1: Methodology

Overview, Design Concepts, Details, and Human Decision Making (ODD+D)

This document provides a description of the model following standard ODD protocols introduced by Grimm *et al* (2006) and further elaborated to ODD+D by Müller *et al.* (2013). The model source code and data is available at https://github.com/nickstares/kiberaTB.

1.1. Overview

1.1.1: Purpose

Tuberculosis (TB) is an opportunistic infectious disease that is present in its latent form in approximately one-third of the world's population, causing around 1.3 million deaths in 2012 alone (Centers for Disease Control and Prevention, 2012; Kwan and Ernst, 2011). Even though it is treatable, TB continues to pose a major health problem in today's society, and has earned the notorious title as the second most common form of death worldwide from an infectious disease, behind only HIV/AIDS. This unfortunate reality can especially be seen in developing nations, where poor health and living standards are causing 95% of all TB cases and deaths (World Health Organization, 2013a). TB is caused by the bacterium *M. Tuberculosis,* and although biological characteristics of it are well known, there is still a lack of information about the transmission dynamics of the disease in a general population. This research attempts to make some headway in this area of study.

Because TB requires person-to-person interactions to spread the disease, the researchers needed to use a modeling tool that could emulate this process as closely as possible. An Agent-Based Model (ABM) was chosen because it is flexible and meets the necessary requirements to simulate interactions between individual particles (in this situation, particles represent residents). The overall purpose for this study was then to create a sophisticated computational ABM framework that could model a society with TB to learn more about how the disease outbreaks, spreads, and can be contained in a population. Reliable and realistic results produced by the model could be used to gain a better understanding of TB, and provide health policy-makers with more information to make better decisions in the global fight against the disease.

1.1.2: Entities, State Variables and Scales

The test case used for this model was Kibera, a roughly 2.5 square kilometer slum located in Nairobi, Kenya. Classified as the largest urban slum in Africa, Kibera is characterized by poor sanitation, congestion, overpopulation and strong ethnic segregation. Residents are also entrapped in a constant state of paucity and forced to live off just a few dollars a day (Mutisya and Yarime, 2011). Limited opportunities for social or economic advancement only further their vicious cycle of poverty (Patterson, 2012). Because of these reasons, it should come as no surprise that Kibera also has some of the highest rates of TB and HIV (Human Immunodeficiency Virus**)** in the world (Tayler-Smith *et al.*, 2011). This location was thus chosen as a test case for its ability to conform to the necessities of the research question, and its relative importance in the fight against TB. It is important to note, however, that even though this model uses Kibera specifically as a test-case, the model itself and general results obtained could just as easily be abstracted and applied to other locations (especially in more developed countries where there is a plethora of the necessary spatial and demographic data).

The model was created in Java using the simulation toolkit MASON and GeoMason (Luke *et al.*, 2005; Sullivan *et al.*, 2010). Space was modeled explicitly and based on actual geographical information of Kibera. Spatial resolution, based on given data, was 10x10 meters, and the temporal resolution was designed to be a timestep of one hour. Geographic Information Systems (GIS) data from OpenStreetMap and Google Maps was used to ensure high reliability in creating the environment. Next, GIS software ArcGIS and QGIS were both used to convert geospatial files into readable formats, and then later to analyze the spatial information provided by runs. Socioeconomic and demographic data was graciously provided by the Map Kibera Project to help calibrate the model (Marras, 2008).

The entities in the model, listed in a lowest to highest hierarchal format, are as follows: (1) residents, (2) facilities (including households and businesses), (3) structures, (4) parcels, (5) population, and (6) environment.

1.1.2.1: Residents

The residents of Kibera are the agents of the ABM. Each resident is mobile, goal-orientated, and has unique character and health attributes that define its role and function in the model. Furthermore, all residents are susceptible to TB infections. Table 1 shows the state variables that make up residents.

Table 1: Resident's state variables and description.

State Variable	Variable Type	Description	Source
Age	Integer	Residents who are not head of the family will be assigned a random age between 5 and 18. If not a head, there is a 25% chance that the resident is an adult (ages 18 to 62) and a 75% chance resident is a child (ages 1 to 17).	(Marras, 2008)
Gender	Integer	A value of 0 means male, while a value of 1 means female. There is a 61.3% chance of being a male and 39.7% chance of being a female in Kibera.	(Marras, 2008)
Position	Parcel	Resident's current position in the model	N/A
Home Location	Parcel	Location for Resident's home	N/A
Goal	Parcel	Location for Resident's goal	N/A
isStudent	Boolean	A value of true means resident attends school. All residents of ages 5 to 18 are automatically considered school eligible.	(Wosyanju, 2009)
isEmployed	Boolean	A value of true means the resident has a job and completes up to 52 hours a week (maximum work time in Kenya), not considering overtime. Employed residents are assigned to work at a business, school, religious center or health center.	(Orao-Obura, 2002)
My School	Facility	If resident is a student, then it is assigned to a School depending on its location.	N/A
My Employer	Facility	If Resident is employed, then this variable will be a reference to the employer. Otherwise it will be null.	N/A

Ethnicity	String	Residents are given an ethnicity when they are first created based on the ethnic distribution of Kibera: Kikuyu (21%), Luyha (14%), Luo (12%), Kalinjin (12%), Kamba (12%), Kisii (6%), Meru (5%), Mijikenda (5%), Maasai (2%), Turkana (1%), Embu (1%), Other (9%)	(CIA, 2013); (Smedt, 2009)

In addition to the attributes listed above, residents also had personal health characteristics specific to TB. These health characteristics are listed in Table 2.

Table 2: Resident's health characteristics state variables and description

State Variable	Variable Type	Description	Source
Body Health	Double	A resident's body health as a percentage. It starts off at 100 and decreases as the resident's health depreciates. When Body Health reaches 0, the resident dies.	N/A
HIV Depreciation	Double	The hourly body depreciation rate due to untreated HIV. The average survival time for an untreated HIV positive patient is from 0-12 years (or 0-105120 hours).	(Peiperl and Coffey, 2013)
TB & HIV Depreciation Rate	Double	The hourly body depreciation rate due to untreated HIV and TB disease. The average survival time for such a patient is less than 6 months (<= 4380 hours)	(Tiemersma *et al.*, 2011)

TB Only Depreciation Rate	Double	The depreciation rate due to only untreated TB disease. The average survival time for someone with untreated TB disease is less than 3 years (<= 26280 hours)	(Tiemersma *et al.*, 2011)
Health Status	Integer	A resident's health status as defined by the SEIR Model. 1= susceptible, 2 = exposed, 3 = infectious, and 4 = recovered/dead	N/A
Contagious Period	Integer	The length of time a resident will be contagious if they develop active TB disease and start taking treatment. The length of time is randomly selected between the typical period of 2-4 weeks (336-2016 hours)	(Centers for Disease Control and Prevention, 2013)
CD4 Cell Count	Double	The number of CD4 cells the resident has (a key determinant of the body's response to TB).	Rodrigues *et al.*, 2003)
Bacilli Count	Double	The number of *M. Tuberculosis* bacilli currently present in the body. Residents contract the bacilli through interactions with other residents who have active TB disease.	Jones *et al.*, 2009
Infection Dose	Double	The number of bacilli necessary for a resident to contract an infection. Randomly selected between 1-10 bacilli.	Herman *et al.*, 2006

1.1.2.2: Facilities

Facilities in this model are what make up this artificial society (the physical entities present in a conventional slum), and consist of the following types: households, businesses, religious facilities, health facilities, schools, restaurants and public water sources. All facilities, except businesses and households, have their own, specific geographical location (obtained from the Map Kibera Project, 2012). Businesses and households are extrapolated onto the environment and based on empirical survey data (Marras, 2008). However, some data for facilities (such as capacity) had to be extrapolated and estimated. Table 3 lists all general state variables for facilities.

Table 3: Facility (Business) state variables and description

State Variable	Variable Type	Description	Source
Employee Capacity	Integer	The number of employees this facility needs to maintain standard business operation	N/A
Capacity	Integer	Max number of agents that can be on this particular parcel	N/A

Businesses extend facilities and Table 4.0 lists specific state variables for them.

Table 4: Facility (Business) state variables and description

State Variable	Variable Type	Description	Source
Popularity Distribution	Double	While businesses are spread throughout the environment based on empirical data, popularity is uniformly distributed.	N/A

Households also extend facilities and Table 5 lists some state variables specific to them.

Table 5: Facility (Households) state variables and description

State Variable	Variable Type	Description	Source
Ethnicity	String	The ethnicity of the household. All members have the same ethnicity	N/A
Water Total	Double	The amount of water in the house. Members can replenish themselves using this water (if enough) rather than going to a public water source every time.	N/A

1.1.2.3: Structures

A structure is the physical location that contains the different facilities. The location of the structure is determined by exact data for facilities that aren't households or businesses. For households and businesses, the structure location is extrapolated based on the empirical survey data (Marras, 2008). Table 6 lists more state variables specific to structures.

Table 6: Structure state variables and description

State Variable	Variable Type	Description	Source
Max Facilities	Integer	The maximum number of facilities that can be held by this structure.	N/A

1.1.2.4: Parcels

Parcels are the cells of the environment and contain structures and their associated facilities and agents. Each parcel is originally given an ethnicity to help assist computation in the Schelling Segregation Model. Table 7 lists the state variables specific to Parcels.

Table 7: Parcel state variables and description

State Variable	Variable Type	Description	Source

Max Structures	Integer	The maximum number of structures that can be held by this parcel.	N/A
Location	Int2D	The (x,y) location of the parcel on the grid.	(OpenStreetMap, 2013)

1.1.2.5 Population

The population consists of the residents of the Kibera slum, with the necessary demographic information coming primarily from the Map Kibera project (Marras, 2008). The full list of population state variables is listed in Table 8.0:

Table 8: Parcel state variables and description

State Variable	Variable Type	Description	Source
Number of Residents	Integer	The total population size was estimated to be 250,000 residents.	(Maron, 2010)
Location	Int2D	The (x,y) location of the parcel on the grid.	(OpenStreetMap, 2013)

Of particular note, the exact population of Kibera is still a topic of dispute, with numbers from different sources ranging from 235,000 to 270,000 residents (Marras, 2008). The research chose a conservative population number of 250,000 residents as it represented the statistical average of the most respected sources to two significant figures (thus minimizing error).

1.1.2.6 Environment

The environment consists of the entire slum of Kibera, and all of the necessary spatial information to define its boundaries and features. GIS data was imported and translated and contained boundaries, road networks, and facilities. State variables are listed in Table 9:

Table 9: Parcel state variables and description

State Variable	Description	Source
Boundaries	Extracted from a shapefile (converted to ASCII) to define the extent of the environment	(OpenStreetMap, 2013)
Road Network	Extracted from a shapefile (converted to ASCII) that defines the available public roads of transportation. However, this is more for visual representation because the time scale used is large enough to automatically move agents between locations without going through specific paths.	(OpenStreetMap, 2013)
Facility Locations	Extracted from a shapefile (converted to ASCII) that gives the location of all the major facilities.	(OpenStreetMap, 2013)

1.1.3: Process Overview and Scheduling

The model process can be divided into three main parts: initialization, stepping process, and health process as shown in Figure 1. The simulation begins by first reading the necessary GIS files and then using the information to create the environment, grids, and displays. Next, sociodemographic information is read and processed to define the individual parameters of the agents (the residents of the Kibera slums). The agents are placed according to a modified Schelling Segregation Model (see section 1.4.2), which helps to create a racially diverse population (Schelling, 1971). It was necessary to conduct this type of segregation to reflect the real-life racial discrimination that occurs in Kibera (Smedt, 2009).

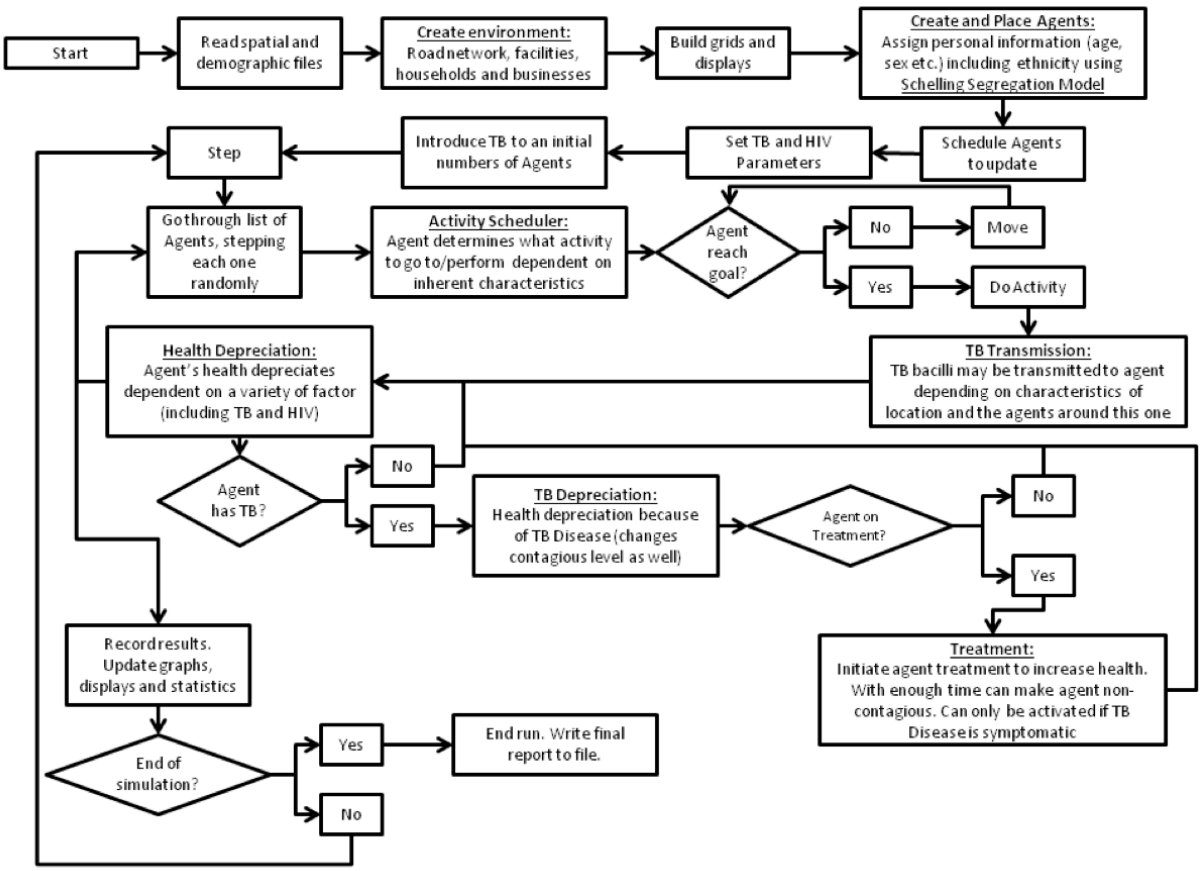

Figure 1: Flow diagram of the ABM

TB is introduced in the model by giving a certain percentage of the population either the latent infection, or the actual TB disease. Once initialized, the agents go through an activity scheduler to determine what activities they will conduct throughout the day. It is important to note that the agents are mobile and goal-oriented, so they will attempt to accomplish as many tasks as possible (and in their best interest) to maximize their utility. Figure 1 shows the full flowchart for the model.

Each simulation timestep in the model is equivalent to one hour (1 timestep/hour), while a year is defined as exactly 365 days (8,760 timesteps). Because of the longer timestep and the smaller region of Kibera, all agents spend exactly one timestep moving to their location. Doing so also saves computation power in path algorithm calculations, and allows for runs of a much longer length.

When the agents reach their goal location and begin conducting their specific activity, they become susceptible to contracting and spreading TB. The transmission is based on absorption and emission rate of a potential victim and the primary host, as well as certain environmental and disease factors (ex. how contagious is the individual). For simplicity this research does not model the transmission of TB among agents while they are moving because this is a much rarer occurrence and doing so would be too much of a computation burden. At the end of each timestep, all agents go through a health depreciation submodel, which determine the effects of the TB (or if applicable, HIV) based on the strength of the agent's immune system. The severity of TB will also be recalculated, and contagiousness may increase or decrease. If enough time has passed since the infection, TB may become symptomatic and can be diagnosed. If an agent only has latent TB, they may also develop active TB during this process.

If an agent's TB has been diagnosed and they have opted to take treatment (which not everyone will do since attrition rates are built in), their health will also improve during the length of the model. TB treatment takes the form of pills, and in this model, that intake occurs automatically. The magnitude of improvement is dependent on which regiment phase the agent is currently on, and the severity of his or her disease. After a certain number of weeks, agents with TB disease become non-contagious, though they can revert back to being contagious if they drop out of treatment. Agents with latent TB infections are all asymptomatic, and as a result, cannot be diagnosed or treated. While a typical treatment plan lasts for 6-9 months, this time is extended if the agent is HIV+ as they have to take additional medication to account for their low CD4 cell count. CD4 cell count for this research's purposes is defined as a numerical representation of the strength of an agent's immune system. Biologically, the value is the number of T-Cells per mm^3.

At the end of each timestep data is collected on each agent, general TB characteristics, as well as general population dynamics. The information is printed out as text files at the end of the model run.

1.2: Design Concepts

1.2.1: Theoretical and Empirical Background

Tuberculosis is an opportunistic airborne infectious disease caused by the bacteria *Mycobacterium Tuberculosis*. Even though it is treatable, TB is still prevalent today with one-third of the world population (~2 billion) infected with the latent form, and 1.3 million deaths occurring from the TB disease in 2012 (Centers for Disease Control and Prevention, 2012; Kwan and Ernst, 2011). Over 95% of these cases occur in developing nations, where poor health conditions facilitate the spread of the disease, and inadequate treatment options lead to unnecessary deaths. TB has therefore earned the notorious title as the second greatest killer worldwide from a single infectious agent (Centers for Disease Control and Prevention, 2012; Todar, 2012; World Health Organization, 2013a).

Mycobacterium Tuberculosis is a nonmotile, rod-shaped bacterium that has a reproduction rate of 15-20 hours (Todar, 2012). The primary form of TB (>90% of cases) is pulmonary TB, and is classified when the bacterium infects the pulmonary alveoli in the lungs (versus extrapulmonary tuberculosis which is when the bacteria spreads outside of the lungs). This causes the infected host patient to emit droplet nuclei with *M. Tuberculosis* bacilli when talking, coughing, etc., which furthers the spread of the infection (Tayler-Smith *et al.*, 2011). An individual needs to absorb only a few of these bacteria to become infected, and once inside, they move up to the lungs with the help of alveolar macrophages (Herman *et al.*, 2006). Because of its dominance as the primary form of TB, this research only studies and models pulmonary TB.

There are two types of TB, a latent infection, and an active TB disease. Latent TB infection is classified anytime an individual actually absorbs and is infected by the bacteria. A latent TB infection is asymptomatic (does not show symptoms) and is non-contagious. A certain

percentage of these cases, however, develop into the active TB disease. This is defined when the rate of TB reproduction surpasses the rate at which the immune system can eliminate the bacteria. Here, TB *is* contagious, and after a certain period of time, symptoms will show through coughing, hemoptysis, and chest pain (World Health Organization, 2013a).

Both forms of TB, the latent infection and active disease, are treatable, though treatment takes a significant amount of time (anywhere between 6-9 months), and requires multiple lines of drugs (Center for Substance Abuse Treatment, 1995; "Transmission and Pathogenesis of Tuberculosis", n.d.). The length of time dissuades many infected individuals from beginning treatment, or they may simply refuse treatment to avoid social stigma that may be associated with it (Tayler-Smith *et al.*, 2011). Furthermore, since symptoms of the disease disappear after several weeks of starting treatment, many patients falsely believe they are cured and quit treatment even though the bacteria may not be completely eradicated. This can lead to a recurrence of the disease and the patient may become contagious once again. The new form of TB that develops is also generally resistant to the traditional lines of defense, and can lead to the development of multi-drug resistant tuberculosis ("Transmission and Pathogenesis of Tuberculosis", n.d.). MDRT was not modeled in this study.

Since TB is an opportunistic infection, it takes advantage of a weakened immune system, and is thus often co-infected with HIV (Bevilacqua *et al.*, 2002). In fact, HIV has been called the driver of TB as a patient contracted with HIV increases the risk of developing active TB disease by 21 to 34 times the normal rate (Kwan and Ernst, 2011; World Health Organization, 2013a). For HIV patients, the number of CD4 cells (white blood cells that fight bacteria such as *M. Tuberculosis*), is critical in determining whether or not TB disease will occur after infection (Bauer *et al.,* 2008). An HIV positive patient has a CD4 cell count <= 350 cells/mm^3 (compared to an average count of 500-1500 cells/mm^3), and untreated HIV continues to diminish the CD4 cell count by an average of 50-80 cells/mm^3 ("CD4 and Viral Load Monitoring", 2011; Bartlett and Gallant, 2007). The risk of developing active TB disease from a latent infection increases dramatically when the CD4 cell count is between 282 and 314 cells/mm^3 (Rodrigues *et al.*, 2003).

1.2.2: Individual Decision Making

All decisions are made by individuals. All agents are independent, mobile, and goal-oriented, and each makes a decision about which activity it chooses to conduct. These decisions are based on personal characteristics (age, sex, location, etc.) as well as more specific health characteristics (such as being HIV positive or dehydrated). A weighted system is used to determine which activity will yield maximum utility for the agent. Once an activity is determined, the location of the goal (x_g,y_g) is chosen to minimize the Euclidean distance *d* from the agent's current position (x_i,y_i):

$$d = \sqrt{(x_g - x_i)^2 + (y_g - y_i)^2}$$

The only exception to the aforementioned condition is if an individual chooses to visit a business. Then the determination of the location includes the popularity of the business (which in turn uniformly distributed). To save computation power and because of its irrelevancy to the research question, no path algorithm was implemented and agents spend exactly one timestep going from one location to another.

If an agent contracts and develops active TB disease, they may begin treatment or may independently choose not to. At any given time during the treatment, they may drop out, with the likelihood of attrition increasing as the disease becomes less and less severe. Drop-out and treatment refusal rates are based on the most current literature regarding attrition rates in Kibera (Tayler-Smith *et al.*, 2011).

1.2.2.1: Goal Selection

Agents choose from a variety of activities to engage in. The following is a description of each of the goals (activities) that agents may select from depending on their personal characteristics.

<u>Goal Selection: Schools</u>

Agents between the ages of 3 and 18 are eligible to go to school. At the start of the model they are assigned the closest school to their home location. Schools are based on the Kibera schedule (data obtained by analyzing a variety of different Kibera school sites) and are open from Monday to Friday from 8 AM to 4 PM. Schools are closed July and August for summer break. An agent that is a student will have a much higher weight to attend this activity (but can be surpassed by other health or family needs).

Goal Selection: Religion

The two primary religions in Kibera (and the two modeled) are Christianity and Islam. An agent that is Muslim follows certain pre-established prayer times: Fajr (5:30 am), Sunrise (7:00 am), Dhuhr (1:00 pm), Asr (4:00 pm), Maghrib (7 pm), Isha (8:00 pm). The frequency of visitation of a religious facility increases with age, as it was assumed that older individuals have a higher tendency to worship. Christian agents have a higher tendency to worship during Sunday than other times of the week. This variable is also dependent on the age of the agent.

Goal Selection: Water

If an agent reaches a certain point of dehydration, it attempts to first utilize the water available for its family (in public places it's assumed this thirst is quenched). If there is not enough water at home, an agent will go to the closest water borehole to refill for itself as well as the family's water supply. The likelihood of taking this increased responsibility is a sin curve – essentially the younger and older agents are less likely to do this, while the healthier middle age groups will be more likely.

Goal Selection: Business

Researchers assume that most agents in the mid-age group are active participants in economic activities, and thus travel through businesses for either work or personal related issues. Agents are more likely to go on a weekend (when they have time off) and the weight is also again dependent on a sinusoidal age group curve. When picking a particular business to go to, agents

are more likely to go to a popular one in the area (determined at the start of the model from a uniform distribution).

Goal Selection: Socialization

Agents, not unlike humans, are social beings and thus are driven to socialize. All agents have a social network created at the start of the model. This network consists of family members, close neighbors, random friends, and members of the same ethnicity. During goal selection, there is a certain chance that an agent will choose to socialize with any other agent in its network. This varies depending on the time and day of the week as well as the age of the agent in question.

Goal Selection: Restaurants

For the purposes of this model, the researchers grouped bars/taverns in the overarching category of restaurants. Agents visit restaurants if they are actually open (certain percentage, the bars, are open extremely late) and with a general tendency to visit at least occasionally.

Goal Selection: Work

In Kibera, of those not attending school, 56.3% of females and 72.9% of men are employed. If an agent is employed, they are randomly assigned an occupation and an associated work location (one of the facilities). Agents work all five days of the week and 25% work on Saturday. A maximum working hour limit is set on each agent (based on Kenyan government regulations). Employed agents have a significantly higher weight for attending work.

1.2.3: Learning

In the model's current form, there is no individual learning included in the decision making process. Decision rules do change on temporal measurements such as day of the week and month of the year. Furthermore, there is no collective learning implanted in the model.

1.2.4: Individual Sensing

Agents are aware of their immediate characteristics (both personal and health), and basic household parameters (ex. household water, family members etc.). If an agent contracts and

develops active TB disease, they will be unaware of the fact until the disease emerges from the incubation period as symptomatic. Since latent TB infection is always asymptomatic, infected agents will never be aware of their contraction. All agents are, however, aware of their HIV status (positive or negative). Basic household needs are communicated with all members, and thus if an agent is thirsty and water is running low, they will replenish the entire household water supply in their trip. The spatial scale of sensing is local, and there are no costs for cognition or gathering individual sensing information.

1.2.5 Individual Prediction

Currently, there is no individual prediction of future conditions.

1.2.6 Interaction

Interaction is pivotal to this model because it is the only means of transmission of TB. Because of the necessity of person-to-person interactions, only direct interaction is modeled among all agents. When an agent conducts a specific activity and goal, there is a certain probability that they will interact with the other agents sharing the same parcel (grid cell). This probability is based on certain emission and absorption rates of the bacteria (see section 1.4.1). An infected agent will cough more, releasing TB bacilli in the air through their saliva and sputum. The absorption of the bacilli is dependent on stochastic measurements and the victim agent's absorption rate (ex. breathing rate, a personal characteristic).

Because of the racial differences present in Kibera, agents also choose to primarily interact with members of the same ethnicity during times of socialization.

1.2.7: Collectives

The two collectives represented in this model are 1) households and 2) ethnic groups, both of which are created at the start of the simulation. Households are analogous to families, and consist of a small group of agents who are related to and spend more time with one another. The second collective is the ethnic group, which consist of interconnected agents of the same ethnicity. Agents choose to interact more frequently with members of the same ethnicity during their

socialization periods. No collectives emerge during the simulation, but there is an observable grouping of agents of similar health statuses.

1.2.8: Heterogeneity

All agents are heterogeneous in the state variables age, gender, ethnicity, religion, and HIV and TB status. Depending on the age and gender variables, agents carry out roles in society as employees or students. The agent's HIV and TB status affects the decision-making process, as an agent's whose health is deteriorating will have a higher tendency to seek out treatment. An agent's ethnicity determines who it socializes with, while the religion variable affects when and how often it attends a religious facility.

1.2.9: Stochasticity

All heterogeneous agent traits are based on a uniform distribution during the initial assignment process. Once the initial facilities and structures are set (based on their GIS information), businesses are dispersed to maximize profitability. To maintain ethnic distribution, the first several agents of an ethnic group are randomly placed throughout the model, with the rest being placed with respect to the preset segregation rules (see section 1.4.1). Households are spatially inserted after agent initialization, and thus they are indirectly stochastic.

1.2.10 Observation

The full visualization window is shown in Figure 2.

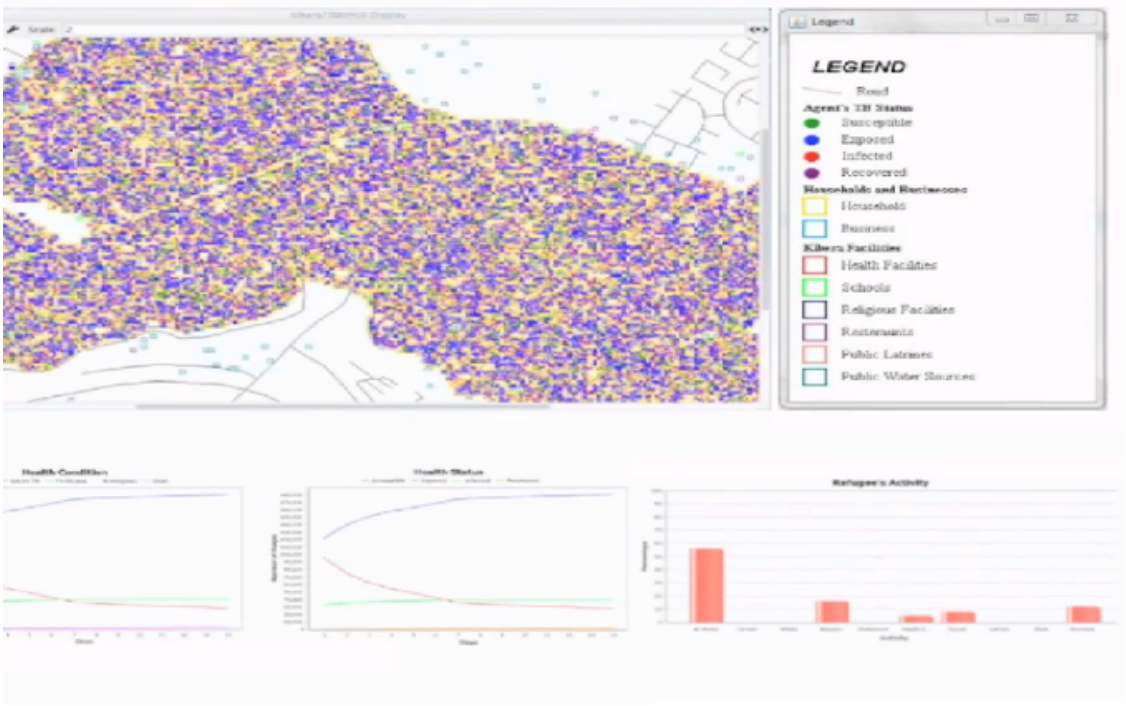

Figure 2: Visualization components of the model. From the left clockwise we show the GUI of the model, legend, health condition and health status charts, and current resident activity.

Data is collected at the end of each timestep and stored in CSV files for later analysis. The information collected includes the percent of the population that is: healthy, infected (latent TB), diseased (active TB), contagious, and dead. On the same timestep, the model is also recording information for which activities agents are currently conducting. In addition, the model records hotspots for infected and diseased agents, the age distribution for infected and diseased agents, and health status (based on the SEIR submodel) in terms of percent. However, the aforementioned measurements are recorded only every simulation day (24 timesteps) to ease computation. Graphs and timecharts are also displayed, but updated every 24 timesteps for the same computation reason. All these records combined are used in detailed analysis and understanding of the model.

1.3 Details

1.3.1: Implementation Details

The ABM was created in Java using the MASON toolkit (Luke *et al.*, 2005) and its geographical GeoMason extension (Sullivan *et al.*, 2010).

1.3.2: Initialization

The model begins by reading and interpreting socioeconomic and spatial data obtained from Map Kibera (2013) and the Map Kibera Project (Marras, 2008). In addition, a separate parameter file contains relevant information regarding TB transmission, growth, and treatment. The full set of input parameters and default values is summarized in Table 10.

Table 10: Default Parameter Values for Typical Runs

Parameter	Value	Source
Initial Number of Agents	250,000	(Marras, 2008)
TB Infection Prevalence Rate	0.35	(Tayler-Smith *et al.*, 2011)
TB Disease Prevalence Rate	0.064	(Tayler-Smith *et al.*, 2011)
HIV Prevalence Rate	0.14	(Patterson, 2012)
Treatment	True	N/A
Age Distribution	45% of the population is under 19	(Marras, 2008)
Latent to Disease Rate	7-10%	(“The Difference Between…”, 2012)
Incubation Period	336–2016 timesteps	(“Exposure to Tuberculosis”, 2005)

Time for Infection to Disease	0-17,520	(Center for Substance Abuse Treatment, 1995)
HIV CD4 Count	350 cells/mm^3	(Rodrigues *et al.*, 2003; "CD4 and Viral Load Monitoring", 2011)
HIV CD4 Count Drop	.00571-.00913 cells/(day*mm^3)	("CD4 and Viral Load Monitoring", 2011)
CD4 Count for TB	282–314 cells/ mm^3	(Bauer *et al.,* 2008)
TB Contagious Period	336-672 timesteps	(Centers for Disease Control and Prevention, 2013)
TB Testing Time	48-72 timesteps	("Testing for TB Infections", 2013)
TB Infection Dose	1-10 bacilli	(Herman *et al.,* 2006)
TB Saliva Bacilli Concentration	650,000 bacilli/ mL	Jones *et al.*, 2009
Saliva Per Cough	6 x 10^{-8} mL/ cough	Jones *et al.*, 2009
Coughs Per Hour	5-15	Jones *et al.*, 2009
Contagious Period After Treatment	336-2016 timesteps	(Centers for Disease Control and Prevention, 2013)

Attrition Rates Before Treatment	30%	(Tayler-Smith *et al.*, 2011)
Attrition Rates During Treatment	20.2%	(Tayler-Smith *et al.*, 2011)

When looking at specifically the parameter "Initial Number of Agents", it's important to point out that the exact number is still a matter of dispute, with sources ranging from 235,000 to 270,000 residents (Maron, 2010). The researchers chose a conservative population number of 250,000 residents as it represented the statistical average of the most respected sources to two significant figures. Furthermore, all above values are capable of being changed by directly manipulating the parameters file.

At time t=0, all parts of the model are in place, and TB is introduced into the population. The end display, sans agents, is displayed in Figure 3. The model display with agents and TB already introduced is displayed in Figure 4. A legend is provided in Figure 2. The model was designed to have similar general characteristics from run to run, while varying through the different possibilities in individual parameter settings.

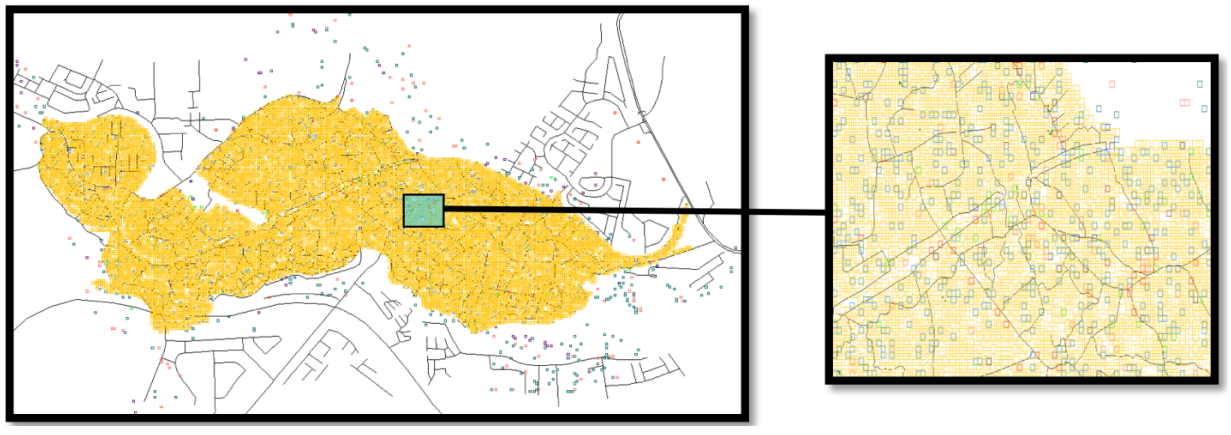

Figure 3: The model at initialization without agents. A zoomed in view is also provided.

Figure 4: The model at initialization with agents and TB already introduced. A zoomed in view is also shown.

1.3.3: Input Data

The model in its current form does not use any external data to represent dynamic processes that change over time.

1.3.4: Submodels

1.3.4.1: Schelling Segregation Submodel

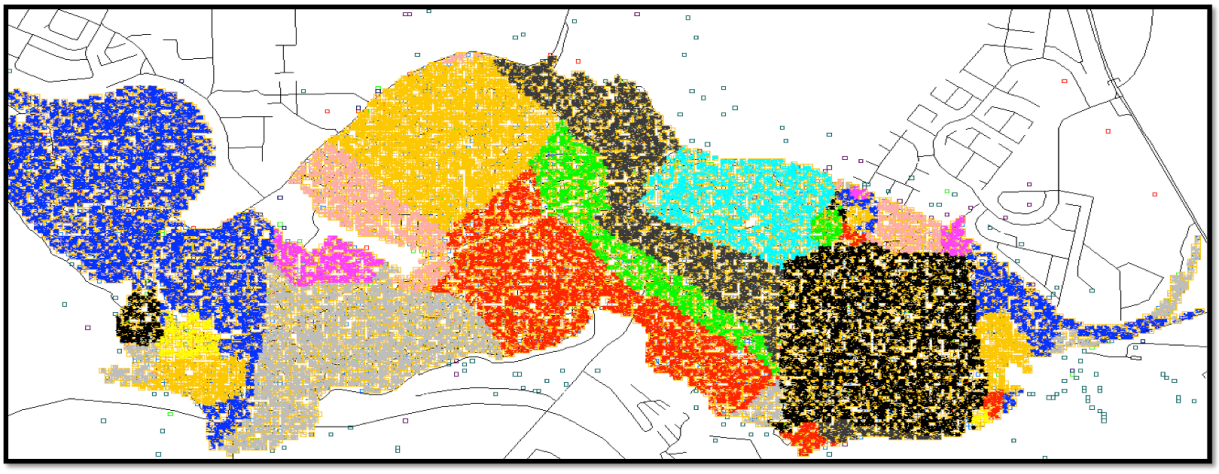

Figure 5: Ethnic distribution as conducted by the Schelling model of segregation. The ethnicities are represented as follows: Kikuyu (Blue), Luhya (Black), Luo (Red), Kalinjin (Orange), Kamba (Light Gray), Kisii (Cyan), Meru (Green), Mijikenda (Pink), Maasai (Magenta), Turkana (Yellow), Embu (Gold) and Other (Dark Gray).

Kibera represents a unique test case because it is very ethnically segregated as shown in Figure 5. There are 12 main ethnicities that divide up Kibera, and residents have a natural preference to socialize, work and interact with members of similar ethnicities. It was thus necessary to segregate the population and create reliable, interconnected social networks. In order to do this, the famous Schelling segregation model was used as an inspiration and was adapted to the ABM (Schelling, 1971). In the modified algorithm, the first families (consisting of a group of networked agents) choose households randomly in the area. The next set of households analyzes the Moore Neighborhood of current households, and determines a numerical "preference" for residing nearby (ethnicity being biggest factor), selecting the location that results in the greatest utility. The same method was applied to all agents' selection of initial locations, and the society was organically and naturally segregated. Figure 5 shows the fully completed segregation.

1.3.4.2 SEIR Submodel

Tuberculosis is an exogenous variable introduced into the model, and because of its complexity, a separate submodel modified of the SEIR model was utilized. The SEIR (Susceptible-Exposed-Infectious-Recovered) model has its roots in traditional mathematical epidemiological models and allows for a sophisticated yet simple way to characterize infectious diseases. Figure 6 shows a flowchart of the completed modified SEIR submodel used.

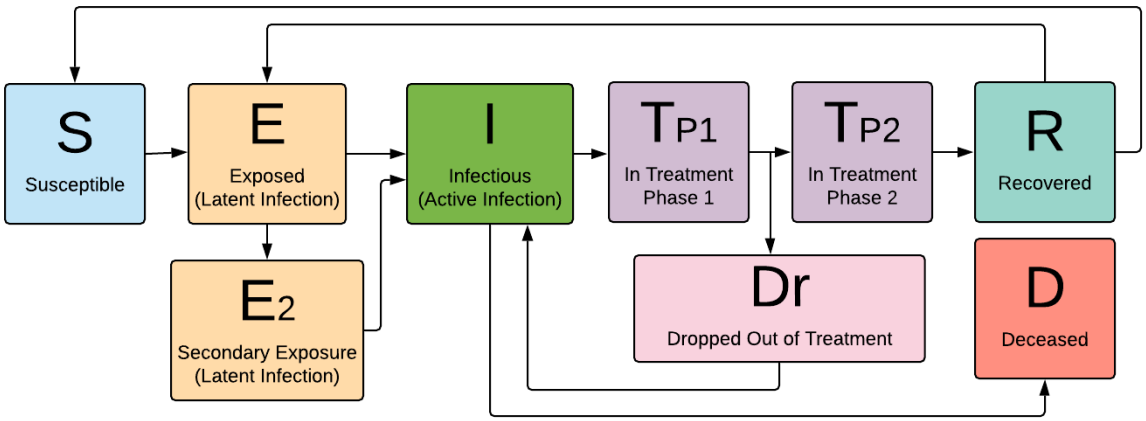

Figure 6: Flowchart for TB progression using a modified SEIR Submodel.

Agents are grouped in different compartments depending on what part of disease progression they are in. Susceptible agents are those that are prone to getting the disease (which is the entire agent population). "Exposed" agents have the latent TB infection (but are non-contagious), while "Infectious" agents have the active TB disease (and are contagious up to a period of time). Actively infected agents may choose to join a two-phase treatment program. It is possible for agents to drop out of the treatment, which returns them to the "Infectious" state. When agents complete the treatment program, they are considered "Recovered". A successful treatment either fully eliminates TB bacteria, or removes just enough to make the infection latent. Those agents with a persistent active infection will eventually become "Deceased". Although it is possible to recover from an active infection without treatment, the model does not currently reflect this self-cure process. Additional detail was added to the model to account for TB's unique characteristics regarding contagiousness, treatment, and HIV.

A key part of this submodel deals with the transmission and health depreciation that occurs as a result of TB. Transmission of the infectious disease specifically only happens when agents are conducting their activities and involved in personal interactions with one another. An individual, *i,* with TB disease will emit *M. Tuberculosis* bacilli at a certain rate of β_i based on the following calculation:

$$\beta_i = \sigma_i \gamma_i \varepsilon_i$$

Where σ_i is the bacilli concentration emitted, γ_i is the saliva per mL, and ε_i is the coughs emitted per hour by agent *i*. Because all interactions occur within a parcel of a certain size, the disease density for a parcel, λ_k must also be calculated in the equation:

$$\lambda_k = \frac{\sum_{i=0}^{n} \beta_i}{\iint dy\,dx}$$

Where *n* is the number of agents in parcel *k* and *dy* and *dx* represent the change in the dimension of the parcel. Using the disease density, the final TB bacteria absorption probability of another agent in the same parcel can be determined. This value labeled $X_{i,k}$ determines how many TB bacilli are absorbed by a victim agent in that parcel during one hour:

$$X_{i,k} \sim U(0, \lambda_k)$$

The total absorbed TB bacillus by an agent is defined as α_i in the equation:

$$\alpha_i = \sum_{k=0}^{p} X_k$$

Where *p* is all parcels that the agent has traveled to. Latent TB infection develops when the bacilli in the body is greater than the agent's infection dose (κ_i), or in other words, it develops when $\alpha_i \geq \kappa_i$

Latent TB is asymptomatic in the body and does not harm the body in any significant way. If an agent develops active TB disease however, then the rate of reproduction of the bacteria surpasses the rate at which the immune system can stop it, leading to depreciation in health. Furthermore, this depreciation is worsened if the agent is HIV positive. The calculation for health depreciation, δ_i is thus defined as:

$$\delta_i = \frac{(\psi_i)(\alpha_i) + (\psi_i)(\mu_i)}{\alpha_i + \mu_i} + T_R$$

Where ψ_i is the agent's total body health, α_i is the TB depreciation factor for agent *i*, and μ_i is the HIV depreciation factor for this agent. Notice that if an agent does not have HIV or TB, then α_i and μ_i will both be 0. T_R is health appreciation due to treatment where R is the regiment being taken by the agent. This value is zero if an agent is not on any treatment regimen.

References

Bartlett, J.G. and Gallant, J.E. (2007), *Medical Management Of HIV Infection 2007***,** Johns Hopkins University Press, Baltimore, MD.

Bauer, A.L., Hogue, I.B., Marino, S. and Kirschner, D.E. (2008), 'The Effects of HIV–1 Infection on Latent Tuberculosis'**,** *Mathematical Modelling of Natural Phenomena*, 3(7): 229-266.

Bevilacqua, S., Rabaud, C. and May, T. (2002), 'HIV-tuberculosis coinfection'**,** *Annales de Médecine Interne*, 153(2): 113-118**.**

Bonabeau, E. (2002), 'Agent-Based Modelling: Methods and Techniques for Simulating Human Systems'**,** *Proceedings of the National Academy of Sciences of the United States of America* (PNAS), 99(3): 7280-7287.

Centers for Disease Control and Prevention (2012), The Difference Between Latent TB Infection and TB Disease, Available at http://www.cdc.gov/tb/publications/factsheets/general/LTBIandActiveTB.htm [Accessed on 26th August, 2014].

Centers for Disease Control and Prevention (2013), *Core Curriculum on Tuberculosis: What the Clinician Should Know,* Centers for Disease Control and Prevention, Atlanta, GA, Available at http://www.cdc.gov/TB/.

CIA (2013), CIA World Factbook, Available at https://www.cia.gov/library/publications/the-world-factbook/ [Accessed on September, 10th, 2014].

Connell, R., Dawson, P. and Skvortsov, A. (2009), *Comparison of an Agent-based Model of Disease Propagation with the Generalised SIR Epidemic Model,* Air Operations Division,

DSTO Defence Science and Technology Organisation, Report DSTO-TR-2342, Fishermans Bend, Victoria.

Epstein, J.M. (2007), *Generative Social Science,* Princeton University Press, Princeton, NJ.

Grimm, V., Berger, U., Bastiansen, F., Eliassen, S., Ginot, V., Giske, J., Goss-Custard, J., Grand, T., Heinz, S., Huse, G., Huth, A., Jepsen, J., Jorgensen, C., Mooij, W., Muller, B., Pe'er, G., Piou, C., Railsback, S., Robbins, A., Robbins, M., Rossmanith, E., Ruger, N., Strand, E., Souissi, S., Stillman, R., Vabo, R., Visser, U. and Deangelis, D. (2006), 'A Standard Protocol for Describing Individual-Based and Agent-Based Models', *Ecological Modelling,* 198(1-2): 115–126.

Herman, P., Fauville-Dufaux, M., Breyer, D., Van Vaerenbergh, B., Pauwels, K., Thi, C.D.D., Sneyers, M., Wanlin, M., Snacken, R. and Moens, W. (2006), *Biosafety Recommendations for the Contained Use of Mycobacterium tuberculosis Complex Isolates in Industrialized Countries,* Division of Biosafety and Biotechnology, Scientific Institute of Public Health, Report No. D/2006/2505/22, Brussels, Belgium.

Jones, R.M., Masago, Y., Bartrand, T., Haas, C.N., Nicas, M. and Rose, J.B. (2009), 'Characterizing the Risk of Infection from Mycobacterium tuberculosis in Commercial Passenger Aircraft Using Quantitative Microbial Risk Assessment', *Risk Analysis,* 29(3): 355-365.

Kennedy, W.B., Hailegiorgis, A.B., Rouleau, M., Bassett, J.K., Coletti, M., Balan, G.C. and Gulden, T. (2010), 'An Agent-Based Model of Conflict in East Africa And the Effect of Watering Holes', *Behavior Representation in Modeling and Simulation (BRiMS) Conference*, Charleston, SC, pp. 274-281.

Kwan, C.K. and Ernst, J.D. (2011), 'HIV and Tuberculosis: A Deadly Human Syndemic', *Clinical Microbiology Reviews,* 24(2): 351-376.

Luke, S., Cioffi-Revilla, C., Panait, L., Sullivan, K. and Balan, G. (2005), 'MASON: A Multi-Agent Simulation Environment', *Simulation,* 81(7): 517-527.

Marras, S. (2008), Mapping the Unmapped. The Map Kibera Project, Available at http://mapkiberaproject.yolasite.com/resources/Kibera_mapping_the_unmapped.pdf [Accessed on 16th October, 2012].

Miller, J.H. and Page, S.E. (2007), *Complex Adaptive Systems,* Princeton University Press, Princeton, NJ.

Müller, B., Bohn, F., Dreßler, G., Groeneveld, J., Klassert, C., Martin, R., Schlüter, M., Schulze, J., Weise, H. and Schwarz, N. (2013), 'Describing Human Decisions in Agent-

based Models – ODD + D, An Extension of the ODD Protocol', *Environmental Modelling & Software,* 48: 37-48.

Mutisya, E. and Yarime, M. (2011), 'Understanding the Grassroots Dynamics of Slums in Nairobi: The Dilemma of Kibera Informal Settlements', *International Transaction Journal of Engineering, Management, & Applied Sciences & Technologies,* 2(2): 197-213.

National Center for Biotechnology Information (1995), *The Tuberculosis Epidemic Legal and Ethical Issues for Alcohol and Other Drug Treatment Providers,* reatment Improvement Protocol (TIP) Series, No. 18, Rockville, MD, Available at http://www.ncbi.nlm.nih.gov/books/NBK64541/.

North, M.J. and Macal, C.M. (2007), *Managing Business Complexity: Discovering Strategic Solutions with Agent-Based Modelling and Simulation,* Oxford University Press, New York, NY.

Orao-Obura, G. (2002), *Kenya: Facing the Challenge of Africa's Integration in the Global Economy: The Role of Multinational Enterprises in the Plantations Sector,* International Labour Organization, Working Paper No. 91, Geneva, Switzerland, Available at http://www.ilo.org/public/libdoc/ilo/2002/102B09_102_engl.pdf.

Patterson, H.L. (2012), *HIV/AIDS in the Slums of Kenya: Intervening Through Effectively Utilizing Volunteers (Doctoral dissertation,).* Master's Thesis, University of Pittsburgh, Pittsburgh, PA.

Peiperl, L. and Coffey, S. (2013), *HIV/AIDS.FAQ: How Long Can People Infected With HIV Expect to Live?,* U.S. Department of Veterans Affairs, Washington, DC, Available at http://www.hiv.va.gov/patient/faqs/life-expectancy-with-HIV.asp.

Rodrigues, D., Cunha, R., Kallas, E. and Salomao, R. (2003), 'Distribution of Naive and Memory/effector CD4+ T Lymphocytes and Expression of CD38 on CD8+ T Lymphocytes in AIDS Patients with Tuberculosis', *Brazilian Journal of Infectious Diseases,* 7(2): 161-165.

Segovia-Juarez, J.L., Ganguli, S. and Kirschner, D. (2004), 'Identifying Control Mechanisms of Granuloma Formation During *M. tuberculosis* Infection Using an Agent-based Model', *Journal of Theoretical Biology,* 231(3): 357-376.

Smedt, J. (2009), '‘No Raila, No Peace!’ Big Man Politics and Election Violence at the Kibera Grassroots', *African Affairs,* 108(433): 581-598.

Sullivan, K., Coletti, M. and Luke, S. (2010), *GeoMason: GeoSpatial Support for MASON,* Department of Computer Science, George Mason University, Technical Report Series, Fairfax, VA.

Tayler-Smith, K., Zachariah, R., Manzi, M., Kizito, W., Vandenbulcke, A., Sitienei, J., Chakaya, J. and Harries, A.D. (2011), 'Antiretroviral Treatment Uptake and Attrition Among HIV-positive Patients with Tuberculosis in Kibera, Kenya', *Tropical Medicine & International Health,* 16(11): 1380-1383.

Tiemersma, E.W., van der Werf, M.J., Borgdorff, M.W., Williams, B.G. and Nagelkerke, N.J. (2011), 'Natural History of Tuberculosis: Duration and Fatality of Untreated Pulmonary Tuberculosis in HIV Negative Patients: A Systematic Review', *PLoS One,* 6(4): e17601.

Todar, K. (2012), Mycobacterium Tuberculosis and Tuberculosis, Available at http://textbookofbacteriology.net/tuberculosis.html [Accessed on 26th August, 2014].

U.S. Department of Health and Human Services (2014), *Guide for HIV/AIDS Clinical Care,* U.S. Department of Health and Human Services, Rockville, MD.

Vanderbilt Occupational Health Clinic (2005), Exposure to Tuberculosis, Available at http://www.vanderbilt.edu/HRS/wellness/OHC/ohctb.pdf [Accessed on 26th August, 2014].

World Health Organization (2013a), TB/HIV Facts 2012-2013, Available at http://www.who.int/hiv/topics/tb/tbhiv_facts_2013/en/ [Accessed on 27th August, 2014].

World Health Organization (2013b), *Tuberculosis,* World Health Organization, Fact sheet No.104, Geneva, Switzerland, Available at http://www.who.int/mediacentre/factsheets/fs104/en/.

Wosyanju, C. (2009), *The System of Education in Kenya,* IUPUI Fulbright Hays Group Projects, Indiana University-Purdue University Indianapolis, Indianapolis, IN, Available at http://international.iupui.edu/kenya/resources/Education-in-Kenya.pdf.