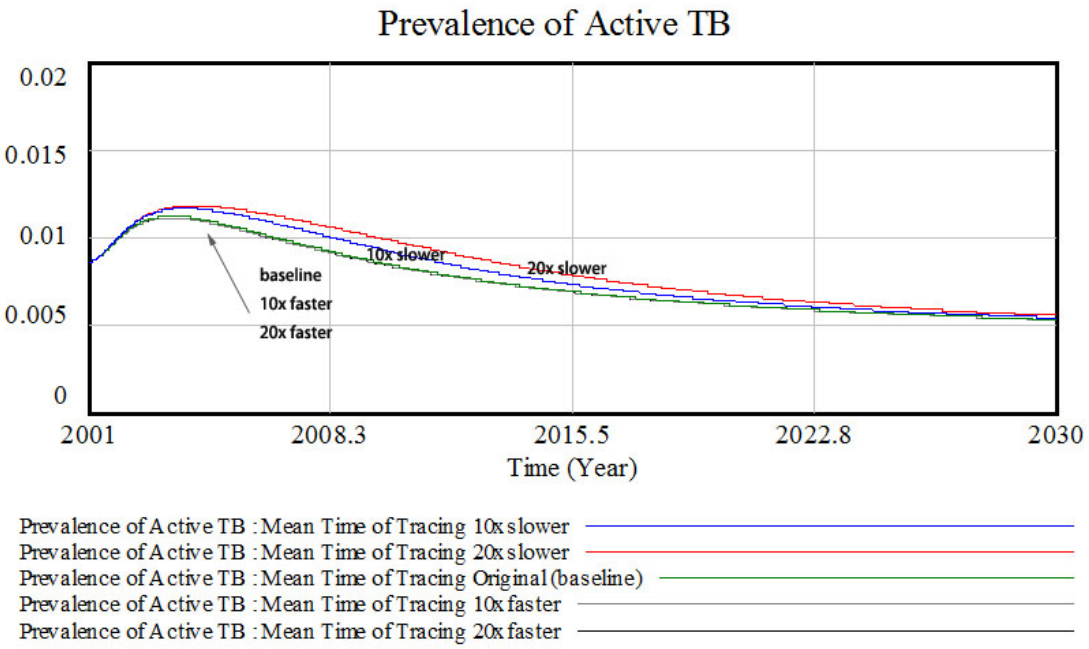


Figure 4.4: Prevalence of Active TB Produced by Different Mean Time of Tracing Contacts

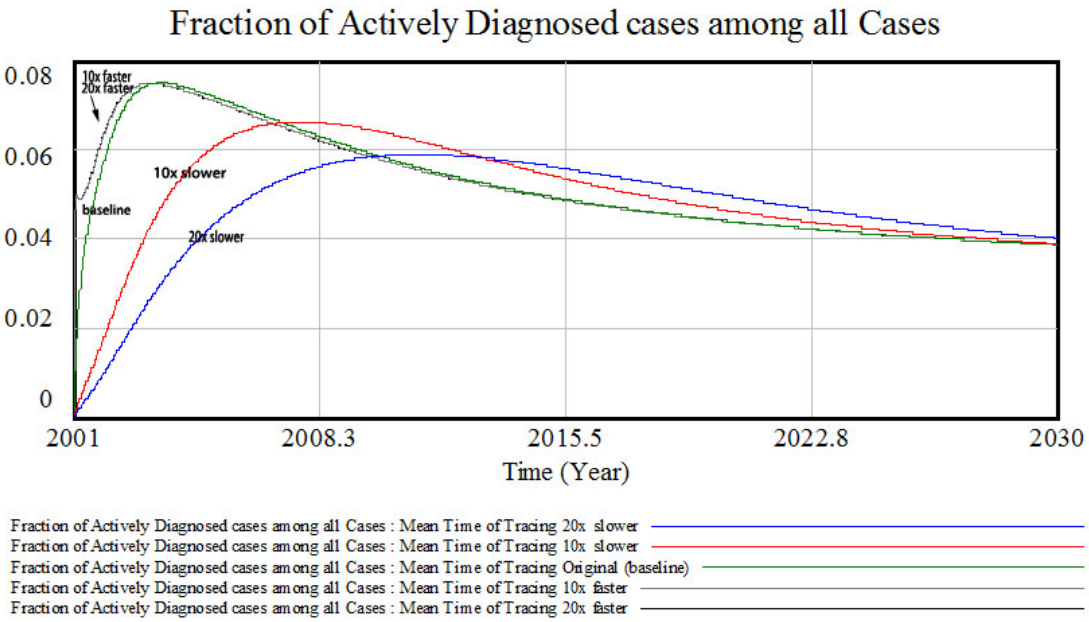


Figure 4.5: Fraction of Actively Diagnosed Cases Among All Incident Cases

except for the initial transient. We observe that changing the mean time of contact tracing slightly around the baseline doesn't provide pronounced benefits in TB control. The impact of decreasing the mean time of contact tracing even by integral factors (e.g. a factor of 2 or 3) is very small, and for small changes, the impact can almost be ignored. Even for larger changes (such as the factors of 10 times shown in the diagram), the impacts are very limited (e.g. prevalence of active TB in 2010: 20x slower=0.98%, 10x slower=0.92%, baseline=0.84%, 10x faster=0.84% and 20x faster=0.84%). While the aggregate character of the model suggests the need for great caution, such results suggest that there may be limited gains to be secured by investments in reducing the time required for contact tracing.

Figure 4.5 illustrates the fraction of actively diagnosed cases among all the incident cases over time. As a whole, all the trajectories increase first and decrease afterward. The year in which the fraction of actively diagnosed cases reaches its peak for each scenario is presented as: 20x slower=2011, 10x slower=2008, baseline=2003.5, 10x faster=2003.2 and 20x faster=2003.2. We observe that the faster contact tracing investigation goes, the sooner the fraction of the actively diagnosed cases reach its peak.

4.6 Discussion and Conclusion

The first question explored in this paper concerned the impact of tracing partial contacts on disease control and prevention. In this area, our models have explored a few alternatives focused on tracing subsets of the contacts. Model results suggest that the impact of contact tracing is pronounced in TB control as we have observed dramatic decrease in incidence rate and the prevalence of active TB. For the scenarios exploring partial tracing, we found that incidence rate decreases when tracing successively larger fractions of contacts; however, the contribution of tracing more contacts declines markedly. This phenomenon of diminishing returns would appear of relevance when developing control policies. The results suggest that strong gains can be obtained when tracing merely partial contacts. Given limited resources, it also tentatively suggests that tracing partial contacts may be more effective and beneficial especially when we set priority for contact tracing to those with high risk (such as those experiencing longer exposure times, or individuals who are known to have experienced past contacts with several other cases).

Our second investigation sought to understand the mean time of contact tracing. Within this area, the model was used to explored a few scenarios. Given model assumptions, TB control effectiveness is relatively insensitive to mean time of contact tracing especially within the large range of 20 times faster to 20 times slower than the baseline. When reducing the mean time of contact tracing and boosting the effort to track TB cases sooner compared to the baseline, the achievement yields only modest gains and the patterns are similar to that obtained at baseline.

This suggests that there may be unexpected degrees of flexibility in the speed of tracking contacts, and raises the interesting possibility that the current guidelines and schedule for contact tracing are sufficient fast to secure most of the benefits to be gained through contact tracing.

By contrast, through experiments with different fractions of contacts coming into investigation, it is found that assumptions regarding the fraction of contacts coming into investigation yield pronounced impacts for the prevalence of TB infection. Bringing larger fractions of contacts of individuals in for contact tracing can yield a reduced prevalence of TB infection. Tentative though they are, the results suggest that the gains exhibit diminishing returns. Depending on the relative cost of undertaking contact tracing and the cost of treating TB, putting into place some basic measure of contact tracing would appear likely to yield the greatest incremental benefit (when compared with enhancing the level of tracing experienced in a program already drawing many contacts).

While deducing the optimal level of contact tracing will require consideration of the cost trade-offs involved. Such diminishing returns suggest that there is likely to be decreasing cost-effectiveness for bringing in larger sets of individuals. While the current model did not seek to capture important known risk factors for TB, such observations would suggest that focusing limited contract tracing effort on a smaller segment of all contacts who have a very high risk of being infected or of developing active TB given infection could yield disproportionate benefit. Such high-risk individuals would include contacts who are known to have had contacts with many other cases, and those in the top risk categories currently used by TB control. With priority setting in tracking the contacts, the striking gains from partial contact tracing seen here will likely be significantly more efficient yet.

We were surprised to find such limited incremental effectiveness due to reductions in the time required for performing contact tracing. In Figure 4.4, we did not observe significant improvements in prevalence of active TB while implementing a faster contact tracing protocol. This phenomenon merits close additional study before any definitive policy-relevant implications are drawn, it raises the provocative possibility that the current speed of contact tracing may be securing the large majority of the benefits that would be gained even by far faster contact tracing. It is undeniable that speeding up the contact tracing process would offer some benefits, but in the context of limited health resources (funds and working staff), model results raise the distinct possibility that the benefits from such speed-ups may be less than the associated opportunity cost.

At a more fundamental level, model experimentation with the scenario results suggests that, when viewing the situation from a cost-effectiveness standpoint, it is important to recognize that contact tracing is, in a sense, "self-limiting". This effect is particularly notable with respect to the speed of contract tracing. Initially, rapid contact tracing will quickly reduce the incidence - and, more slowly, the prevalence - of active TB. The more rapid is the contact tracing involved, the sooner will these rates fall. However, whether these gains are achieved sooner or later, the

results tend to converge. This convergence reflects the fact that reduced active TB prevalence in the population limits the efficiency of contact tracing by imposing larger numbers of contacts to be traced before a case of infectious TB can be located. It is entirely possible that finding an active TB cases needs to test hundreds of contacts and require many resources and much money. As a result, regardless of how quickly the prevalence of active TB initially decreases, these initial decreases will slow significantly due to the inefficiency of contact tracing, and will converge to similar levels of prevalence (an effect seen clearly in Figure 4.4 and Figure 4.5).

At this point, the greatest gains in efficiency may be secured by effective risk prioritization, which could allow the given investment in resources to secure significant additional benefit [4]. By contrast, while its dependence on passive methods of diagnosis (in which an infective individual presents for care, typically as a result of symptoms) can allow significant reservoirs of infective cases to circulate in the population, the efficiency of passive diagnosis is less adversely affected by decreasing prevalence of TB than is contact tracing with an important exception being the reduced likelihood that a given physician will recognize TB within a population in which TB circulates at low levels. In this sense, contact tracing can easily become a "victim of its own success", with its initial successful reduction of the prevalence of TB tending to reduce over time the fraction of all active TB cases that are brought in via active diagnosis, and over time shift larger burden of case-finding to passive methods. It is important to emphasize that a situation in which low fraction of case-finding occurs due to active methods of diagnosis, should not automatically be taken as an indictment of the limited efficiency of contact tracing per se. It may well be the case that the presence of contact tracing processed is what has allowed for the lowering of prevalence to the current point, and the maintenance of the system at low levels of active TB. In so doing, it is likely that it will make itself appear "inefficient" to the amount of effort entailed in finding new infectious cases - but this apparent "inefficiency" belies the key role of contact tracing in achieving the current situation.

While the preliminary and aggregate character of the model suggest the need for great caution in interpreting model results in a quantitative fashion, the simulation outcomes suggest some very tentative public health related implications in disease control and policy formulation. This study has focused on contact tracing investigation within the context of an aggregate simulation model. While simulation of contact tracing can very naturally be carried out at an individual level, System Dynamics modeling can assist in the process of policy optimization by running different scenarios to predict potential control outcomes. Modeling within this area raises the potential for maintaining more judiciously chosen programs and enhancing the reliability and validity of deploying new prevention strategies.

However, the model described in this paper exhibits significant drawbacks. Firstly, our implementation of contact tracing ignores the underlying contact patterns within the population. Due

to the nature of the aggregate model, it cannot readily be used to investigate existing or prospective alternative contact tracing protocols, because it deals with people in an aggregate way rather than as individuals. In addition, the model does not capture the population heterogeneity in risk across the population. Among other factors, the assumption of a homogeneous population limits our capacity to investigate targeted contact tracing strategies, which could be highly important for enhancing the efficiency of contact tracing. We are currently working to address this shortcoming using simulations on individual-based model with network structure. An individual-based model would provide a natural vehicle for examining the impact of expediting contact tracing according to the levels of risk factors. A further limitation of the current model is the failure to represent BCG. Although we recognize the potential importance of protection conferred by BCG, we have not represented its impact on preventing TB infection in our model for several reasons. The first of these reasons is limited data: many individuals in the TB database employed lack BCG information recorded in their files. In addition, the BCG data that is available is imprecise: most traced individuals have limited recollection as to whether or not they underwent inoculation with BCG. Finally, recognizing the considerable variability in the epidemiological data on BCG efficacy [51], the degree of efficacy of the BCG vaccine is unclear and controversial.

Chapter 5

Estimating the Effectiveness of Contact Tracing on Tuberculosis Outcomes in Saskatchewan Using Agent-based Modeling

Chapter 4 introduced a System Dynamics model to evaluate the effectiveness of contact tracing in TB control at the aggregate level. However, due to the limitations of aggregate models, the impacts of population heterogeneity and network structure haven't been explicitly examined in that aggregate model. In this chapter, we employ another aggregate System Dynamics TB model [37], and transfer it into a stochastic agent-based TB model with contact tracing strategies to further investigate prospective contact tracing protocols.

5.1 Scheme of the System Dynamics Model of TB Transmission in Saskatchewan

Before proceeding, the prototype of aggregate TB model which we employ in our agent-based TB model is introduced. The System Dynamics TB model in [37] is a well calibrated aggregate TB model with implementation of 2 important risk factors (namely age and ethnicity) and TB preventions for the population in Saskatchewan (e.g. BCG and TLTBI). Figure 5.1 demonstrates the aggregate model structure. Here rectangles denote states in which a person could possibly stay, and solid arrows denote transitions via which a person can update their states. The dotted arrows which are directed into the state are either birth or immigration; a new person (either an immigrant or a new-born) can enter the associated states via such pathways. In addition, those dotted ones directed out of the state denote death either from active TB or reasons other than TB. There are 11 stocks used to characterize the TB related status; each of those stocks is additionally further stratified by age group and ethnicity. In this model, age group and ethnicity constitute risk factors; people in the younger age group experience a higher risk of TB infection given exposure, and First Nations people also hold a relatively higher chance of TB infection and progression. Moreover, age and ethnicity are also considered when representing the TB transmission; the force of infection

is formulated by constructing contact matrices to capture mixing patterns – such as the fact that people (all other things being equal) will tend to have greater levels of contact with others of similar ethnicity or age.

According to Figure 5.1, uninfected individuals can either go to the "BCG vaccinated" stock via vaccine or to the "high-risk latent" stock via acquiring a new TB infection [37]. People who remains in the "BCG vaccinated" stock still can acquire TB infection but with a lower chance compared to those without vaccine. Those who are present in the "high-risk latent" state can develop active TB via primary progression, receive treatment for latent TB infection (TLTBI), or transition to the "low-risk latent" stock. Individuals in the "low-risk latent" stock can return to the "high-risk latent" stock when they are re-infected. Persons who are protected by TLTBI can either progress to active TB or return to the "low-risk latent" stock after the waning of protection. States associated with Active TB are stratified by infectiousness, diagnosis and treatment. In addition, people who were previously treated for active TB but lack current disease will remain in the "latently infected with previous treatment" stock.

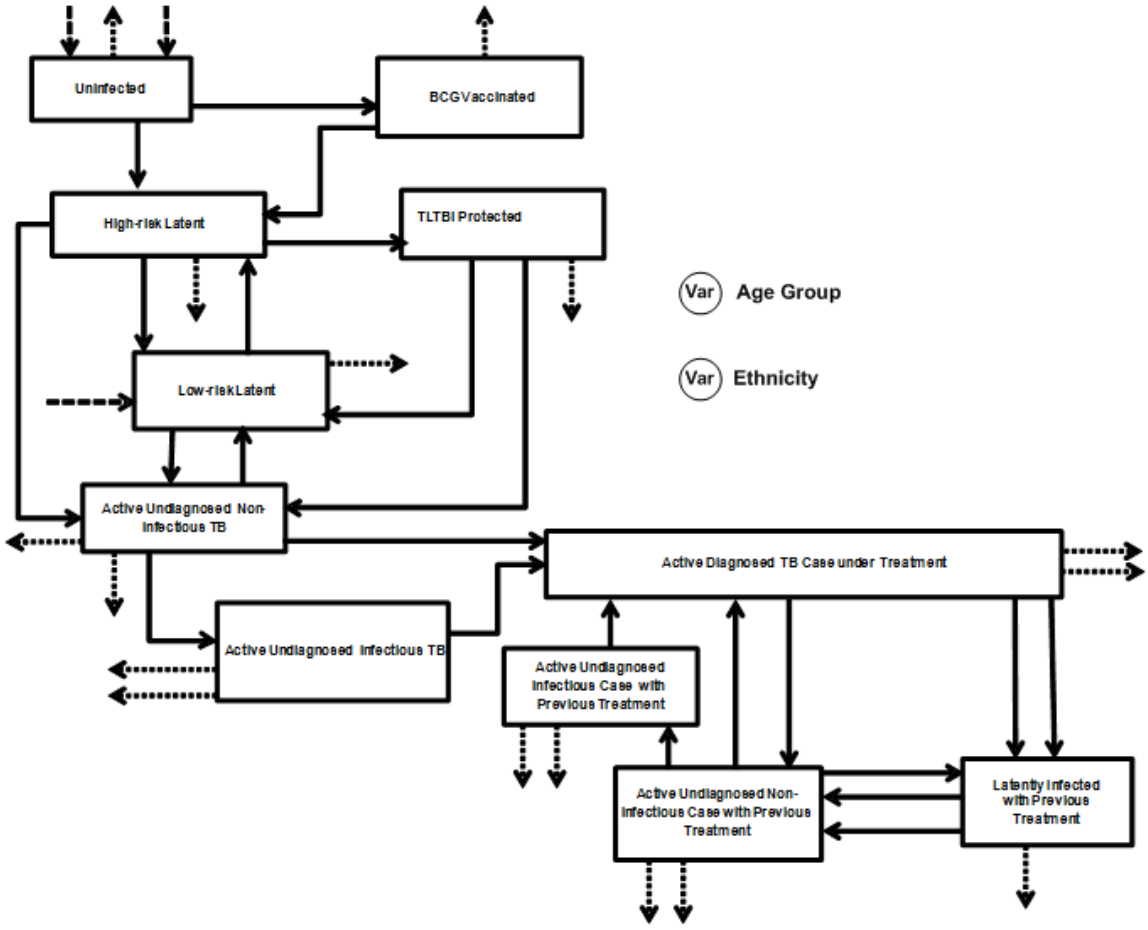


Figure 5.1: Model Structure of System Dynamics TB model in [37]

5.2 Agent-based TB Model Regarding Contact Tracing

We employ the model structure illustrated in Figure 5.1 and TB related parameters for TB transmission components in our agent-based model to characterize TB transmission in Saskatchewan. Contact tracing procedures and statistical data are gathered from the Saskatchewan TB Control Program and supplemental datasets. AnyLogic 6.2.2 Software is used to construct the stochastic agent-based model and conduct simulations. The UML diagram of the overall model structure is depicted in Figure 5.2. The Person class is defined with sophisticated behaviors and numerous characteristics with respect to TB transmission, diagnosis and treatment, aging, contact tracing investigation, and personal status (such as ethnicity, sex, etc.). The MYSQLDB class is created to store the model results and meta-data associated with different scenarios, and the Main class is the root of the simulation which stores all the parameters and statistics of results. The detailed implementation of each component will be illustrated in the following sections.

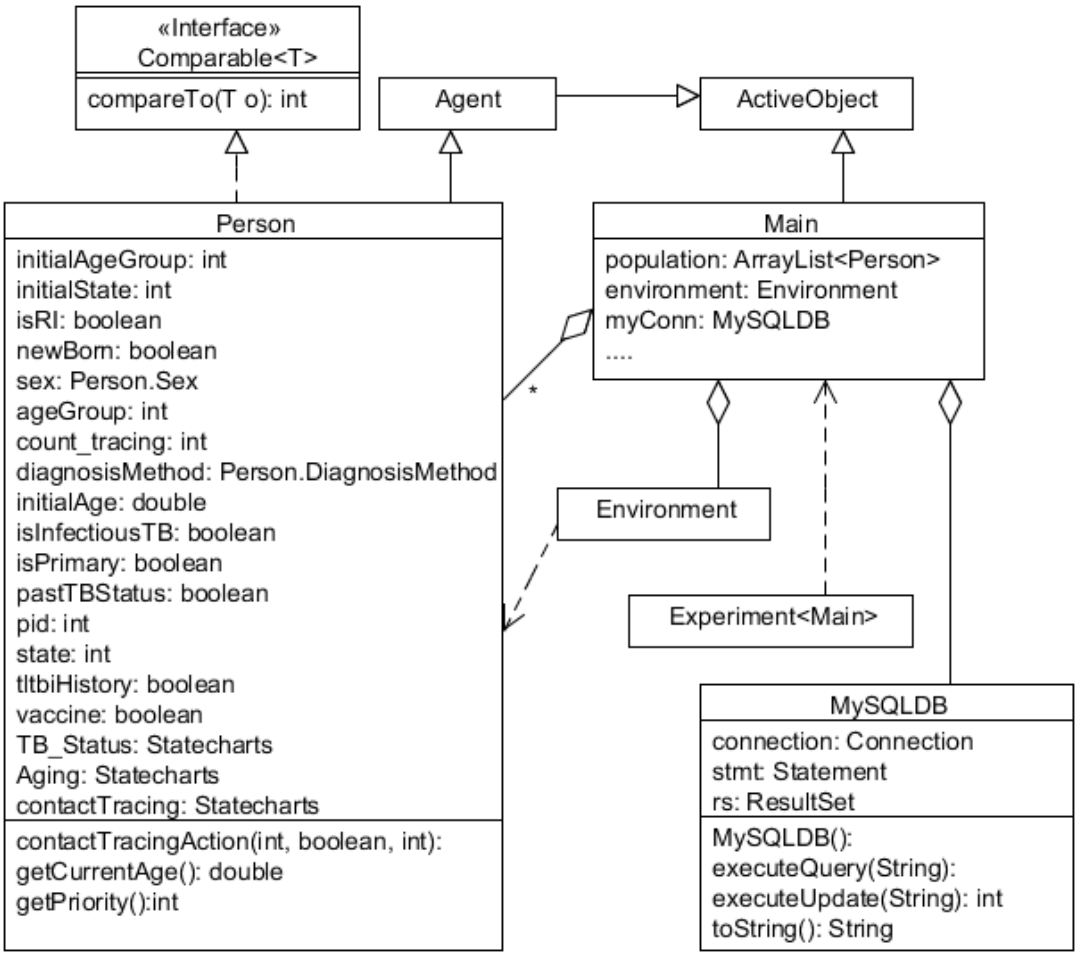


Figure 5.2: Hierarchy of the Overall Model Structure in UML Diagram

5.2.1 Network Structure

The network structure in the context of the TB transmission model is composed of nodes and edges, which represent independent individuals (with different TB related status) and their links with other individuals, respectively. We can consider this network as an undirected graph. The number of edges not only defines the degree of an individual, but also depicts the interactions among the individuals as well as the transmission pathways through which the TB infection can spread. For example, a link between 2 individuals indicates that TB can be transmitted via this connection if one of them is an undiagnosed infectious active TB case.

The underlying contact structure plays a crucial role in infectious disease spread; however, with limited knowledge of the actual network architecture, it is difficult to determine and explore the exact network structures and human behaviors in many cases. Given the advantages of agent-based modeling, we can easily investigate the consequences of assumptions about the underlying network structure of the population, and represent the individuals within a network environment. Such representation provides the opportunity to explore the questions regarding how TB transmission and its control strategies are affected by the underlying social network structure.

We investigate the impact of assuming that the underlying contact network adheres to a random, scale-free and small world network structure. By simulating contact tracing on such an underlying network, we can obtain a contact tracing network over the exact contact network; estimating such networks might help understand the difference between the hypothetical contact network (for which we don't have knowledge) and the contact tracing network (about which we collect information during the contact tracing process).

In many studies, the network models for disease spread assume a closed population – that is, they assume a fixed number N of nodes without modifying the N. In contrast, the social networks in the real world represent open systems and are subject to dynamics in the form of removal and insertion of nodes and edges. In our model, the contact network is a dynamic one where the population is open and changed by birth and death over time. Although it would obviously be more realistic to have breaking and forming of relationship due to reasons other than birth or death, without information collected at such detailed level, we decided to keep the links amongst individuals static for the duration of the simulation, with the exception of those edges changing because of births and deaths. The characteristics of different network architectures are maintained in such an open population over time by keeping the average connections per individual stable for all of the simulated network topologies, and by employing preferential attachment based on the node's degree for the scale-free network.

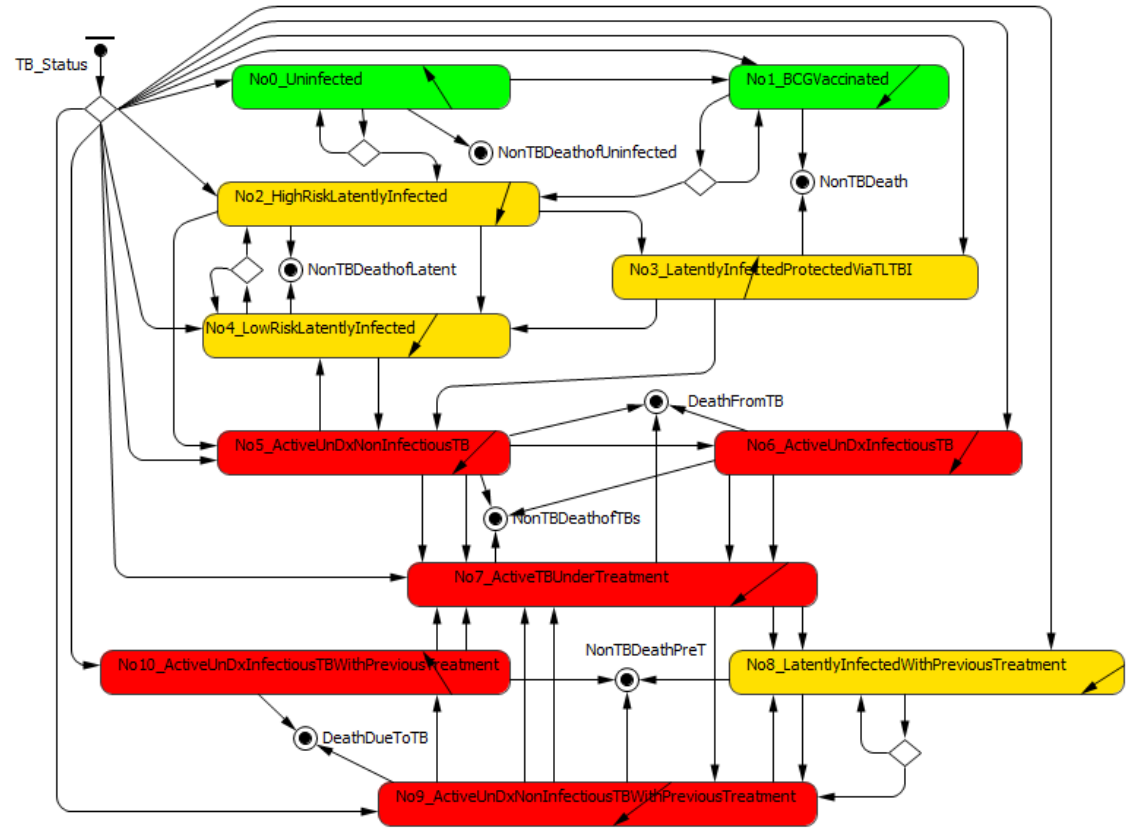


Figure 5.3: TB Transmission Statechart in Agent-based Model

5.2.2 TB Transmission Statechart

Each individual in the network environment has 3 statecharts which represent TB status, aging and contact tracing. Figure 5.3 shows the TB status statechart. Similar with the stocks in the System Dynamics TB model, there are 11 state in the TB status statechart: "Uninfected" (U); "BCG Vaccinated" (V); "High Risk Latently Infected" (E_h); "Latently Infected Protected Via TLTBI" (P_{tltbi}); "Low Risk Latently Infected" (E_l); "Active UnDx Noninfectious TB" (TB_{uninf}), *i.e.* individuals who are currently diagnosed; "Active UnDx Infectious TB" (TB_{inf}), *i.e.* individuals with infectious active TB who can transmit infection, but who are currently undiagnosed; "Active TB Under Treatment" (R); "Latently Infected With Previous Treatment" (E_{ex}); "Active UnDx Noninfectious TB With Previous Treatment" ($TB_{ex,uninf}$); "Active UnDx Infectious TB With Previous Treatment" ($TB_{ex,inf}$). The transitions between states are illustrated in Figure 5.3, and the details of the transitions are presented below.

- *Vaccination* U(Uninfected) $\rightarrow$ V(BCG Vaccinated). BCG is a vaccine against TB. The BCG rate per year (stratified by ethnicity) in this agent-based model is derived from the historical BCG vaccination count data in Saskatchewan and the aggregate TB transmission model [37].

- *Infection* U(Uninfected) $\rightarrow$ E_h(HighRiskLatentlyInfected). In our model, the chance of acquiring TB infection depends on a contact process (operationalized via infection message receiving) with neighboring nodes (infectious nodes in particular) in the network, and the probability of having been infected (β) given such exposure. The term βc (the same one used in [37]), stratified by ethnicity, is used when calculating the force of infection (e.g. βc for registered Indian is 38.0299, βc for non-registered Indian is 28.9106), these values are derived from calibration with historical data in Saskatchewan. In addition, the relative risk (RR) of TB infection for different age categories and ethnicity is also introduced to capture individual heterogeneity in terms of TB infection (e.g. RR of TB infection for age between 0 and 4 is 11.1094, RR of TB infection for age between 5 to 9 is 14.7316, and it equals to 1 for the rest age groups), these values are obtained via calibration.

- *Infection* V(BCG Vaccinated) $\rightarrow$ E_h(High Risk Latently Infected). This transition is analogous to the transition $U \rightarrow E_h$, and it is a network-dependent process. The difference between this transition and the infection transition $U \rightarrow E_h$ extends from the relative risk of acquiring TB infection. Given TB exposure, the relative risk of been infected is lower for vaccinated individuals (especially children) compared with those without vaccine, BCG vaccination significantly decreases the risk of TB by 50% on average, and vaccination of newborns and infants reduces the risk even more [7, 8]. In our model, the RR of TB infection given BCG is implemented as 1 for each age group. Compared with the RR of TB infection absence of BCG, it is lower for the children under 9 years old, but is maintained the same for the individual above age 9.

- E_h(High Risk Latently Infected) $\rightarrow$ E_l(Low Risk Latently Infected). The model divides latently infected persons into 2 categories (a high risk group versus low risk group) regarding different level of risk with respect to developing active TB. Such classification of latently infected individuals gives a better representation of TB progression than does assuming a uniform risk of progression to active TB among all latently infected individuals. According to the facts regarding TB [43, 45, 54], around 10% of infected people will develop Active TB in their lives; approximately 5% of them will develop Active TB in the first 2 years after acquiring TB infection; the remaining 5% will develop Active TB later in their lives. This transition applies a mean time of 2 year delay to distinguish different levels of risk of developing Active TB.

- *Primary Progression* E_h(High Risk Latently Infected) $\rightarrow$ TB_{uninf}(Active UnDx Noninfectious TB). This is a network-independent transition, and the natural process of primary progression is captured via active TB progression after a shorter latency period in state E_h (less than 2 years since infection).

- *Re-infection* E_l(Low Risk Latently Infected) $\rightarrow$ E_h(High Risk Latently Infected), E_{ex}(Latently Infected With Previous Treatment) $\rightarrow$ $TB_{ex,uninf}$(Active UnDx NonInfectious TB With Previous Treatment). A small fraction of the individuals in state E_l and E_{ex} are represented as having been re-infected (e.g. reinfected by another string of mycobacteria), and this transition is dependent on network and on the probability acquiring infection upon contact with an infectious TB case. The proportion of latently infected individuals susceptible to exogenous re-infection is 25% in our model [20].

- *Treatment for Latent TB* E_h(High Risk Latently Infected) $\rightarrow$ P_{tltbi}(Latently Infected Protected Via TLTBI). This is a contact tracing related transition. Investigation of contacts of the traced active TB case will trigger treatment for latent TB infection via message sending and receiving. Contacts who are found in the state E_h with age less than 35 are treated as being eligible for this treatment.

- P_{tltbi}(Latently Infected Protected Via TLTBI) $\rightarrow$ E_l(Low Risk Latently Infected). This transition represents the delay between state P_{tltbi} and the state E_l. It is estimated that the mean time of protection given by treatment for latent TB (TLTBI) is 7 years [37].

- P_{tltbi}(Latently Infected Protected Via TLTBI) $\rightarrow$ TB_{uninf}(Active UnDx Noninfectious TB). The proportion of protected population with TLTBI at risk of developping active TB is derived as 32% [37].

- *Natural Recovery* TB_{uninf}(Active UnDx Noninfectious TB) $\rightarrow$ E_l(Low Risk Latently Infected), $TB_{ex,uninf}$(Active UnDx Noninfectious TB With Previous Treatment) $\rightarrow$ E_{ex}(Latently Infected With Previous Treatment). These transitions represent natural recovery from active TB, and the calibrated rate of natural recovery from active TB is 0.1 per person per year for registered Indian, and 0.1638 per person per year for non-registered Indian.

- *Reactivation* E_l(Low Risk Latently Infected) $\rightarrow$ TB_{uninf}(Active UnDx Non Infectious TB). Low risk infected individuals can reactivate and progress to active TB (known as endogenous reactivation) via this transition.

- *Progressing to Infectious State* TB_{uninf}(Active UnDx Noninfectious TB) $\rightarrow$ TB_{inf}(Active UnDx Infectious TB), $TB_{ex,uninf}$(Active UnDx Noninfectious TB With Previous Treatment) $\rightarrow$ $TB_{ex,inf}$(Active UnDx Infectious TB With Previous Treatment). Individuals in TB_{uninf} or $TB_{ex,uninf}$ can become infectious TB with a mean delay time of 1.29 years.

- *Passive Diagnosis of Noninfectious Cases* TB_{uninf}(Active UnDx Noninfectious TB) $\rightarrow$ R(Active TB Under Treatment), $TB_{ex,uninf}$(Active UnDx Noninfectious TB With Previous Treatment) $\rightarrow$ R(Active TB Under Treatment). The passive diagnosis process is represented with a delay

with a mean time of 2.9 years for noninfectious case without previous TB, while the mean time of diagnosis for those with previous TB is adjusted by a coefficient of 0.5.

- *Active Diagnosis of Noninfectious Cases* TB_{uninf}(Active UnDx Noninfectious TB) $\rightarrow$ R(Active TB Under Treatment), $TB_{ex,uninf}$(Active UnDx Noninfectious TB With Previous Treatment) $\rightarrow$ R(Active TB Under Treatment). These transitions are triggered by contact tracing procedures via message interactions within the contact network. Contact tracing can help identify undiagnosed active cases. As represented, contacts who are found with active TB by contact tracing are diagnosed actively. Contact tracing might follow on the diagnosis procedure to further investigate potential infected contacts or identify the source of infection; however, the follow-on activity depends on protocol and underlying objective of the contact tracing investigation.

- *Passive Diagnosis of Infectious Cases* TB_{inf}(Active UnDx Infectious TB) $\rightarrow$ R(ActiveTB Under Treatment), $TB_{ex,inf}$(Active UnDx Noninfectious TB With Previous Treatment) $\rightarrow$ R(Active TB Under Treatment). Similar to the diagnosis process of noninfectious active TB cases, a delay is used to represent these 2 diagnosis transitions. Compared with the mean time of diagnosis of noninfectious TB cases, the mean time of diagnosis of infectious TB cases without previous TB, which is treated as 0.1 year for non-First Nations people and 0.22 year among First Nations individuals, is much shorter due to the fact that individuals at this stage are more likely to have obvious symptoms and seek health care examination. In the default contact tracing protocol, these transitions also trigger the contact tracing to further investigate the next generation of infected individuals.

- *Active Diagnosis of Infectious Cases* TB_{inf}(Active UnDx Infectious TB) $\rightarrow$ R(Active TB Under Treatment), $TB_{ex,inf}$(Active UnDx Noninfectious TB With Previous Treatment) $\rightarrow$ R(Active TB Under Treatment). The occurrence of these transitions are activated by contact tracing, and they are network-dependent. Depicting the diagnosis of infectious cases by contact tracing, these 2 transitions are not only the result of contact tracing, but also another starting point of contact tracing investigation.

- *Treatment* R(Active TB Under Treatment) $\rightarrow$ E_{ex}(Latently Infected With Previous Treatment). There are 2 transitions which remove nodes from state R to state E_{ex}, and they are triggered by the rate of annual likelihood of treatment completion and the rate of treatment default occurring with latent TB (taking into account the fact that some of the relapsed individuals still have non-infectious active TB).

- *Relapse* E_{ex}(Latently Infected With Previous Treatment) $\rightarrow$ $TB_{ex,uninf}$(Active UnDx Noninfectious TB With Previous Treatment), R(ActiveTB Under Treatment) $\rightarrow$ $TB_{ex,uninf}$(Active

UnDx Noninfectious TB With Previous Treatment). Removal of a person from state R to $TB_{ex,uninf}$ refers to the relapse happening among the persons in state R. Prior to the Directly Observed Treatment (DOTS), there was a high relapse rate among the active TB cases due to failure of treatment or low compliance. Individuals within state E_{ex} can also relapse to active TB.

- *Birth and Death.* Birth and death can change the network structure by forming or breaking the connections of the nodes. The birth and death rate in our model are applied the same as those in [37] which are derived from historical Saskatchewan birth count data and historical death rate.

Most of the rates used in the transitions above vary according to different age group or ethnicity to provide more accurate measure of the TB progression and its interaction with risk factors.

5.2.3 Contact Tracing Statechart

Given the advantages of individual-based modeling techniques, we can implement contact tracing at an individual level with more detailed representation of the contact tracing procedure in the real world, which we hope will eventually provide a more robust and accurate estimation of the effectiveness of contact tracing.

Contact Tracing Procedure

Starting from an infectious TB case, diagnosed either actively through previous contact tracing or passively due to illness and the presence of obvious symptoms, a number of potential contacts will be identified by contact tracing investigation. Contacts whose records demonstrate that they are associated with previously positive skin tests (also known as a "TST test" or "Mantoux test"), usually will not be re-examined again unless they are the contacts of a primary TB case (in which case they may be examined for the sake of reverse contact tracing), since they are known as tuberculin reactors. The remaining contacts are typically notified via a letter that they are sought for examination. The objective is to have 95% of the contacts examined within 30 days since the TB case is identified [43], and to maintain a low level of loss to follow-up in both skin test and clinical review.

Figure 5.4 depicts the flow chart with respect to contact tracing and the decision making processes. Usually there are 3 steps for investigating a contact, namely a first skin test, a potential second skin test, and clinical review. However, the examination procedure for a particular individual does not necessarily transition through all these 3 steps; the next step test depends highly on the previous test outcome as well as on the personal history information. For example, a contact doesn't need a second skin test if we know he or she is TST positive in the first test. Treatment for

either active TB or latent TB are assessed by the clinical review, which includes standard testing and diagnosis procedures, such as X-ray, culture and smear test. Contacts of primary TB cases – even those previously skin test positive – will be examined to infer the source of infection. Loss of contacts follow-up can happen in each step of the contact investigation procedure.

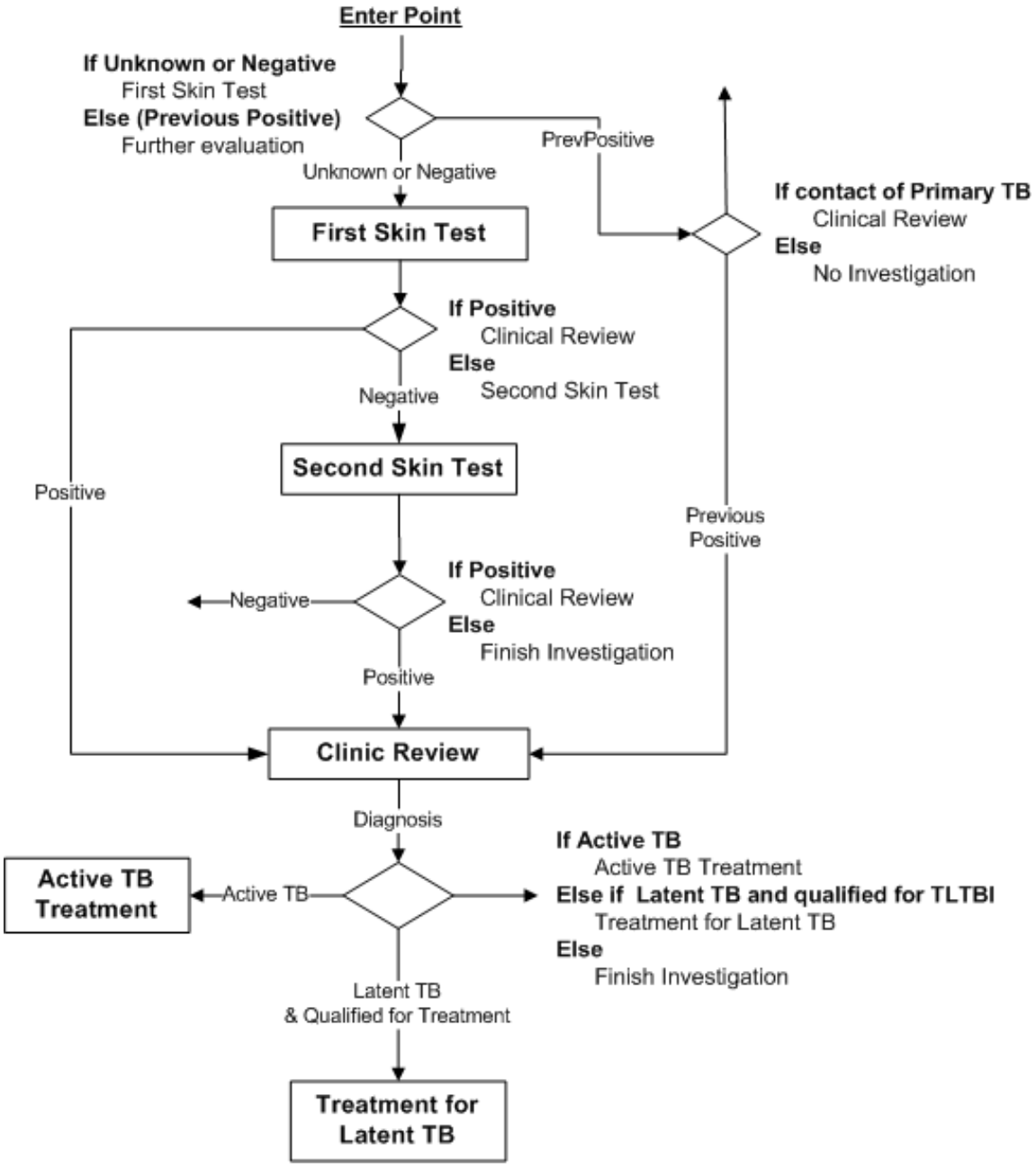
According the contact surveillance information from the years 2004 to 2008 in Table 5.1, the objective of investigating 95% of the contacts within 30 days is not well accomplished. The cross-year average percentage of the contacts examined following the current surveillance procedure within 30 days, 90 days, 180 days and 365 days of identifying the active TB cases are 22.8%, 73%, 94.2%, 99.4%, respectively. Obviously, there is still great effort required to meet the stated scheduling goals of contact tracing investigations.

Table 5.1: The Cumulative Percentage of Contacts Examined at Each Interval (Contact Surveillance of Saskatchewan from 2004 to 2008, obtained from Saskatchewan TB Control)

Year/Days	**30**	**90**	**180**	**365**
2004	21	81	97	99
2005	22	71	95	99.8
2006	24	77	99	100
2007	31	78	96	100
2008	16	58	84	98
Average	**22.8**	**73**	**94.2**	**99.4**

Figure 5.5 illustrates a scenario of investigating an infectious active TB case, and the overall information collected from Saskatchewan TB control regarding the lost of follow-up and tracing procedure. Based on the contact tracing outcome from Saskatchewan TB Control between years 2001 to 2006, roughly 15% of the contacts were unavailable for skin testing. Among those who were skin tested, 51.54% of contacts with negative result in their first skin test were lost follow-up in their second skin test and a potential clinical review, while 38.92% of those with positive results missed their clinical review for active TB disease. Such loss might significantly undermine the effectiveness of contact tracing, and the analysis of the impact of this loss to follow-up merits further investigation.

Contact tracing is currently limited to two categories of the clients: Those who have infectious TB or primary TB. Suppose an infectious active TB cases is diagnosed such as that shown in Figure 5.5; the nurse will create a contact list for the case, and ask for his or her contacts in the 30 days before the diagnosis. A contact list sorted by the amount of exposure time is recommended. The objective of tracing infectious TB cases is to find infected people, people with active disease and people with infectious disease. The main potential infected contacts for this case include preschool



Note: Contact tracing for primary case is different. Since the target is to find the source of infection, once the presumed source is found, contact tracing is discontinued. If the contact is not positive in a TST test, then no need for the 2nd skin test.

Figure 5.4: Contact Tracing Protocol and Tests

children, persons who breathed the same indoor air as the case as well as contacts of the case within 30 days before diagnosis. Usually a letter of notice will be mailed to the contacts, then contacts are scheduled to finish their skin test first, those whose skin tests are significant will be brought into the system for further assessment for disease. There is variations in the duration between case diagnosis and contacts investigation, and relevant descriptive statistics is shown in Table 5.1 and Figure 5.5. The diagnosis of persons with active disease by contact tracing is classified as active diagnosis. However, the procedure for tracing primary TB is somewhat different from that of tracing infectious cases. The target for this process – termed "Reverse contact tracing" – is to identify the source of infection, and the focus is contacts age 15 years and older as well as contacts during 30 days period to diagnosis. When the source is believed to be known up front, the contact skin testing trace is not required. In addition, contact skin testing will be discontinued when the source is found. For example, the second skin test is not necessary if the first test is not significant because these contacts are not the source (they were evidently not infected and not suffering from active TB at the time of the contact).

Contact Tracing Implementation

In order to capture the principles of contact tracing in conjunction with TB progression, a contact tracing statechart is created in the Person class to represent details of the protocol of investigating contacts. Figure 5.6 illustrates the statechart of contact tracing within our agent-based model, and it interacts with the TB transmission statechart to simulate the investigation procedure. Here, the diagnosis related transitions in TB transmission statechart assist in initializing a prioritized queue of contacts for a diagnosed active TB cases (mainly primary TB and infectious TB case), then further assessment on contacts will be conducted (such as eligibility for investigation, previously investigated or not); afterward message of notice will be sent to the qualified contacts. Usually only a fraction of the contacts will be named by the TB case. The priority score of each contact is calculated based on the RR of TB infection by age, ethnicity and number of times an individual has been reported as a contact. A Higher score refers to higher risk as well as higher ranking in the queue list of the contacts to be interviewed.

Individuals can be in one of five states with regards to their contact tracing status which is shown in Figure 5.6, namely "PotentialContact", "ReceivedNotice", "Mantoux1stSkinTest", "Potential2ndSkinTestandclinicalReview" and "PreviousPositive". Initially, individuals are sitting in the "PotentialContact" state. Upon receiving the contact notice messages, the contacts will go to "ReceivedNotice" state. Then the "Mantoux1stSkinTest" state represents a situation where people are currently waiting for their 1st skin testing. While "Potential2ndSkinTestandClinicalReview" state represents the clinical review procedure for those positive in the 1st TST test; for those who are negative in the first TST test, it stands for the second skin test and a potential clinical review

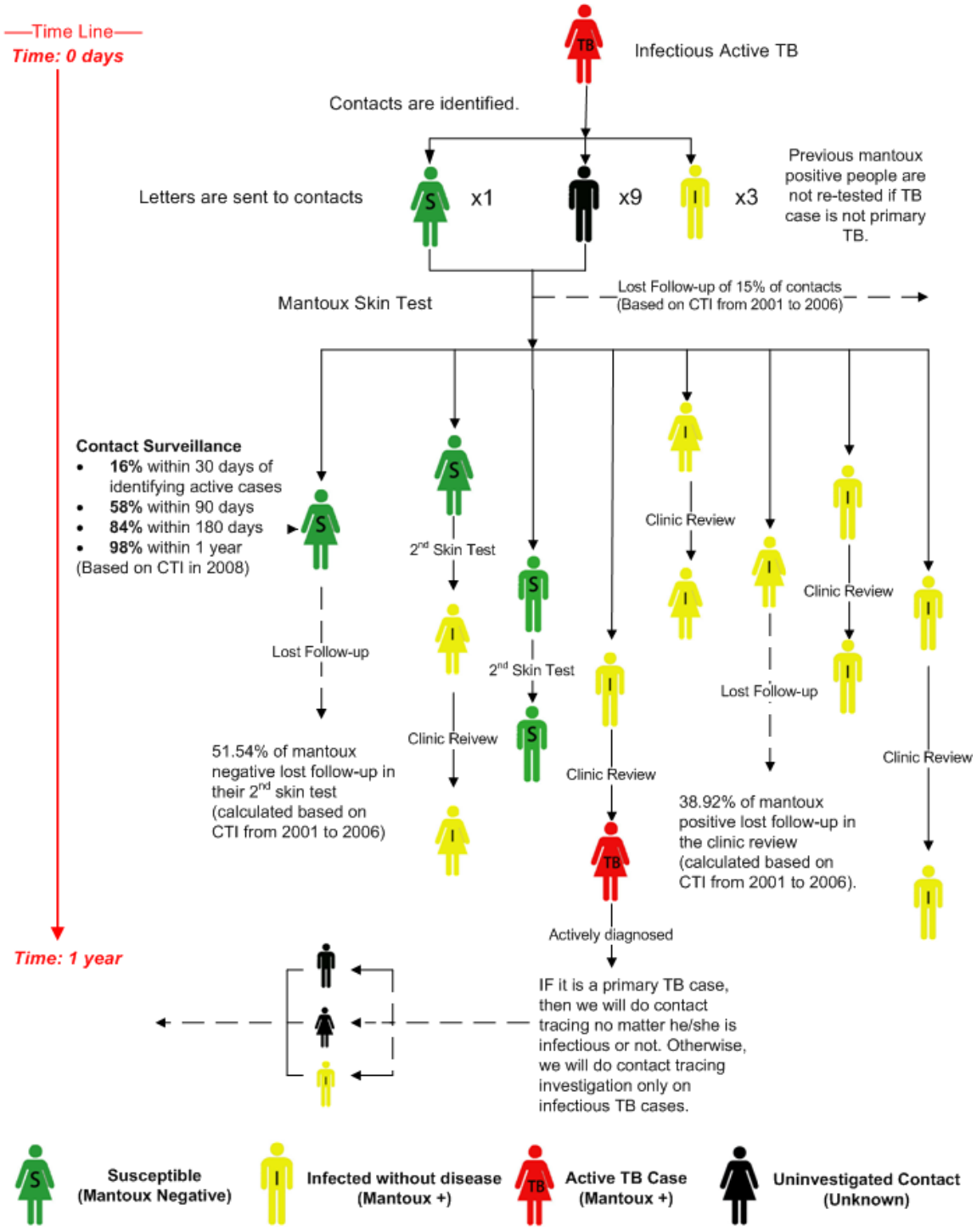


Figure 5.5: Contact Tracing Investigation(CTI) Procedure and Facts

procedure. According to the historical data, we lost 4.4% contacts between the second skin test and clinical review. Since the lost is modest, we assume there is no lost between the second skin test and clinical review to simplify the implementation without losing too much information. "PreviousPositive" is a state for the past TB cases and those previously investigated with positive skin test results. The implementation of the transitions in the contact tracing statechart is as follows.

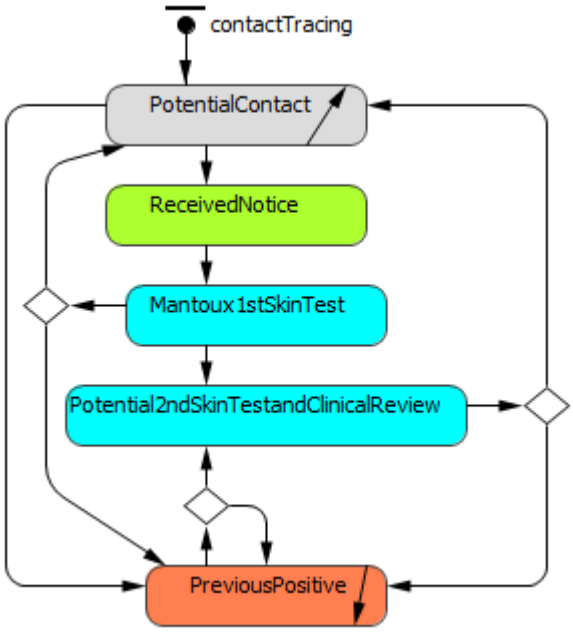


Figure 5.6: Contact Tracing Statechart in Person Class

- "PotentialContact" → "PreviousPositive". This transition is triggered when an active TB case is diagnosed and ready for treatment. An active TB case is known as tuberculin reactor and shows positive results in the skin test.

- "PotentialContact" → "ReceivedNotice". This is a message and network dependent transition. When contact tracing is enabled, diagnosis of an active TB case might trigger this transition by investigating his or her contacts. Notice messages are sent by the index case via the contact network.

- "ReceivedNotice" → "Mantoux1stSkinTest". This is a timeout transition. The speed of investigating contacts in the model is based on information regarding contact surveillance speed in 2008, shown in Table 5.1.

- "Mantoux1stSkinTest" → *Branch* → "PotentialContact". Lost follow-up after the first skin test is captured via this transition. Negative contacts can follow this transition and go back the "PotentialContact" state, since they are uninfected and fully eligible for future investigation.

- "Mantoux1stSkinTest" → *Branch* → "PreviousPositive". This transition represents loss of contacts with positive results in their first TST test, and they are known as previous positive even just finishing the first skin test. We assume that these individuals can be investigated again only if they are nominated as a contact of a primary case for reversing contact tracing.

- "Mantoux1stSkinTest" → "Potential2ndSkinTestandClinicalReview". A delay between the first skin test and clinical review is represented in this transition.

- "Potential2ndSkinTestandClinicalReview" → *Branch* → "PotentialContact". If a contact shows negative results again in the second TST test, then the investigation is done and he can go back to "PotentialContact" state.

- "Potential2ndSkinTestandClinicalReview" → *Branch* → "PreviousPositive". If an individual shows significant result in the potential second skin test or the clinical review, then he or she will go to "PreviousPositive" state afterward. Treatment for Latent TB infection and active diagnosis of new TB cases are assessed in this transition as well.

- "PreviousPositive" → "Potential2ndSkinTestandClinicalReview". This transition is used in reverse contact tracing only. Those who are previously skin test positive will be reviewed again to assess the source of infection of a primary TB case. This is a message dependent transition with chance of losing follow-up in the clinical review.

5.2.4 Parameterization

The TB model is calibrated to fit the TB epidemic in Canadian province of Saskatchewan. Key TB related parameters applied in our model are summarized in Table 5.2. TB transmission parameterization was based on updated estimates from calibration of the model reported in [37], while the contact tracing central parameters are from Saskatchewan TB Control (as presented in Table 5.1, Figure 5.5).

Since age and ethnicity can contribute to the risk of TB infection and progression, the relative risk (RR) of infection and primary progression are also calibrated and introduced. For example, children and First Nations people possess higher risk in TB infection. In addition, the guidelines for contact tracing in Saskatchewan also suggest that children under age 5 are assumed as infected contacts if they are exposed, although they might look and feel well. The parameters estimated via calibration assist in reproducing the historical behavior patterns which eventually enhance the confidence regarding model reliability.

The Relative risk (in terms of the number of times an individual has been reported as a TB contact) doesn't affect the TB diffusion directly, instead it is applied when assessing the priority of contacts in the investigation process, which plays an indirect role in shifting the behaviors of the

Table 5.2: Major Model Parameters and Estimates

Parameters	**Value**	
	Registered Indian	*Non Registered Indian*
βc (Average number of infections an infectious person causes within the course of their illness in a fully susceptible population absent intervention)	38.0299	28.9106
Relapse to active TB rate	0.0062	0.001
Natural Recovery Rate	0.1	0.1638
Mean Time Until Discovery of Undiagnosed Infectious TB	0.2221	0.1
Primary Progression Rate	0.0142	0.0109
Mean Time Until developing infectious TB	1.29	
Mean Time Until Discovery of Undiagnosed noninfectious TB	2.9012	
Reactivation Rate	0.001562	
Previously Treated Death Rate Coefficient	5	
Average Connections per Person in the Network	60	

system. The most important RR values implemented in our model are depicted in Table 5.3. The first 5 are taken from an updated (but not yet published) calibration of the model reported in [37], and the balance are calculated from literature [3].

Table 5.3: Major RR Parameters and Estimates in Our Model

Relative Risk (RR)	**Values**
RR of TB Infection for age 0 to 4	11.1094
RR of TB Infection for age 5 to 9	14.7316
RR of primary progression for age 0 to 4	2.3495
RR of primary progression for age 5 to 9	1
RR of primary progression for age 10 to 14	1.0444
RR of TB Infection while being reported as a contact 2 times	2.1481
RR of TB Infection while being reported as a contact 3 times	2.7396
RR of TB Infection while being reported as a contact 4 or more times	3.9381

5.3 Experiment Design

Dynamic models are designed to understand the emergent system behavior associated with counterfactual scenarios. Our individual-based TB transmission model with contact tracing serves as a tool to simulate the impact of different control scenarios and their potential impact on disease prevention, especially among Aboriginal peoples. Each scenario is dependent on a set of assumptions with respect to different network characteristics, contact investigation target populations, and prioritized tracing protocols. Our scenarios are primarily seek to address two questions.

- 1. Given the current situation with high loss to follow-up in the contact tracing program, if we delivered an intervention capable of reducing the loss to follow-up to some ideal level, what would be the impact on TB incidences rate and prevalence of TB infection over time? Upon the current data collected from contact tracing investigation, there are roughly 30% to 40% lost follow-follow-up in the skin test and clinic review. Such loss directly reduces efficiency in delivering treatment for latent TB infection for high risk latently infected people, as well as in identifying active diagnosis of TB cases. Given the current situation, it is beneficial to understand how much we have lost and how much we can improve by applying additional guidelines to reduce such loss.

- 2. Given a situation without explicit prioritization of contact tracing, what would happen to TB incidence rate and TB infection prevalence if we were to devote enough effort to perform

prioritized contact tracing that places an elevated priority on high risk people? Since risk factors of TB infection given exposure, such as age and ethnicity, have been well studied and validated, it is interesting to explore the opportunities brought in by distinguishing and investigating individuals according to their risk of acquiring TB infection. In addition, some studies, which applied social network analysis on contact tracing network [3], suggest that the number of times a contact has been reported in the contact investigation program is another important indicator of likelihood of TB infection; and taking such a consideration into account when deciding a contact's priority in the process of investigation might yield a favourable disease control outcome. Within our model's context, we'd like to examine the assumption of investigating contacts prioritized by number of times a contact has been reported.

5.3.1 Population Size Selection

To address the questions above, we simulate the system describing a stylized Aboriginal community in the Canadian province of Saskatchewan that is treated as having an initial population size of roughly 15,000 individuals (90% Registered Indian) for a period of 20 years.

Because we are assuming a stylized population – rather than mimicking a particular region – it will be desirable to examine the impact of assumed population size on simulation outcome. It bears noting that because of the quantized nature of infection (in the model – as in the real world – one doesn't have 0.25 infectives) and non-linear interactions within the model, it should not be assumed that model results will scale linearly with population size and structure.

Due to the potentially significant impact of network structure on infection transmission dynamics, we need to specify a particular network structure within which we will examine the model behavior (measures of TB spread in particular). While an investigation of the underling TB-relevant contact network structure extant in Saskatchewan is an important line of research, it lies outside the scope of this thesis. For this set of scenarios, we assume the underlying network structure for the population is scale-free since this gives the highest level of heterogeneity in terms of connections compared with other assumptions of the network types in our model.

Assuming a scale-free network, we are with in some sense making a "worst-case" assumption regarding the potential for TB spread given the higher degree of the variability of this spread. In addition, the contact tracing program is disabled in these scenarios. Cumulative incidence cases for a period of 20 years is used to measure the impact of population size on TB spread. The fraction of the Registered Indian in the community is maintained as 90% in all the realizations. The population size for each scenario is 5,000, 10,000, 15,000 and 20,000 respectively. Each scenario is simulated for 30 realizations, each associated with a different random seed, and the results of the simulation is demonstrated in Table 5.4.

Table 5.4: Scenarios Regarding Population Size within Scale-free Network without Contact Tracing Investigation

	Cumulative Incidence Cases (Active TB)				
Population Size	*Mean*	*Max*	*Min*	*Std. Deviation*	*Coefficient of Variation(C.V.)*
5,000	136.74	263	67	49.50	0.362
10,000	272.81	388	162	65.25	0.239
15,000	423.81	658	293	76.33	0.180
20,000	535.63	736	389	97.34	0.182

Figure 5.7 shows the mean and standard deviation for scenarios regarding different population size. Within the range examined, when the simulated population size is scaled by a specific factor, the mean of the cumulative incidence cases is also scaled by the same coefficient, while the standard deviation of all the scenarios scales sublinearly.

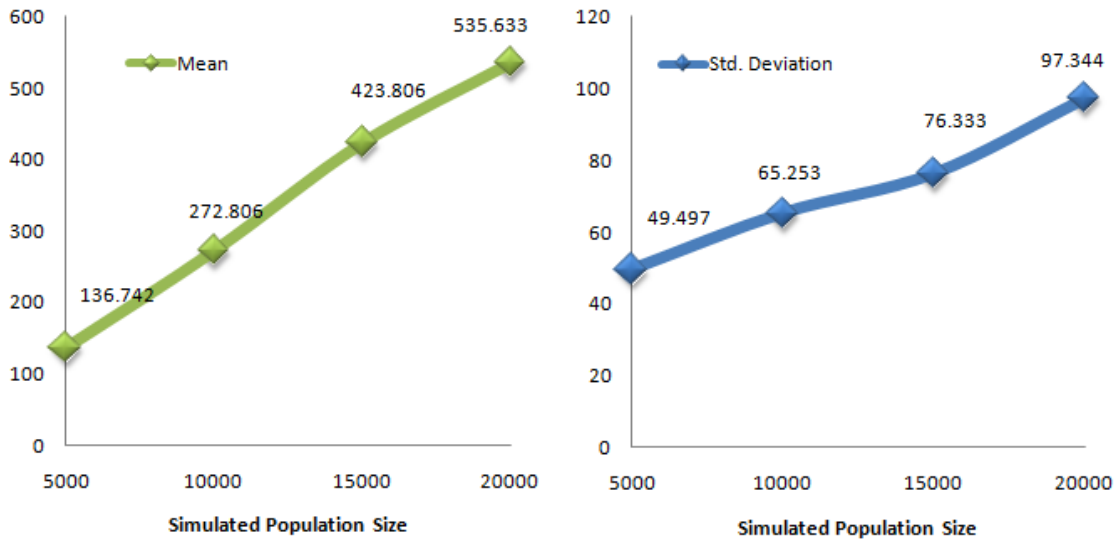


Figure 5.7: Mean and Standard Deviation for Scenarios with Different Population Size without Contact Tracing

Because the above scenarios have different mean and deviation, in order to compare them in a meaningful way to determine the impact of population size and have a simulated population size stable enough to conduct further investigation of contact tracing without heavy fluctuation that might alter the diffusion behavior, we introduce the coefficient of variation (C.V.), where $C.V. = \frac{Std.Deviation}{Mean}$, to estimate the fluctuation level of incidence cases for scenarios with different

population size. This ratio cancels the units of the variable, and smaller C.V. indicate that the measured values are less dispersed. Results shown in Figure 5.8 suggest that the $C.V.$ decreases for larger population size, and reaches some equilibrium level when the population size is 15,000 or bigger.

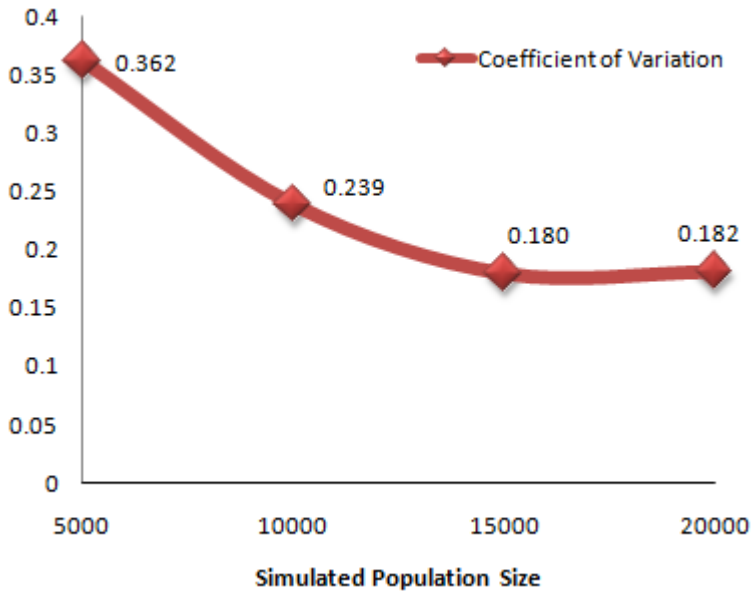


Figure 5.8: Coefficient of Deviation for Scenarios with Different Population Size

In summary, alternating the population size in conjunction with a highly heterogeneous network structure does not alter the behavior of the system in an unpredictable way. When the simulated population size increases, the measure of incidence cases becomes more stable with less fluctuation. Based on these experiments, we posited that having a population size of 15,000 would appear appropriate enough to capture the characteristics of different scenarios.

5.3.2 User Interface for Scenarios Exploration

An interface was developed within this model to make the simulation easier and efficient for the user to simulate desired scenarios. Figure 5.9 provides a snapshot of the user interface designed for customizing the settings for different scenarios, helping to facilitate conducting experiments with different sets of parameter values. Our user interface provides flexibility in exploring the experiments with respect to different network types (random, small world, scale-free) and network-related settings (average connections per agent), contact tracing targets (all active TB cases, only infectious cases, infectious cases and primary TB cases), the criterion to use to set the relative priority for investigating contacts (ethnicity, age, number of times been reported), the fraction of contacts to be interviewed(with lower priority contacts never being sought for follow-up), and population size

Figure 5.9: Contact Tracing Simulation Scenarios Design

(specified separately for each ethnicity as a fraction of the real population in Saskatchewan).

5.3.3 Baseline Scenarios

Before exploring future policy alternatives, the first set of scenarios is the "Baseline scenarios" that reflects the status quo without sophisticated and advanced disease control programs. The parameters for these baseline scenarios maintain their default values. The baseline scenarios, serving as the reference scenarios, will be compared with a set of customized scenarios regarding different research questions to evaluated and explore the optimized disease control programs.

Table 5.5: Baseline Scenarios Settings without Contact Tracing Investigation(CTI)

Id	Network Type	CTI	Tracing Target	Lost follow-up	Tracing Fraction
R_0	**Random**	**Disabled**	**N/A**	**N/A**	**N/A**
W_0	**Small World**	**Disabled**	**N/A**	**N/A**	**N/A**
S_0	**Scale-free**	**Disabled**	**N/A**	**N/A**	**N/A**

In our model, the default settings for baseline scenarios are established as different network assumptions (random, small world, scale-free network respectively) without contact tracing investigation program for 20 years' simulation, seen in Table 5.5. Because of absence of contact tracing program, active diagnosis and treatment for latent TB infection are not applicable in the baseline scenarios.

5.3.4 Alternative Scenario Definitions

To explore the alternative futures regarding the 2 questions listed above, a set of experiments will be designed under each assumption of the underlying network structure with different schemes of contact tracing. To account for the stochastic variability in results, each scenario consists of 30 realizations with different random seeds, and the time window for each realization is 20 years.

Scenarios Assuming a Random Network

In addition to the baseline scenario under the assumption of a random network, 5 additional scenarios are simulated in random network with different assumptions with regards to contact tracing. Question 1 is well examined within this set of experiments. Table 5.6 exhibits the settings for each scenario. For the "Lost follow-up" column, a value of "30% to 40%" means that contact tracing is implemented as the data we obtained from TB control in Saskatchewan, while 10% indiates the standard level of lost follow-up required to achieve published guidelines. All the other

parameters (*i.e.* population size, transmission rate, RR) are maintained the same for all the scenarios in the random network except those mentioned in Table 5.6.

Table 5.6: Scenarios' Settings under the Assumption of the Random Network with Fraction of contacts to investigate equal to 90%

Id	CTI	Tracing Target	Lost follow-up	CTI Speed
R_0	**Disabled**	**N/A**	**N/A**	**Normal**
R_1	Enabled	Infectious & Primary TB	30%-40%	Normal
R_2	Enabled	Infectious & Primary TB	10%	Normal
R_3	Enabled	Infectious TB	30%-40%	Normal
R_4	Enabled	Infectious TB	10%	Normal
R_5	Enabled	Infectious & Primary TB	10%	Faster(within 30 days)
R_6	Enabled	Infectious & Primary TB	30%-40%	Faster(within 30 days)

The set of scenarios with the random network assumption mainly focus on simulating the impact of lost follow-up on TB control outcome. R_1 and R_2 simulate the effect of different contact loss levels in the contact tracing with investigation targets on both infectious TB and primary TB cases. In a similar fashion, R_3 and R_4 also address the issue of loss of contacts, but with the assumption that the investigation only takes place on infectious TB cases.

In addition, it is worth looking into the outcome by applying the ideal guideline of contact tracing. So scenario R_5 is designed to simulate the optimal contact tracing scheme which is to investigate 90% of the contacts within 30 days of the diagnosis of an active case. Similarly, scenario S_6 simulates the faster contact tracing with a different follow-up lost level. The "Normal" speed of the contact tracing is implemented as that we obtained from historical data in 2008 (seen in Table 5.1), while the "Faster" speed of CTI indicates that the contacts are interviewed and tested within one month on average.

Scenarios Assuming a Small World Network

Under the assumption of a small world network, scenarios regarding the impact of follow-up loss are also examined. In addition to the small world baseline scenario without contact tracing, 2 more scenarios are designed and simulated with different level of follow-up loss. The settings of the scenario are depicted in Table 5.7.

Scenarios Assuming a Scale-free Network

In addition to questions with respect to the impact of level of follow-up loss, prioritized contact tracing (Question 2) is evaluated within the assumption of a scale-free network. Table 5.8 shows

Table 5.7: Scenarios' Settings under the Assumption of the Small World Network with Fraction of Contacts to Investigate Equal to 90%

Id	CTI	Tracing Target	Lost follow-up	Tracing Fraction
W_0	**Disabled**	**N/A**	**N/A**	**N/A**
W_1	Enabled	Infectious & Primary TB	30%-40%	90%
W_2	Enabled	Infectious & Primary TB	10%	90%

the parameters of different contact tracing protocols. S_0 is the baseline scenario without contact tracing. S_1 and S_2 address the issue of follow-up loss when 90% of contacts been investigated and when both infectious TB and primary TB cases are traced. Scenarios S_3, S_4, S_5, S_6, and S_7 evaluate the impact of different prioritization schemes (e.g. age, ethnicity, number of times been reported as a contact, etc.) for contact tracing when we assume a lower fraction of contacts to investigate (45%). S_8 is designed to evaluate fast contact tracing, which is to have 90% contacts skin tested within 30 days of diagnosis, and it is comparable with S_2 where contacts are investigated at normal speed.

Table 5.8: Scenarios' Settings under the Assumption of the Scale-free Network

Id	CTI	Tracing Target	Lost follow-up	Priority	Tracing Fraction
S_0	**Disabled**	**N/A**	**N/A**	**N/A**	**N/A**
S_1	Enabled	Infectious & Primary TB	30%-40%	None	90%
S_2	Enabled	Infectious & Primary TB	10%	None	90%
S_3	Enabled	Infectious & Primary TB	10%	None	45%
S_4	Enabled	Infectious & Primary TB	10%	Age	45%
S_5	Enabled	Infectious & Primary TB	10%	Ethnicity	45%
S_6	Enabled	Infectious & Primary TB	10%	Reported Times	45%
S_7	Enabled	Infectious & Primary TB	10%	Age & Ethnicity	45%
S_8	Enabled	Infectious & Primary TB	10%	None	90% Fast
S_9	Enabled	Infectious TB	30%-40%	None	90%
S_{10}	Enabled	Infectious TB	10%	None	90%

5.4 Results

5.4.1 Results of Baseline Scenarios

Table 5.9 shows the cumulative incident cases for a period of 20 years under each of the baseline scenarios in turn, and each scenario is simulated for 30 realizations. Given assumptions of different underlying network structures for the population, the cumulative incident cases varies significantly. Scale-free network baseline scenario (S_0) gives the highest mean of cumulative incident cases with large standard deviation, while the lowest mean of the cumulative incident cases among these 3 baseline scenarios is given by the small world baseline scenario (W_0), which also gives the lowest standard deviation. The coefficient of variation (C.V.) is calculated to give a more understandable estimate of the fluctuation level of our measurement (cumulative incident cases); a smaller C.V. indicates the measurement values are less dispersed (when considered relative to the size of the mean) than those with larger C.V.. According to our results based on 30 realizations, the scale-free network baseline scenario gives the highest C.V., and the lowest C.V. is obtained from the small world network baseline scenario.

Table 5.9: Average Cumulative Incident Cases for 20 Years in Baseline Scenarios Absence of Contact Tracing with Implementation of Different Network Structures

	Cumulative Incident Cases (Active TB)				
Scenario Id	*Mean*	*Max*	*Min*	*Std. Deviation*	*Coefficient of Variation*
R_0	250.633	300	199	25.591	0.102
W_0	181.433	211	151	14.516	0.08
S_0	425.633	614	289	74.659	0.175

The mean prevalence of TB Infection for baseline scenarios are illustrated in Figure 5.10. In the absence of any contact tracing protocols, the prevalence increases over time in the scale-free network. In random and small world networks, a declining trend is observed over time, and the prevalence of TB infection in a small world decreases faster compared with that in a random network.

5.4.2 Results of Scenarios Assuming a Random Network

Table 5.10 depicts the cumulative incident cases for scenarios with a random network assumption. Different protocols of contact tracing are assessed especially for the issue regarding follow-up loss. Based on the mean of the cumulative TB cases, the baseline scenario gives the highest number of TB cases across all the scenarios. Contact tracing investigation helps reduce the cumulative

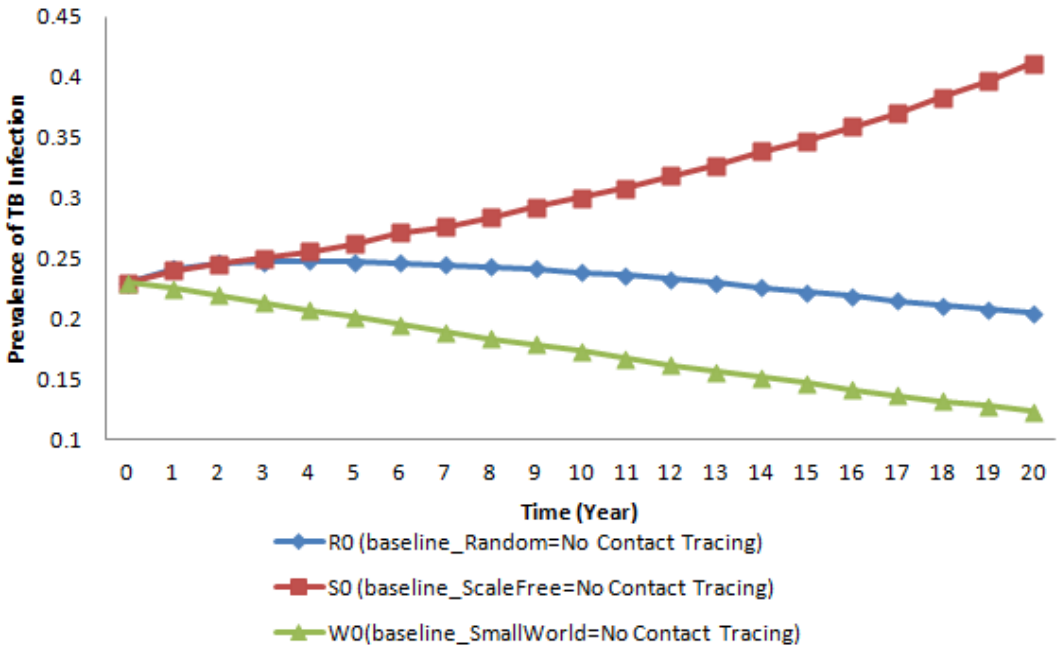


Figure 5.10: Prevalence of TB Infection in the Baseline Scenarios Absence of CTI with Implementation of Different Networks

TB cases, although the average number of reduced TB cases are quite limited. Targeting both infectious and primary TB cases, scenarios R_1 and R_2 show that lowering the loss level of contacts to 10% can reduce around 18 cases on average. If we restrict the aim of contact tracing to infectious TB cases only, maintaining a 10% loss of contacts level can prevent roughly 16 active TB cases on average. Comparing scenario R_1 with R_3 (which differs only in tracing just infectious cases), the cumulative incident cases don't vary much. Scenarios R_2 and R_4 also doesn't exhibit significant difference regarding the average count of cumulative incident cases, despite the broader population beting traced in R_2.

Regarding the ideal protocol for contact tracing, having 90% of the contacts skin tested within 30 days of diagnosis of active TB cases as well as maintaining 10% loss of contacts – scenario R_5 – doesn't show a large difference in cumulative TB cases compared with those of R_2 where 90% of the contacts are tested continuously within a year (rather than within 30 days). While comparing R_1 with R_6 (maintaining the same level of follow-up loss 30% to 40% but with difference only in the speed of CTI), the cumulative TB incident cases decrease dramatically from 231.5 to 223.4 on average.

Figure 5.11 gives the realization-mean prevalence of TB infection among the population over time. All trajectories demonstrate the declining trend over time. The prevalence of TB infection in baseline R_0 is above that of other scenarios where contact tracing is enabled. Lower level of follow-up loss ultimately gives lower prevalence, seen in R_2, R_4 and R_5. Having reverse contact

Table 5.10: Average Cumulative Incident Cases for 20 Years under the Assumption of Random Network

	Cumulative Incident Cases (Active TB)				
Scenario Id	*Mean*	*Max*	*Min*	*Std. Deviation*	*C.V*
R_0	250.633	300	199	25.591	0.102
R_1	231.5	299	197	29.821	0.129
R_2	212.667	262	184	17.149	0.079
R_3	232.333	285	187	26.259	0.113
R_4	216.933	266	181	21.284	0.098
R_5	213.3667	277	167	22.51	0.105
R_6	223.4	286	173	24.47	0.11

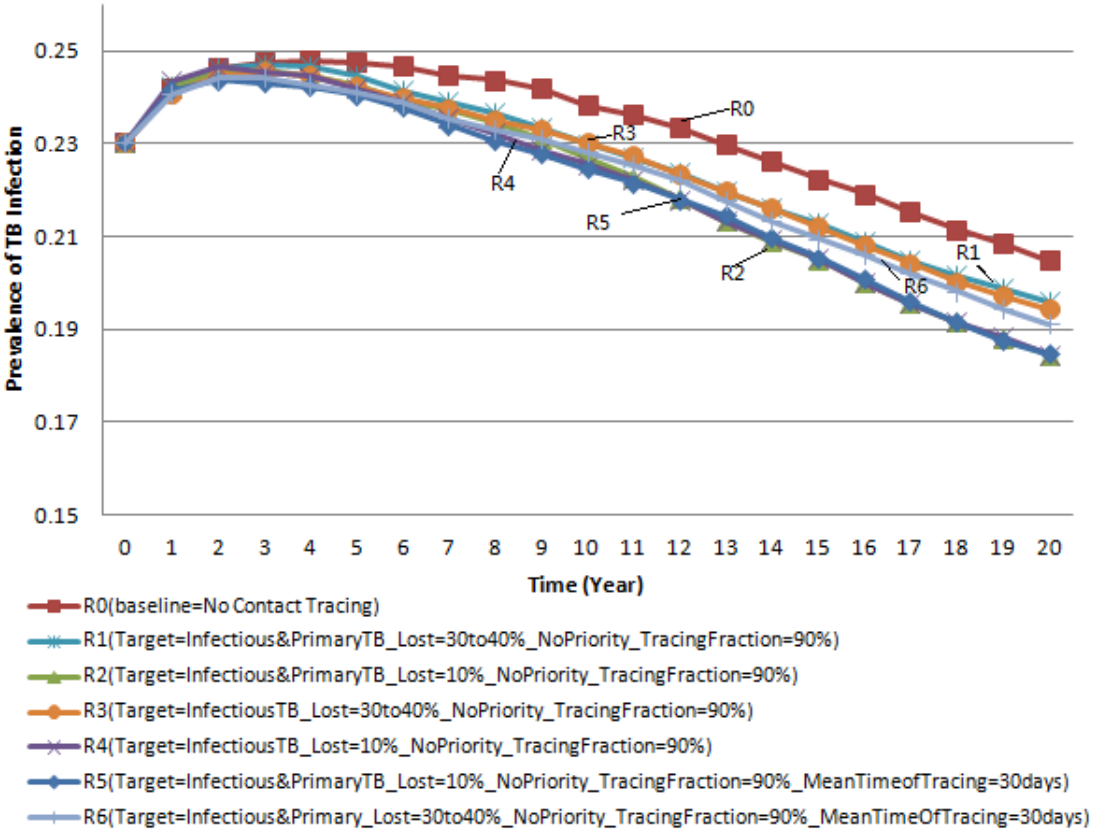


Figure 5.11: Prevalence of TB infection for Scenarios Regarding Random Network

tracing (which targets on primary TB cases to assess the source of infection) enabled doesn't appear to bring observable difference in TB infection prevalence; the TB infection prevalence in R_1 and $R3$ are almost overlapped; R_2 are R_4 are also overlapped in prevalence of TB infection. Since the objective of investigating primary TB cases is not to lower or reduce the TB incident cases or prevalence of infection directly, so the contribution given by such reverse contact tracing can't be easily assess by these 2 measurements.

In addition, maintaining a faster speed of contact tracing (30 days) and 10% follow-up loss in R_5 doesn't produce significant improvement in prevalence compared with the current speed of contact tracing (R_2). However, under 30% to 40% follow-up loss, faster contact tracing (R_6) seems to be able to lower the prevalence of TB infection while compared with R_1.

5.4.3 Results of Scenarios Assuming a Small World Network

Table 5.11 shows the results of small world network scenarios regarding cumulative incident cases. Regarding the loss of contacts, the mean of cumulative TB cases doesn't vary dramatically across scenarios W_0, W_1 and W_2. On average, a small number of active TB cases can be prevented by having contact tracing enabled (W_0 versus W_1). The improvement in reducing active TB cases is also small (only 6 cases) when a 10% of loss of contacts is maintained (see W_1 versus W_2).

Table 5.11: Average Cumulative Incident Cases for 20 Years under the Assumption of Small World Network

	Cumulative Incident Cases (Active TB)				
Scenario Id	*Mean*	*Max*	*Min*	*Std. Deviation*	*C.V*
W_0	181.433	211	151	14.516	0.08
W_1	174.267	199	136	15.0	0.086
W_2	168.867	197	137	16.033	0.095

Figure 5.12 illustrates the TB infection prevalence in the small world network. The difference in scenarios W_0, W_1,and W_2 are not remarkable, which indicates that increasing the contact investigation level didn't make explicit contribution to the prevalence of TB infection.

5.4.4 Results of Scenarios Assuming a Scale-free Network

Under the assumption of a scale-free network, contact tracing strategies varying the level of contact loss, prioritized contact tracing, and more rapid follow-up are investigated. Figure 5.12 shows the effect of different contact tracing protocols on the cumulative incident cases in the population. In absence of any contact investigation activities, the average cumulative incident cases is 425.6 cases

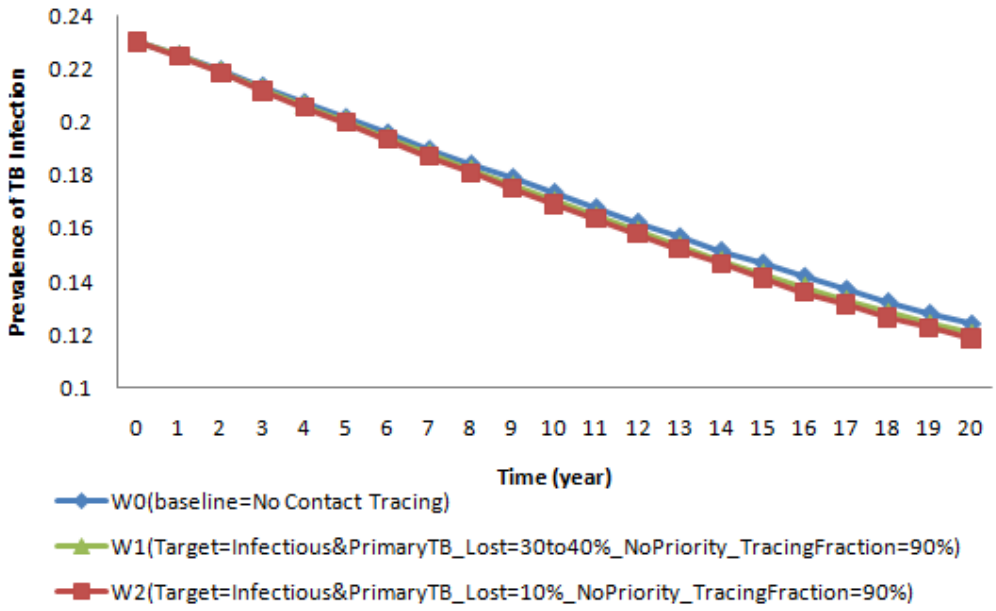


Figure 5.12: Average Prevalence of TB infection for Scenarios regarding Small World Network

– the highest across all the scenarios – and the gap regarding total number of incident cases between baseline S_0 and scenario S_1(where contact tracing is enabled with 90% contacts been investigated) is distinct. Having 90% of contacts investigated, reducing the percentage of loss in contacts from current level (30% to 40%) to 10% can bring the total incidence numbers down from 311.8 cases to 279.1 cases on average, seen in scenarios S_1 and S_2. Similarly, with scope of contact tracing targeting on infectious TB cases only, lowering the loss of contacts in CTI can result in a much lower cumulative incident case (from 315 cases to 271.6 cases on average for 20 years, seen in S_9 and S_{10}).

While investigating the impact of modifying the target of contact tracing via comparing S_1 with S_9 or comparing S_2 with S_{10}, the mean of cumulative incident cases decrease.

Results of prioritized contact tracing are shown in scenarios S_3 (no priority), S_4 (age priority), S_5 (ethnicity priority), S_6 (prioritized by number of times been reported as contacts) and S_7 (age and ethnicity priority). In each of these scenarios, 45% of contacts are investigated with 10% loss in follow-up. Given the relative risk associated with the youngest ages in terms of TB infection, age prioritized contact tracing – where children under 14 will be targeted first – gives the lowest cumulative incident cases (283 cases on average). And it is almost the same average level achieved by investigating 90% of contacts without priority (279.1 cases on average in S_2). Comparing with the scenario S_3 without priority (318.7 cases), prioritized contact tracing by age (S_4), ethnicity (S_5), or a combination of age and ethnicity (S_7) gives lower cumulative incident cases on average, while prioritized contact tracing by number of times a contact has been previously named generates

higher cumulative incident cases (S_6) (363.2 cases). The mixing priority scheme by age and ethnicity (S_7) is better than priority with ethnicity only (S_5).

The contribution given by speeding up the contact tracing (improving the speed of contact tracing to 30 days of diagnosis) can be observed from scenarios S_2 and S_8. On average, 14 active TB cases can be prevented in the stylized community over the course of 20 years when much faster contact tracing is implemented.

Table 5.12: Average Cumulative Incident Cases for 20 Years under the Assumption of Scale-free Network

	Cumulative Incident Cases (Active TB)				
Scenario Id	*Mean*	*Max*	*Min*	*Std. Deviation*	*C.V*
S_0	425.633	614	289	74.659	0.175
S_1	311.767	429	217	49.646	0.159
S_2	279.1	392	211	49.682	0.178
S_3	318.667	403	207	48.093	0.151
S_4	283	364	193	40.403	0.142
S_5	302.233	486	194	64.917	0.215
S_6	363.2	508	239	70.19	0.193
S_7	291	383	190	53.018	0.182
S_8	265.5	400	185	44	0.166
S_9	315	438	184	49.2	0.156
S_{10}	271.6	387	192	41.57	0.153

Figure 5.13 shows the average prevalence of TB infection in the scale-free network when altering assumptions regarding follow-up loss and tracing target. Decreasing the contacts loss in the clinical review stage can eventually reduce the prevalence of TB infection dramatically (seen in S_1 and S_2), while restricting the target on infectious TB cases slightly increase the prevalence of TB infection (seen S_1 and S_9, or S_2 and S_{10}).

Figure 5.14 shows the prevalence of TB infection with respect to different speed of contact tracing. On average, faster contact tracing (S_8) didn't lower the prevalence of TB infection dramatically compared with normal speed contact tracing (S_2).

Figure 5.15 illustrates the impact of prioritized contact tracing on prevalence of TB infection. Scenarios S_0, S_2 and S_3 shows the impact of assuming that different fractions of contacts enter the contact tracing system. The gains with respect to TB infection prevalence from tracing of contacts 45% (S_3 vs. S_0) is bigger than that from tracing the second 45% of contacts (S2 vs.

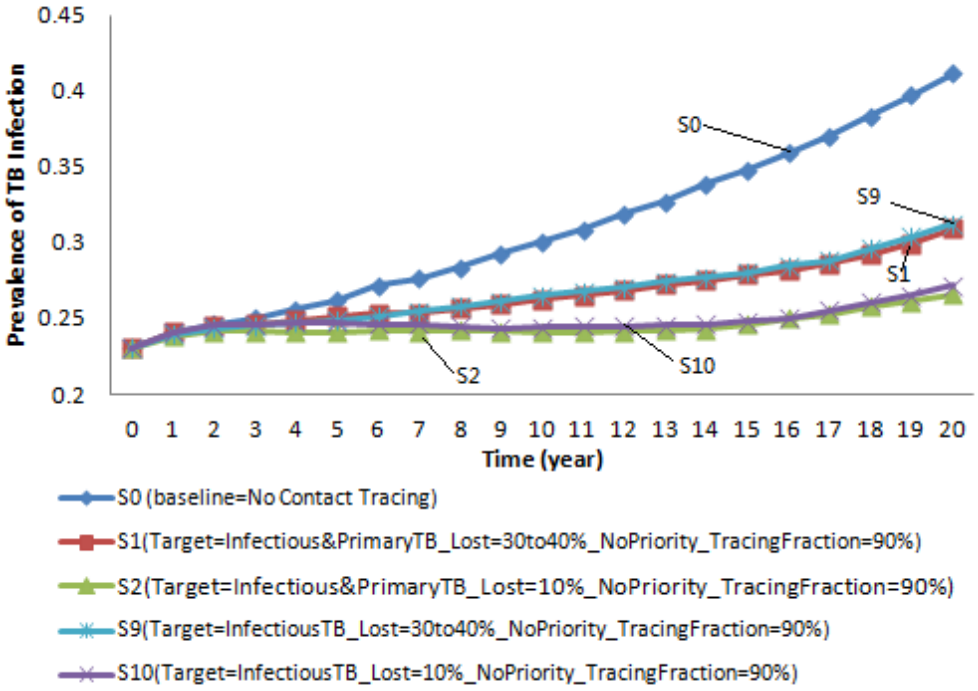


Figure 5.13: Averaged Prevalence of TB infection for Scenarios regarding Lost Follow-up and Investigation Target in the Scale-free Network

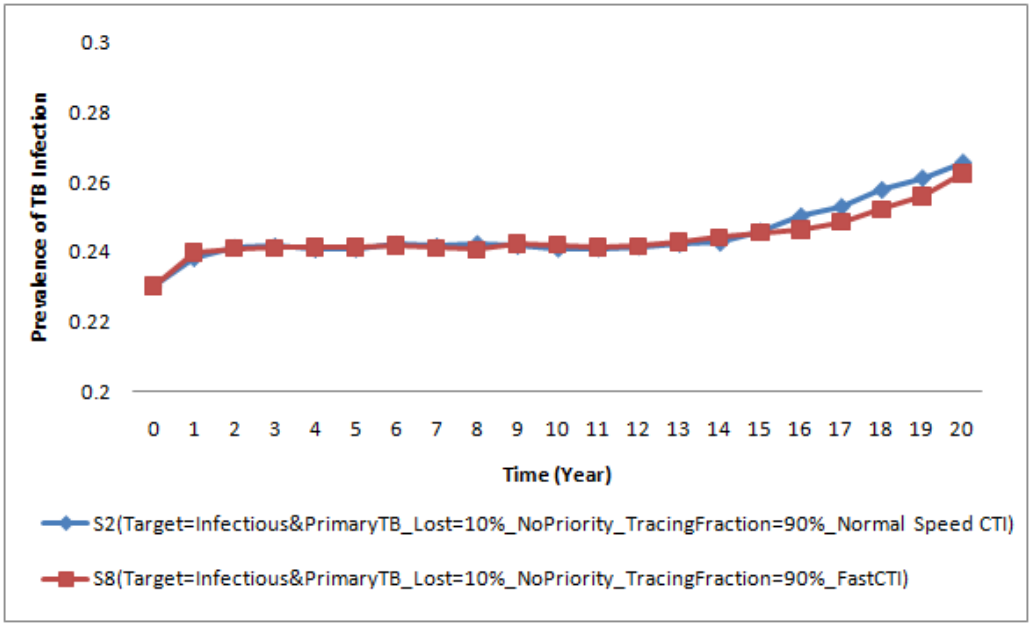


Figure 5.14: Averaged Prevalence of TB infection for Scenarios regarding the Speed of Contact Tracing in the Scale-free Network

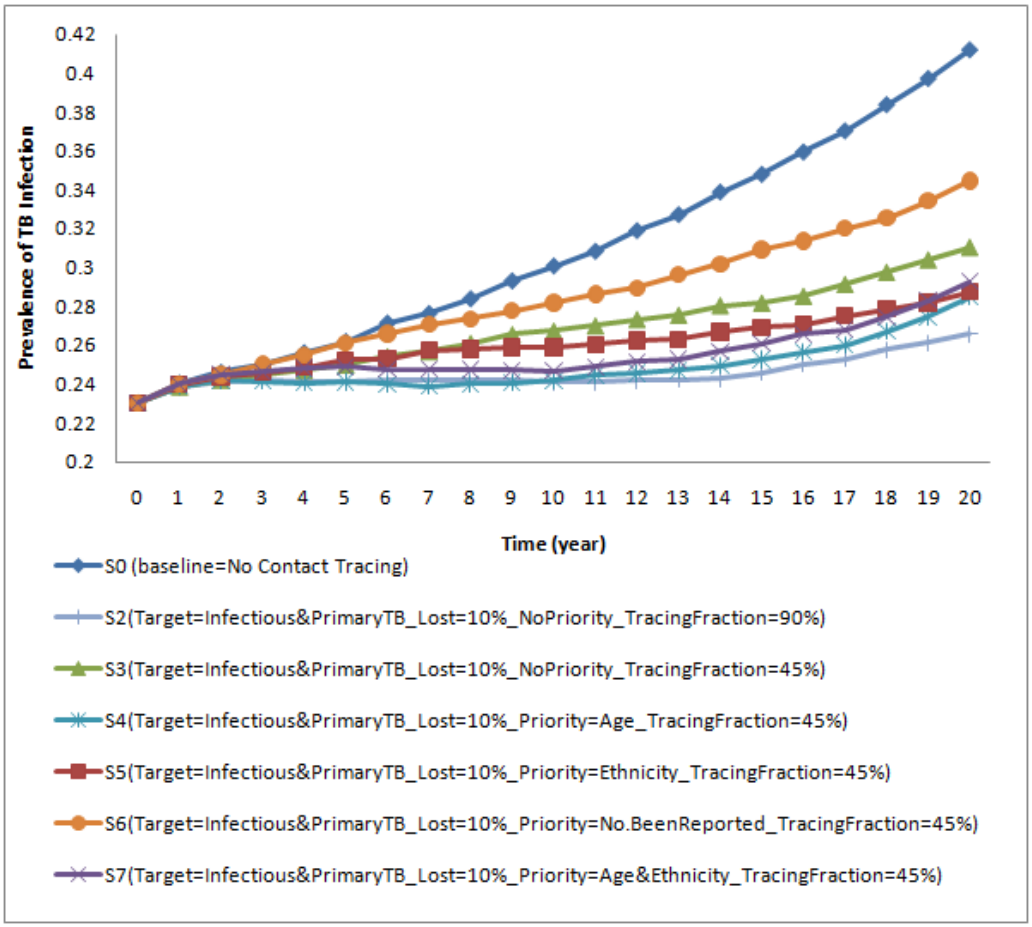


Figure 5.15: Prevalence of TB infection for Scenarios regarding Prioritized Contact Tracing in the Scale-free Network

S3), so diminishing returns is observed, even in the absence of prioritization of contact tracing. In addition, the prevalence of TB infection in scenarios S_4, S_5 and S_7 fall into an envelop with an upper boundary given by scenario S_3 (45% contacts investigated without priority) and the lower boundary given by scenario S_2(90% contacts traced without priority as well), although scenario S_4 (with pure age prioritization) comes close to the bottom of that boundary. Scenario S_6 (where priority is based on number of times identified as a contact) gives the highest prevalence of TB infection compared with other prioritized investigation scheme, and it is even higher than contact tracing without priority (S_3).

5.5 Conclusion and Discussion

The role of contact tracing is examined on different contact network topologies with a variety of combination of investigation priorities, target groups and tracing rates. Given our individual-based network model representing the contact tracing program for TB control, qualifying the trade-offs associated with different protocols of contact tracing at an individual level (e.g. prioritized contact tracing strategies, contacts loss) becomes possible, and such a model can further examine and enhance our understanding of the current situation so as to aid in the formulation of effective control strategies.

Starting from different assumptions regarding the underlying network of the population, different TB diffusion patterns are readily observed. For the baseline assumptions, scale-free networks lead TB infection to spread aggressively if no actively diagnosis or control programs are implemented, while the TB infection tends to decrease in random and small world networks even without any control programs like contact tracing. In terms of cumulative incident cases among the baseline scenarios, the highest incident cases on average is obtained from the scale-free network; this scenario also yields the highest sample standard deviation as well, results are consistent with the high heterogeneity of the scale-free network (where most of the nodes in the network have a limited number of connections, and only a few have many connections). By contrast, the incident case counts obtained from the small world and random networks are much lower and possess smaller sample standard deviation. According to the alternative scenarios measuring the effectiveness of contact tracing by prioritization or maintaining a lower level of loss in contacts, contact tracing shows its effectiveness in reducing the prevalent and incident cases explicitly in scale-free network, while such gains from random network and small world network are not remarkable as those from scale-free network, with the gains from the small world network being very small. Ultimately speaking, network architecture is capable of shifting the patterns of TB transmission.

Our first question is to investigate the impact of follow-up loss on the effectiveness of contact tracing given real contact tracing data collected by Saskatchewan TB control. Assumptions regard-

ing the true underlying network are made with additional parameters sets with respect to different contact investigation protocols. Suppose we reach the standard level (having 90% of contact going through the whole procedure, and most notably without missing the potential clinical review) versus the current level of contacts loss, and we are interested in understanding how many incident cases we can prevent and how much we can reduce the prevalence of TB infection on average by minimizing loss in contacts. According our simulation results of contact tracing program with different network assumptions, the benefits we obtained while enabling contact tracing investigation from scale-free and random network are explicit, while those from small world network is not strongly observable. This suggests to us that it is worth spending extra efforts in bring contacts in, especially in the clinical review phase, to verify their eligibility for latent TB treatment or diagnosis of active disease if the true network is either scale-free or random.

Prioritized contact tracing – which gives priority to high risk groups – provide advanced capability in managing our limited resources. Our second research question is to investigate effectiveness of contact tracing in terms of prioritization, and to measure the gains by adhering to prioritized tracing protocols. Prioritized contact tracing is evaluated within a scale-free network in particular. With 45% contacts investigated and 10% follow-up loss, we examined contact tracing prioritized by age, ethnicity, age and ethnicity, count of times the contact has previously been nominated as a contact. According to the results shown in Table 5.12 and Figure 5.15, the cumulative incident cases on average listed from lowest to highest is S_4 (age priority), S_7 (age and ethnicity priority), S_5 (ethnicity priority), and S_6 (reported times as contacts). Compared with that from S_3 (45% contacts investigated with no priority and 10% follow-up loss), prioritized protocols of contact tracing – except the one prioritized by the count of times an individual has been reported as a contact – give a better TB control outcome. In particular, the average cumulative incidences cases from S_4 (age priority) is almost the same as that from contact tracing protocol without priority and tracing twice the number of contacts. This suggests potential TB control strategies, especially from the cost-effectiveness perspective. We can achieve almost the same level incident cases by tracing 45% of contacts instead of 90% contacts, which eventually will save resources and money – especially given limited human resources and funding. Given the prioritization by age, ethnicity or a combination of these 2 risk factors, we can also achieve a lower level of TB infection prevalence compared with that from scenario S_3 – when no priority is implemented with 45% contacts investigated – although the prevalence is still higher than that from the scenario S_2 when 90% of contacts are traced without priority.

Under the assumption of a random network, reverse contact tracing (targeting primary TB cases) is evaluated by adjusting different tracing targets. According the results from R_1 versus R_3 and R_2 versus R_4 (seen in Table 5.6), reverse contact tracing (with objective of finding the source of infection) doesn't generate noticeable impact on both average cumulative incident cases

and TB infection prevalence. If the main target of contact tracing is to lower the prevalence and incidence rate, then reverse contact tracing is inefficient in preventing TB diffusion. Given the limited contribution by reverse contact tracing, it can be disabled for effectiveness unless it is designed for other purposes.

Scenarios regarding faster contact tracing (R_7 and S_8) in both random and scale-free network are simulated. Each of these scenarios assumes that contacts finish the procedure of investigation within 30 days. However, the results turns out that faster contact tracing doesn't contribute strongly to the TB control outcome (prevalence of TB infection in particular). This results is consistent with the answer to the same question we examined in the System Dynamics model in Chapter 4. In a scale-free network, faster contact tracing can bring the cumulative incident cases down to a lower level by reducing 14 TB cases on average over the 20 year period. Based on the cross-checked results regarding faster contact tracing, we conclude that, given a random social network of the community, there may not be a need to speed up the procedure of contact tracing, since it is inefficient from a cost-effectiveness perspective. While in a scale-free network, there is trade-off when implementing faster contact tracing. If the major objective is to control the prevalance of infection, faster contact tracing is not efficient; however, given the goal of reducing the incidence rate, then faster contact tracing can bring limited contribution to a lower level of cumulative TB cases.

Chapter 6

Conclusions

6.1 Summary

The objective of this thesis is to compare and contrast representation of TB Transmission dynamics using System Dynamics modeling and agent-based modeling, so that we can enhance our understanding of the dynamics of TB diffusion in terms of casual effects, risk factors, social networks and evaluation of preventive and treatment strategies. Agent-based models and aggregate models are compared in the context of TB by setting up a variety of scenarios. Such comparison helps reveal trade-offs between these two modeling methodologies. In addition, contact tracing strategies are evaluated in a System Dynamics model of TB diffusion by conducting sensitivity analysis. From an cost-effectiveness view, our model can explore different alternatives when deploying the contact tracing strategies to optimize the outcome in terms disease control and limited resource as a whole. Moreover, the impact of explicit incorporation of networks – a closer approximation to the underlying pathways of transmission – are investigated in an agent-based TB model. Such a model allows us to conduct comparison with the results of the aggregate models while charactering the disease transition. In addition, our capacity to perform simulations of contact tracing within social network structures provides extra support in estimating the efficiency of ongoing and alternative control policies and enriches our understanding of disease transmission with respect to network topologies. Such epidemiological models can assist in sorting alternatives and optimizing TB prevention and control programs.

6.2 Deliverables of the Research Work

The deliverables of the thesis work include:

- An agent-based TB model with smoking impact, oriented towards exploring pros and cons of the aggregate modeling and agent-based modeling methodologies in the course of TB transmission.
- An aggregate System Dynamics TB model with respect to contact tracing investigation in a small community of Canadian province of Saskatchewan.

- An network-based TB model oriented towards investigating the impact of contact tracing timing, breadth and prioritization strategies at an individual level to explore the optimal health policies and preventions in TB control.

6.3 Thesis Contribution

The main contribution of this thesis are:

- Understanding of pros and cons of both aggregate modeling and agent-based modeling methodologies under the context of TB transmission. This study has investigated the impact of different representations of BCG vaccine, reactivation and network on TB transmission in these 2 models. Different results generated from these 2 models enrich our understanding about different methodologies and the cost in terms of model development and extension.

- Through simulating contact tracing procedure in a System Dynamics model, the efficiency of contact tracing program in a small community of Saskatchewan has been investigated. Regarding the question about how many contacts we should trace, our model depicts diminishing returns in terms of incidence rate and prevalence when the fraction of contacts has been investigated goes up. In addition, we also study the speed of contact tracing as another indicator of the efficiency of contact tracing; scenarios with different speed settings suggest that the speed doesn't contribute dramatically to the TB infection prevalence or incidence rate; nevertheless, contact tracing plays a role in maintaining a lower prevalence and incident level.

- The third model has extended an existing, well–calibrated mathematical model of TB Dynamics for Saskatchewan population to an network-based model, supporting our examination of more detailed features regarding efficiency of contact tracing and the impact of social network on TB transmission. Given different assumption of the underlying social network of the population, we have investigated issues of follow-up loss of contacts, speed of contact tracing and different targets of contact tracing. Given the assumptions regarding the population network, we re-examined the same questions which we explored in the System Dynamics TB transmission model with contact tracing to cross-verify our findings. Results of our simulation suggests that the efficiency of contact tracing varies greatly given different network structure; network architecture alters the dynamic patterns of TB diffusion which eventually affects the contact tracing. Some preliminary results reveals that the speed of contact tracing doesn't contribute dramatically in reducing the burden of TB, while the fraction of contacts brought into contact tracing programs plays an important role, the contribution in terms of reducing the cumulative incident case and prevalence is distinct when the underlying social network

of the population is assumed to be scale-free. The contribution of prioritized contact tracing given the assumption of a scale-free network is also remarkable, which should be emphasized in practice especially from a cost-effectiveness view. Follow-up loss in contacts is also examined, and its efficiency varies from different network.

6.4 Future Work

The agent-based model in Chapter 3 can be extended by implementing dynamic network (rather than static network) to gain insights into how TB transmission is affected by network dynamics. A customized network representing preferential attachment (regarding ethnicity, age or other factors) could be developed to more accurately address the contact mixing patterns instead of using a random mixing network.

Given the System Dynamcis model in Chapter 4, risk factors (e.g. age, ethnicity) of TB tranmission and sophisticated mixing patterns should be addressed to more accurately capture TB dynamics in community 1.

Based on the simulation studies in Chapter 5, more simulations should be conducted in conjunction with statistical analysis to further verify the preliminary findings. Since the presented outcome are based on a limited set of realizations (30 runs per scenario). Robustness of our findings under different settings (e.g. disease parameters, control strategies parameters) need to be examined further to confirm such differences are not simply the results of chance. Our measurements regarding the scenarios are the average values of either cumulative incidence cases or TB prevalence across realizations. Further statistically analysis needs to be undertaken to estimate the reliability of our simulation results. For example, Student's t-test can help examine whether the mean from each scenario is statistically different.

Moreover, scenarios regarding different mean times of contact tracing should be simulated to gain insights into the sensitivity of TB control outcome on the speed of contact tracing. Faster contact tracing is only examined within random and scale-free networks, and it doesn't produce significant contribution in both cumulative incident cases and prevalence of TB infection. Similar scenarios should also be examined within a small world network. Insights into the impact of faster contact tracing is very important, given the current objective of contact tracing in Saskatchewan is to examine 95% of the contacts within 30 days of diagnosis of an active TB case. To that end, it would be relatively straightforward to examine the impact of assuming different speeds of contact tracing (rather than just the two levels of delay examined here).

We only employ 2 measurements (average cumulative incident cases and TB prevalence) when assessing effectiveness of contact tracing in our agent-based model; additional measurements or criteria can be included to further evaluate contact tracing efficiency from different angles. In addi-

tion, we extend the existing calibrated model by adding contact tracing procedure which might alter the dynamics of the original model. Futher calibration work is needed to verify that our enriched TB model with contact tracing can reproduce the historical patterns among the Saskatchewan communities.

In addition to the scenarios we examined in the agent-based TB model with contact tracing, there are more opportunities to more accurately address TB control given our implementation at an individual level in agent-based model, e.g. targeted BCG vaccination, waning BCG and waning TLTBI. Waning of protection given by latent TB treatment or vaccination can be implemented with dependency on the number of years that have elapsed since prophylaxis or BCG instead of assuming a static averaged rate per year (as is currently assumed).

Another useful avenue of future work is to capture more details of the mobility at the individual level in the agent-based model with contact tracing. For example, our model currently assumes that the network of the population is a more static one, where the network is altered only by birth and death. A more realistic model could have heterogeneous individuals with different levels of mobility and implement disease transmission via dynamic contact instead of static connections. Another opportunity employing mobility of individuals is to bring breaking and forming of relationship into the static network. However, data regarding the mobility of people living in a typical First Nations community in Saskatchewan is quite limited. Although dynamics of mobility can introduce more realistic scenarios, we should be cautious that it might undermine the reliability of the results due to additional assumptions about people's contact pattern, in particular contact patterns of First Nations people in Saskatchewan.

Since the built-in function in Anylogic is applied when implementing the network structure, different random seeds will generate different network in each realization, an further improvement we can do in the future on the agent-based model in Chapter 5 is to initialize one network using a fixed random seed, then run different contact tracing protocols on the same network, like pairwise scheme to remove the noise or bias introduced by different network generated each time. Comparison based on pairwise schemes can introduce extra robustness of our simulations.

A further line of work could lend insight into the structure of the underlying contact network, which can be accomplished by comparing characteristics (degree distribution, degree of locality) in the TB control contact tracing network with the same characteristics in the simulated contact tracing network that emerges from the simulation model. Given the large volume of simulation results regarding the network in Chapter 5, the simulated contact tracing network data can be analyzed and compared with the true network data collected by Saskatchewan TB control. Such comparison might help us identify the distinguishing trademarks of TB spread on different network structure, and it might provide insight into true network or disease parameters, which can be important when deciding and selecting the most desirable protocol for contact tracing.

References

[1] R. M. Anderson and R. M. May. *Infectious Diseases of Humans: Dynamics and Control*. Oxford University Press, 1991.

[2] J. P. Aparicio and J. C. Hernández. *Preventive treatment of tuberculosis through contact tracing*. Contemporary Mathematics, Volume 410, 2006.

[3] A. Al-Azem. *Social Network Analysis in Tuberculosis Control Among the Aboriginal Population of Manitoba*. PhD dissertation, University of Manitoba, Canada, 2007.

[4] A. Al-Azem, V. Hoeppner and N. Osgood. *Social Network Analysis (SNA) advantages in tuberculosis (TB) control in high TB incidence community in Saskatchewan*. In Proceedings INSNA Sunbelt XXX. Trento, Italy. July, 1, 2010.

[5] A.L. Barabási and R. Albert. *Emergence of Scaling in Random Networks*. Science, 286 (5439), 509-512, 1999.

[6] E. Bonabeau. *Agent-based modeling: Methods and techniques for simulating human systems*. Proc. Nat. Acad. Sci. U.S.A., 99 (Supplement 3):72807287, 2002.

[7] G.A. Colditz, T.F. Brewer, C.S. Berkey, M.E. Wilson, E. Burdick, H.V. Fineberg and F. Mosterller. *Efficacy of BCG vaccine in the prevention of tuberculosis: Meta-analysis of the published literature*. Journal of the American Medical Association, 271(9), pp.698-702, 1994.

[8] G.A. Colditz, C.S. Berkey, F. Mosteller, et al. *The efficacy of bacillus Calmette-Guerin vaccination of newborns and infants in the prevention of tuberculosis: meta-analyses of the published literature*. Pediatrics, 96 1 (Part 1), pp. 2935, 1995.

[9] G.W. Comstock and C.E. Palmer. *Long-term results of BCG in the southern United States*. Am Rev Resp Dis 93 (2): 17183, 1966.

[10] V. Darley and A.V. Outkin. *A NASDAQ Market Simulation: Insights on a Major Market from the Science of Complex Adaptive Systems*, World Scientific, Singapore, 2007.

[11] E. W. Dijkstra. *Selected writings on Computing: A Personal Perspective*. New York, NY, USA: Springer-Verlag New York, Inc.. pp. 60-66. ISBN 0-387-90652-5, 1982.

[12] N. Eagle, A. Pentland, and D. Lazer. *Inferring friendship network structure by using mobile phone data*. Proceedings of the National Academy of Sciences, 106(36):15274-15278, 2009.

[13] K. T. D. Eames and M. J. Keeling. *Contact tracing and disease control*. Proc. R. Soc. Lond. B 270, 25652571, 2003.

[14] J.W. Forrester. *World Dynamics*. Cambridge:Wright-Allen Press, 1971.

[15] G. P. Garnett and R.M. Anderson. *Sexually transmitted disease and sexual behavior: insights from mathematical models*. J. Infect. Dis. 174, S150-S161, 1996.

[16] B.T. Grenfell, O.N. Bjornstad and J. Kapper. *Travelling waves and spacial hierarchies in measles epidemics*. Nature 414, 716-723, 2001.

[17] N. Gilbert. *Agent-Based Models (Quantitative Applications in the Social Sciences)*. Sage Publications, Inc., 2005.

[18] U.S. Government Accountability Office. *Global Health: Challenges in Improving Infectious Disease Surveillance Systems*. GAO-01-722, Aug 31, 2001.

[19] P.D. Hart and I. Sutherland.*BCG and vole bacillus vaccines in the prevention of tuberculosis in adolescence and early adult life*. Final Report of the Medical Research Council. Brit Med J 2: 29395, 1977.

[20] K. M. Hassmiller. *The Impact of Smoking on Population-Level Tuberculosis Outcomes*. PhD dissertation, University of Michigan, USA, 2007.

[21] M. Hashemian, K. Stanley, and N. Osgood. *Flunet: Automated tracking of contacts during flu season*. Proceedings of the 6th International workshop on Wireless Network Measurements (WiNMee 2010), 557-562, 6pp, 2010.

[22] K. M. Hassmiller. *The association between smoking and tuberculosis*. Salud Publica Mex.48(Suppl 1):S20116, 2006.

[23] O. Hein, M. Schwind and W. König. *Scale-Free Networks The Impact of Fat Tailed Degree Distribution on Diffusion and Communication Processes*. Wirtschaftsinformatik. 48(4), pp.267-275, 2006.

[24] R. Huerta and L.S. Tsimring. *Contact tracing and epidemics control in social networks*. Phys. Rev. E 66, article number 056115, 2002.

[25] J.M. Hyman, J. Li and E.A. Stanley. *Modeling the impact of random screening and contact tracing in reducing the spread of HIV*. Math. Biosci., 181, 17-54, 2003.

[26] J.A. Jacquez and C.P. Simon. *The stochastic SI model with recruitment and deaths - Comparison with the closed SIS model*.Mathematical Biosciences. 117(1-2), pp.77-125, 1993.

[27] S.W. Jaffry and J. Treur.*Agent-Based and Population-Based Simulation: A Comparative Case Study for Epidemics*.Proc. of the 22th European Conference on Modelling and Simulation, ECMS'08. European Council on Modeling and Simulation, , pp. 123-130, 2008.

[28] M. J. Keeling and K. T. D. Eames. *Networks and epidemic models*. Journal of the Royal Society Interface 2, 295-307, 2005.

[29] K.M. Khalil, M.H. Abdel-Aziz, M.T. Nazmy and A.M. Salem.*An agent-based modeling for pandemic influenza in Egypt*. INFOS2010: 7th International Conference on Informatics and Systems, Cairo, Egypt, AOM 24-30, 2010.

[30] A. Mahamoud, Y. Liu, et al. *Tuberculosis Modelling Project: Reconstruction of Kristen M. Hassmillers Model*. CMPT858 report, Department of Computer Science, University of Saskatchewan, 2009.

[31] A. Meng and N. Osgood. *The System Dynamics Longitudinal Analysis System: Quantifying the Hidden Trajectories of System Dynamics Models*. Extended abstract and presentation at System Dynamics Winter Conference 2011, Invited contributions only. Austin, Texas. 2pp. 2011.

[32] M. Morris and M. Kretzschmar. *Concurrent partnerships and transmission dynamics in networks*. Soc Networks, 17:299-318, 1995.

[33] J. Muller, M. Kretzschmar and K. Dietz. *Contact tracing in stochastic and deterministic epidemic models*. Math. Biosci. 164, 3964, 2000.

[34] M. Newman. *The structure and function of complex networks*. Physical Rev. E 66(1) 016128-1-016128-11, 2002.

[35] Northern Saskatchewan Health Indicators Report 2004. http://www.athabascahealth.ca/images/reports/2004%20Health%20Indicators%20Report%20revision1.pdf(2010-1-13).

[36] N.D. Osgood. *Representing Heterogeneity in Complex Feedback System Modeling: Computational Resource and Error Scaling*. Proceedings, 22nd International Conference of the System Dynamics Society, Oxford. 46p, 2004.

[37] N. Osgood, A. Mahamoud, K. Hassmiller, Y. Tian, A. Al-Azem and V. Hoeppner. *Estimating the Relative Impact of Early-Life Infection Exposure on Later-Life Tuberculosis Outcomes in a Canadian Sample*. Research in Human Development. 8(1). pp.26-47, 2011.

[38] L. Perez and S. Dragicevic.*An agent-based approach for modeling dynamics of contagious disease spread*. International Journal of Health Geographics, 8:50, 2009.

[39] Public Health Agency of Canada. *Tuberculosis FACT SHEETS*. http://www.phac-aspc.gc.ca/tbpc-latb/fa-fi/trans-eng.php (2010-1-12).

[40] H. Rahmandad and J. Sterman. *Heterogeneity and Network Structure in the Dynamics of Diffusion: Comparing Agent-based and Differential Equation Models*. Management Science.54(5), pp.998-1014, 2008.

[41] S. Riley. et al. *Transmission dynamics of the etiological agent of SARS in Hong Kong: impact of public health interventions*. Science 300, 1961-1966, 2003.

[42] Saskatchewan Anti-TB League Report, 1972.

[43] Saskatchewan TB Control. *Contact Tracing Procedure Documents*.

[44] J. Sterman. *Business dynamics: Systems thinking and modeling for a complex world*. Boston: Irwin/McGraw-Hill, 2000.

[45] Stop TB Partnership and World Health Organization. *Tuberculosis Facts*. World Health Organization, 2005.

[46] Y. Tian and N. Osgood. *Comparison between Individual-based and Aggregate Models in the context of Tuberculosis Transmission*. In Proceedings, the 29th International Conference of the System Dynamics Society., Washington, D.C. 29pp, 2011.

[47] G.M. Vazquez-Prokopec, et al. *Usefulness of commercially available GPS data-loggers for tracking human movement and exposure to dengue virus*. Int J Health Geogr, 2009.

[48] D.M. Vickers and N.D. Osgood.*Current crisis or artifact of surveillance: Insights into rebound Chlamydia rates form dynamic modeling*. BMC Infect. Dis. 10:70. 2010.

[49] D.M. Vickers and N.D. Osgood. *A unified framework of immunological and epidemiological dynamics for the spread of viral infections in a simple network-based population*. Theor. Biol. Med. Model. 4, 4962, 2007.

[50] E. Vynnycky. *An Investigation of the Transmission Dynamics of M. tuberculosis*. PHD dissertation, London University, 1996.

[51] H.A. Ward. *Risk Factors in the Progression from Tuberculosis Infection to Disease*. Master dissertation, University of Saskatchewan, Canada, pp.84, 2004.

[52] D.J. Watts and S.H. Strogatz. *Collective dynamics of 'small world' networks*. Nature 393, 440-442, 1998.

[53] What is system Dynamics? http://www.systemdynamics.org/what_is_system_dynamics.html (2010-1-12).

[54] World Health Organization. *Tuberculosis (TB)*. http://www.who.int/tb/en/ (2010-1-10).

[55] World Health Organization. *The Stop TB Strategy: Building on and enhancing DOTS to meet the TB-related Millenium Development Goals*. Geneva: World Health Organization, 2006.

[56] World Health Organization. *WHO Tuberculosis Factsheet*. World Health Organization, March 2010.

[57] M. Wooldridge and N.R. Jennings. *Intelligent agents: Theory and practice*. Knowledge Engineering Review, Vol.10, pp. 115-152, 1995.

[58] E. Ziv, C.L. Daley and S.M. Blower. *Early therapy for latent tuberculosis infection*. Amer. J. of Epidemiology. 153(4), 381-385, 2001.