A Four-Strain Model of Drug-Resistant Tuberculosis in the United States

Olivia Justynski

Abstract

Tuberculosis (TB) is a disease of global epidemiological concern. It is estimated that one-third of the world's population (including 11 million people in the US) are currently infected with TB [11]. Insufficient or irresponsible treatment of TB with antibiotics can select for drug-resistant bacteria, which are much more difficult to treat [9]. In particular, Multi-Drug-Resistant (MDR) and Extensively-Drug-Resistant (XDR) cases of TB require treatment regimens that are expensive, long lasting, toxic, and often unsuccessful [24]. Despite the importance of drug-resistance to understanding the current state of TB epidemiology, many published models of TB do not take resistance into account.

In this project, a compartmental mathematical model of TB epidemiology is presented. The model consists of four strains of TB, including one drug-susceptible strain, two strains that are each resistant to a single drug, and one MDR strain. This model fits accurately to several sets of relevant data collected by the CDC in the years 2000-2013, improving upon some previous predictions for the transmission of TB in the US. It also predicts the efficacy of various interventions with the goal of reducing the incidence of TB and MDR TB in particular.

The effects of interventions on TB epidemiology are modeled by modifying relevant parameter values starting at the year 2015 and comparing the projected incidence of TB. The most promising interventions for reducing TB and MDR TB incidence are decreasing treatment time, decreasing the potential for new infections via quarantine, and decreasing LTBI cases in the immigrant population. However, complete elimination of TB is not feasible for the foreseeable future.

A Four-Strain Model of Drug-Resistant Tuberculosis in the United States

Olivia Justynski

An honors thesis presented for the degree of
Bachelor of Arts

Department of Mathematics and Statistics
Mount Holyoke College
May 5, 2016

Acknowledgement

I would like to express my thanks to my advisor, Dylan Shepardson, for his indispensible help over the course of this project.

I would also like to thank my committee members, Robert Dorit and Mark Peterson, for their time and insight.

I additionally thank my collaborators, Chenyue Lu and Ellie Mainou, with whom this project was begun.

Finally, thank you to the Mount Holyoke College Mathematics and Statistics department for a wonderful four years.

Contents

List of Figures

INTRODUCTION

0.1 TB in the US

Tuberculosis (TB) is caused by *Mycobacterium tuberculosis*, which usually affects the lungs, though infection is also possible in the brain, kidneys, or spine [9–11]. TB is spread by bacteria in the air; when a person with active TB coughs, sneezes, shouts, or sings, the bacteria from their lungs or throat are expelled and can infect others who breathe them in [9–12]. Symptoms of TB include weakness, decreased appetite, weight loss, fatigue, fever, night sweats, chest pain, coughing, and hemoptysis [9–12]. According to the most recent data, in the US, there were 9,421 reported cases of TB in 2014 and 555 TB deaths in 2013 [14].

One-third of the world's population is estimated to be infected with TB, including 11 million in the US, making it an important concern of public health [11]. Most of these individuals have Latent TB Infection (LTBI), which is caused by TB bacteria existing at low levels in the body without causing symptoms or other physical evidence of TB disease; these individuals are not infectious [10, 11]. Out of all individuals with LTBI, only 5-10% will progress to active TB disease if left untreated, though treatment (which lasts between three and nine months) can prevent

progression to active disease [10, 11].

Active TB results from either reactivation or exogenous infection. Reactivation occurs when an individual with LTBI progresses to active disease. Exogenous infection occurs when an individual (who may be uninfected or already latently infected) is newly infected with TB and immediately progresses to active disease with a trivial latent period.

Active (symptomatic) TB is usually treated with a combination of antibiotic drugs, including Isoniazid and Rifampin [29]. The World Health Organization (WHO) recommends a program called DOTS (Directly Observed Therapy - Short-course strategy), which requires the patient to be supervised by a healthcare worker while taking antibiotics in order to ensure adherence to the regimen [29]. Treatment can last for six to twelve months, but may be extended beyond this timeframe for various reasons [11]

There are three priorities for controlling TB in the US that are generally agreed upon: 1) Individuals with active disease must be identified and treated to prevent further transmission; 2) Those at risk for exogenous infection due to proximity to an active patient must be treated for either active or latent TB; 3) Treatment for LTBI must be made available in order to prevent LTBI cases from progressing to active disease, which requires effective testing [1, 5, 6, 8, 27].

In 2014, two-thirds of new cases of TB in the US occurred in the foreign-born population [14]. In fact, generally most US cases of TB occur in foreign-born individuals; cases in foreign-born individuals are also more likely to be due to reactivation than those in US-born individuals (83.7% of foreign-born cases are due to reactivation), which suggests that individuals may become latently infected in

their home country and progress to active disease after they have relocated to the US [5, 13, 27, 28].

Currently, all prospective permanent entrants to the US are tested for active TB via a chest X-ray and a skin or blood test to detect immune response to *M. tuberculosis* [23]. If the chest X-ray is normal but the skin or blood test indicates the presence of bacteria, the results are indicative of LTBI. In this case, further treatment is recommended but not required, and the individual is cleared for entry to the US [23]. In order to decrease LTBI rates in the US, Cain *et al.* recommend screening specifically for LTBI in foreign-born entrants from countries with high levels of endemic TB and providing preventative treatment before entry to the US [5].

If such a test were to be implemented, it would have to be done with rigorous cultural sensitivity, as it has been argued that such screening is ultimately both ineffective and discriminatory [4, 6]. Cain *et al.* conclude, however, that it will be impossible to eliminate TB in the US until the LTBI burden in the foreign-born population is confronted [6].

0.2 Drug-Resistance

Drug-susceptible TB is treated with a combination of four antibiotics: Isoniazid, Rifampin, Ethambutol, and Pyrazinamide; these first-line drugs are the least toxic and most effective of anti-TB drugs [24]. If these drugs are administered incorrectly; that is, if the patient does not take all the medicine, the doctor prescribes it incorrectly, the patient's access to the drugs is unreliable, or the drugs themselves are of poor quality, then the patient may acquire drug-resistance [9].

If a patient has drug-resistant TB, additional drugs will be administered that may be more toxic, less effective, more expensive, and have longer treatment times [24]. Generally, as drug-resistance increases, the survival and cure rates for TB strains decrease considerably [22]. Despite accounting for a small number of total cases of TB, drug-resistant strains are much more difficult and costly to treat [17].

Multi-Drug-Resistant (MDR) TB is resistant to Isoniazid and Rifampin, which are the two most powerful drugs used to treat TB [9, 12]. In 2014, 1.3% of TB cases in the US were MDR [14].

Extensively-Drug-Resistant (XDR) TB is a very rare type of MDR TB [9, 12]. XDR TB is resistant to Isoniazid, Rifampin, at least one fluoroquinolone (a type of antibiotic), and at least one injectable second-line drug [9, 12]. Treating XDR TB therefore requires the use of other drugs that are not as effective, have greater risk of side effects, and cost more [12]. These drugs are also generally harder to administer; for example, they may be injectable rather than taken orally [30]. Between 1993 and 2011, there have been 63 reported cases of XDR TB in the US [12].

Out of worldwide new TB cases, it is estimated that 3.3% are MDR, of which 9.7% are XDR [24]. Alarmingly, among cases that have received prior treatment for TB, 20% are MDR, which indicates that unsuccessful treatments are associated with drug-resistance [24]. In the US, resistant strains are more common in foreign-born than US-born individuals [5].

Some strains of XDR TB are resistant to even more drugs, and it has been proposed to extend the classification system to include these "Extremely" or "Totally" drug-resistant strains; however, no classification beyond XDR TB has yet been defined by the international community [21].

One study in Iran found fifteen cases of XDR TB that were resistant to all drugs tested and could not be successfully treated; they defined these cases as Totally drug-resistant (TDR) [34]. All of the TDR cases occurred in individuals who had previously received treatment for TB and were genetically unrelated to any of the other TDR cases. Since the cases of TDR TB were unrelated, it is unlikely that they resulted from a few original TDR strains that were then transmitted to the other patients. Rather, these highly resistant cases likely arose independently in each of the fifteen cases; that is, failed treatment resulted in the evolution of TDR bacteria in fifteen separate cases. Velayati *et al.* suggest that the prevalence of these TDR TB cases could be a result of the common use of anti-TB drugs for other respiratory diseases in Iran [34].

This project focuses on resistance to Isoniazid and Rifampin, as these are the most common anti-TB drugs, and resistance to Isoniazid is the most common form of drug-resistance; however, it is important to note that MDR TB may include resistance to many more drugs beyond these [25]. A study at Hospital Pulido Valente (a hospital in Portugal that focuses on pulmonary medicine) found that on average, MDR TB cases exhibited resistance to seven anti-TB drugs, not just the two required to define MDR cases [35].

The *M. tuberculosis* genome spontaneously accumulates mutations that cause drug-resistance [16]. However, the mutation rate is slow enough that multiple mutations for drug-resistance would almost certainly not arise in the same bacterium by chance in order to produce MDR TB; rather, MDR TB evolves by a bottlenecking process upon unsuccessful antibiotic treatment [17]. Indeed, it has been observed clinically that TB strains initially resistant to a single drug can acquire resistance

to other drugs during treatment [20].

Even with excellent adherence to treatment regimens, however, MDR can still occur [29]. Srivastava *et al.* hypothesized that this is due to pharmacokinetic variability, that is, the inconsistent rates at which individual patients will absorb anti-TB drugs [29]. Patients are given combinations of anti-TB drugs with the goal of avoiding monotherapy, or treatment by a single drug, which is likely to lead to resistance. However, if a patient absorbs some drugs unexpectedly quickly, the treatment may not be in effect for the predicted timeframe. If a group of patients is given a combination of anti-TB drugs on the same dosage and schedule, some subset of patients will absorb the drugs at a more rapid rate [38], and so for an unintended period of time will be exposed to only one type of drug (monotherapy) or be effectively untreated, creating an environment conducive to the acquisition of drug-resistance [29].

0.3 Hill Model

I have used the compartmental differential equations model of TB transmission developed by Hill *et al.* in 2012 as a basic template for my model of TB. A compartmental model separates a population into groups, called compartments, based on their disease state and other descriptive factors. The Hill model divides the US population into US-born (USB) and foreign-born (FB) subpopulations, which have different rates of LTBI prevalence and other relevant factors. These two groups also exhibit preferred mixing; that is, USB and FB individuals are more likely to contact other USB and FB individuals, respectively. This means that most exogenous infections occur within these groups, rather than across them [19].

An outline of their model is as follows, and the schematic is shown in Figure 1. Most individuals reside in the Susceptible compartment. Upon exogenous infection, an individual may move to a compartment for Acute LTBI or Chronic LTBI. Acute LTBI progresses to active disease within two years, while Chronic LTBI progresses more slowly. An individual in the Chronic LTBI compartment can also be re-infected and move to the Acute LTBI compartment. Individuals in both LTBI compartments may progress to either Infectious TB or Non-Infectious TB, at which point their TB is active. Treatment or self-cure can result in movement back into the Susceptible compartment from either of the LTBI or active TB compartments. All births enter the Susceptible compartment; FB arrivals may enter the Susceptible or LTBI compartments. Individuals exit from every compartment due to unrelated death; TB death only affects the active TB compartments [19].

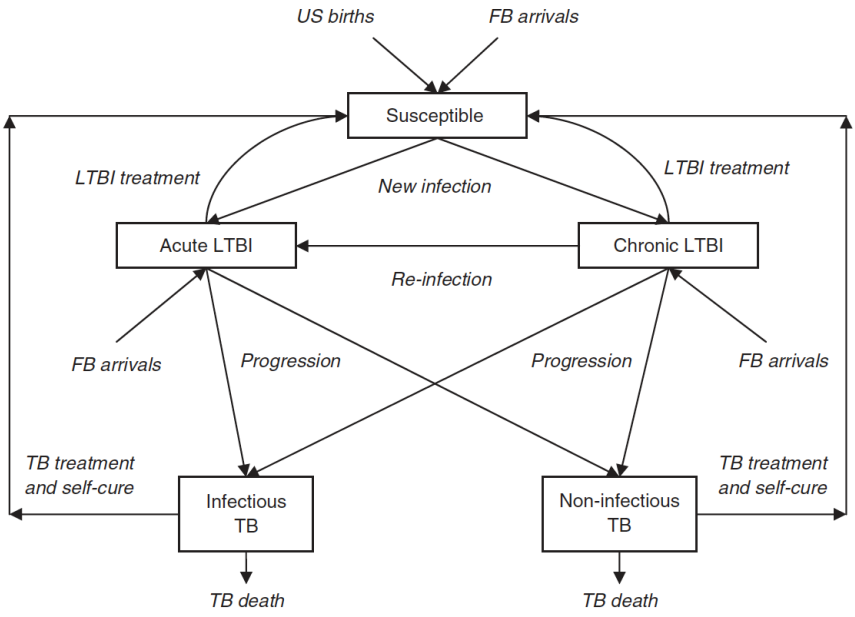

Figure 1: Schematic of the Hill model from their 2012 paper [19]

In their analysis, Hill *et al.* fit their model to data collected by the CDC on the incidence of active cases in each subpopulation for the years 2000-2008. They focused primarily on the possibility of eliminating TB from the US by the year 2100, where elimination is defined as an annual incidence of fewer than one case per million of population. They found that elimination by this date would be possible for the USB population, but not for the FB population, and therefore, not for the overall US population. Even if transmission of TB were completely halted, elimination would not be reached by 2100, because the influx of LTBI cases from immigration is too great [19].

There are two main areas in which I have extended the Hill model. The first concerns the addition of drug-resistance. Drug-resistance is a critical component of TB epidemiology and dynamics. Resistant cases have different rates of treatment and cure than drug-susceptible strains, and the Hill model does not yet account for these qualities. The second is the method of fitting the model. Hill *et al.* fit only to data for incidence of active TB [19]. I fit my model to these data, as well as other data from the CDC, including cumulative TB death. In Section 6.1, I will compare the Hill model's prediction to these data and show that the model overestimates the death rate due to TB. Therefore, its predictions could be potentially unreliable, though defense of the Hill model's TB death rate is possible. In contrast, my model was fit to the TB death data, which may increase accuracy.

0.4 Resistance Models

In contrast to the model presented by Hill *et al.*, some published mathematical models of TB transmission have included dynamics of drug-resistance. Several significant models and their findings are presented below.

Cohen *et al.* 2004: This model takes into account three strains of TB: Drug-susceptible, fit MDR, and unfit MDR. The fitness of a strain describes the bacteria's ability to survive and infect new hosts, among other characteristics. It is hypothesized that MDR bacteria undergo a fitness cost when they acquire resistance, so Cohen *et al.* investigate whether even an unfit MDR strain could be epidemiologically dangerous. They conclude that even with such a cost, an MDR strain would still sufficiently outcompete drug-susceptible strains to be a public health danger [15].

Basu *et al.* 2007: This model focuses on TB in and around South African hospitals, where many cases of TB are exogenously acquired. The model takes into account three strains of TB: Drug-susceptible, MDR, and XDR. Basu *et al.* conclude that without serious changes to the strategies used to combat TB, XDR TB will become a major public health threat. However, the model does indicate that some strategies to prevent infections within hospitals could drastically improve this outlook [2].

Bhunu 2011: Like Basu *et al.*, this model includes drug-susceptible, MDR, and XDR TB [2, 3]. Bhunu introduces the possibility of a Quarantine compartment for XDR cases, as well as different treatment plans. He finds that quarantining XDR cases reduces transmission, but acknowledges that this has practical and ethical

conflicts [3].

Generally, these models assume a direct change from drug-susceptible TB to MDR TB. However, it is known that resistance is acquired discretely, starting with resistance to a single drug [20]. My model takes this into account by including strains that are resistant only to a single antibiotic anti-TB drug, which may then progress to MDR TB with additional acquisition of resistance.

Chapter 1

SINGLE-STRAIN MODEL

1.1 Model

I begin by constructing a model for a single strain of drug-susceptible TB. This model takes the form of a simple SEI (Susceptible, Exposed, Infectious) compartmental model. The single-strain model consists of a series of ordinary differential equations coded in RStudio. I will not include the equations for the single-strain model; however, the equations for the four-strain model are included in Section 3.3.

This model does not yet take drug-resistance into account. A schematic diagram of this model is shown in Figure 1.1.

There is a single Susceptible compartment, which contains all uninfected individuals. This compartment includes all individuals who have never been infected as well as all individuals who have completely recovered from an infection.

There is a single Exposed compartment, which contains all individuals with LTBI. These individuals harbor *M. tuberculosis* in their lungs, but they are neither

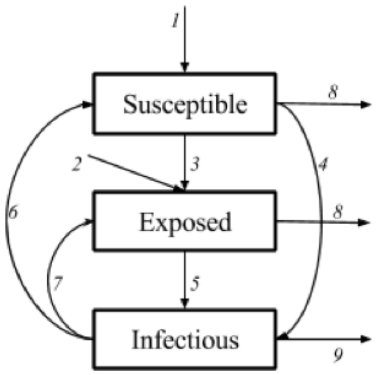

Figure 1.1: Schematic of the single-strain compartmental model
1: Individuals enter the Susceptible compartment through birth and immigration. *2*: Immigrants who have LTBI enter the Exposed compartment directly. Individuals with active TB are not permitted to immigrate to the US, so immigration does not directly affect the Infectious compartment. *3*: Individuals leave the Susceptible compartment to enter the Exposed compartment when they become latently infected by contact with an Infectious individual. *4*: Individuals leave the Susceptible compartment to enter the Infectious compartment when they become immediately actively infected through contact with an Infectious individual. *5*: Individuals with LTBI progress to active disease after some period of time and enter the Infectious compartment. *6*: If treatment is successful, Infectious individuals return to the Susceptible compartment. *7*: If treatment is unsuccessful, Infectious individuals have not been cured and therefore return to the Exposed compartment. *8*: Individuals exit the Susceptible and Exposed compartments due to unrelated deaths. *9*: Individuals exit the Infectious compartment due to unrelated deaths as well as deaths from TB.

symptomatic nor infectious.

There is a single Infectious compartment, which contains all actively infected individuals. These individuals are symptomatic and infectious. Since this model concerns the US population, it is assumed that all of these individuals are currently receiving treatment. It is also assumed that patients are actually infectious only for the early portion of their treatment.

Note that going forward, "Infectious" will specifically denote individuals in an

Infectious compartment. The lowercase "infectious" will denote the quality of being able to infect others. All infectious individuals are within an Infectious compartment, but not all individuals in an Infectious compartment are infectious. There are some individuals who are actively infected (and therefore, placed in the Infectious compartment), but are not able to infect others. Therefore they are Infectious but not infectious.

It is assumed that there is homogeneous mixing within the US population. This is important because it dictates that each individual is equally likely to make contact with any other individual in the population. Therefore, the rate of new infections is directly proportional to the proportions of Susceptible and Infectious individuals in the population. This departs from the Hill model, which assumed preferred mixing within the groups of US-born and Foreign-born [19]. Since my model does not make this distinction, I assume that there is no such preference.

1.2 Fit

Using an algorithm in RStudio that generates randomized sets of parameters within some realistic range, I fit this model to two sets of data collected by the CDC for the years 2000-2013 [14]. As shown in Figure 1.2, the incidence of TB per year and cumulative TB death fit well with the model's prediction.

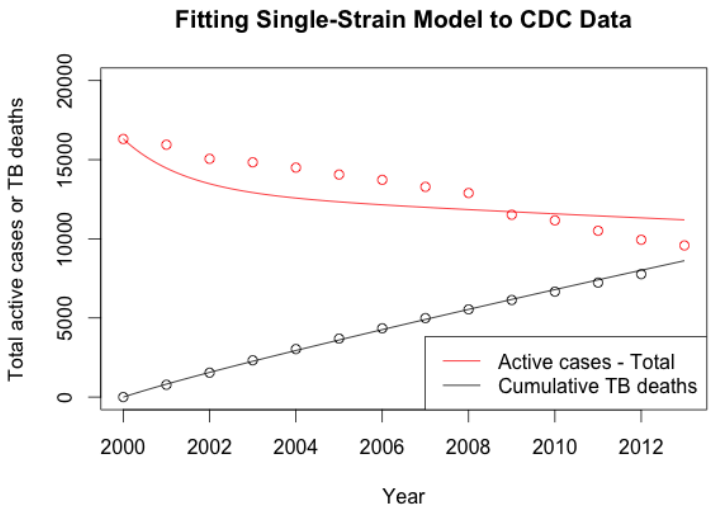

Figure 1.2: Single-strain model fit to CDC data for total incidence of active cases and cumulative TB death for the years 2000-2013
Open circles represent CDC data points [13], while solid lines represent the model's prediction.

Chapter 2

TWO-STRAIN MODEL

2.1 Model

Next, I construct a model that incorporates two strains of TB. For simplicity, one strain is assumed to be entirely drug-susceptible, while the other is assumed to be MDR, that is, resistant to both Isoniazid and Rifampin. This model also follows the format of an SEI compartmental model, but with additional Exposed and Infectious compartments. The two-strain model consists of a series of ordinary differential equations coded in RStudio. I will not include the equations for the two-strain model; however, the equations for the four-strain model are included in Section 3.3.

A diagram of this model is shown in Figure 2.1.

There is a single Susceptible compartment, similar to the single-strain model.

There are two Exposed compartments, which together contain all individuals with LTBI. The TB of individuals in the Exposed (Drug-susceptible) compartment is not resistant to any antibiotic drugs, while the TB of individuals in the Exposed

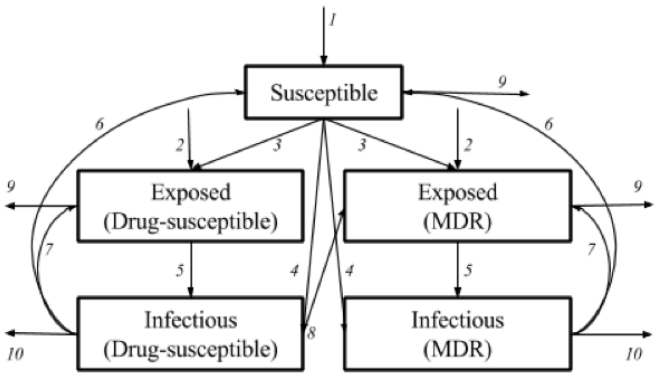

Figure 2.1: Schematic of the two-strain compartmental model
1: Individuals enter the Susceptible compartment through birth and immigration. *2*: Immigrants who have LTBI enter one of the Exposed compartments directly. *3*: Individuals leave the Susceptible compartment to enter an Exposed compartment when they become latently infected through contact with an Infectious individual. Exogenously infected individuals always enter the compartment of the same strain as the individual who infected them. *4*: Individuals leave the Susceptible compartment to enter an Infectious compartment when they become immediately actively infected through contact with an Infectious individual. Again, these individuals always enter the compartment of the same strain as the individual who infected them. *5*: Individuals with LTBI progress to active disease after some period of time and enter the corresponding Infectious compartment. *6*: If treatment is successful, Infectious individuals return to the Susceptible compartment. *7*: If treatment is unsuccessful, but resistance is not acquired, Infectious individuals return to the corresponding Exposed compartment. *8*: Individuals who fail treatment for active drug-susceptible TB may acquire resistance. In this case, they will exit the Infectious (Drug-susceptible) compartment and enter the Exposed (MDR) compartment. *9*: Individuals exit the Susceptible and both Exposed compartments due to unrelated deaths. *10*: Individuals exit both Infectious compartments due to unrelated deaths as well as deaths from TB.

(MDR) compartment is resistant to at least Isoniazid and Rifampin.

This does not account for all possible cases. For example, in reality, an individual could exist who is infected with TB that is resistant to Isoniazid but not Rifampin. This model ignores the presence of these individuals as a necessary simplification; this will be rectified in the four-strain model presented in Chapter 3.

There are two Infectious compartments, divided similarly to the Exposed compartments.

The main difference in dynamics between the single-strain and two-strain models is in the acquisition of resistance. An individual in the Infectious (Drug-susceptible) compartment is assumed to be receiving treatment. The treatment, once completed, will have been either successful or unsuccessful. If treatment fails, the individual is still infected with TB, but the infection is no longer active. In some such cases, the ineffective treatment can select for drug-resistant bacteria. At this point, the individual enters the Exposed (MDR) compartment (See Figure 2.1, *8*).

2.2 Fit

Using a randomization algorithm in RStudio similar to that used for the single-strain model, I fit this model to three sets of data collected by the CDC for the years 2000-2013 [14]. As shown in Figure 2.2, the total incidence of active cases of TB per year, incidence of MDR cases per year, and cumulative TB death fit well with the model's prediction.

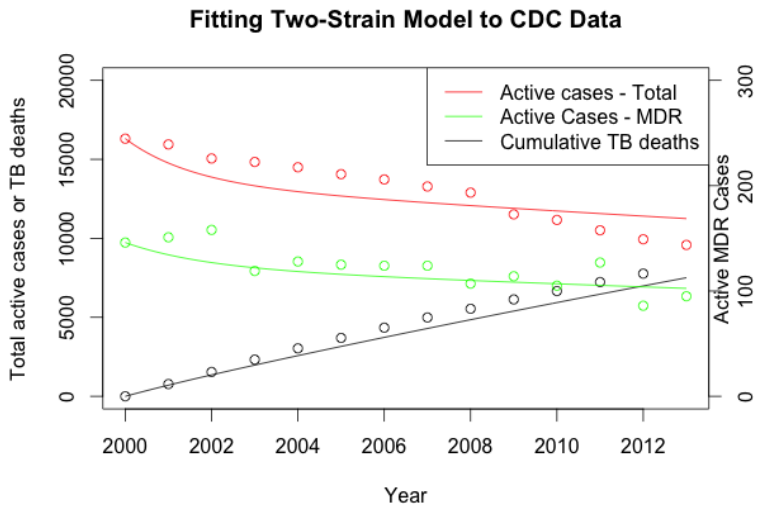

Figure 2.2: Two-strain model fit to CDC data for total incidence of active cases, MDR cases, and cumulative TB death for the years 2000-2013

Open circles represent CDC data points [13], while solid lines represent the model's prediction. Note that the right-hand axis applies to MDR cases, which occur at much lower quantities than total active cases and TB death do.

Chapter 3

FOUR-STRAIN MODEL

3.1 Model

Finally, I construct a four-strain model, which will be the main focus for the remainder of the project. My model incorporates four strains of TB. The first strain is entirely drug-susceptible. The H-resistant strain is resistant to Isoniazid but not Rifampin; the R-resistant strain is resistant to Rifampin but not Isoniazid. The fourth strain is resistant to both Isoniazid and Rifampin, making it MDR. Cases of TB that are resistant to these two drugs as well as additional antibiotic drugs will also be placed in the MDR category. Thus, in this model, all possible combinations of resistance are accounted for. The four-strain model consists of a series of ordinary differential equations coded in RStudio. These equations are included in Section 3.3. A diagram of this model is shown in Figure 3.1.

As before, there is a single Susceptible compartment.

There are four Exposed compartments that collectively contain all individuals

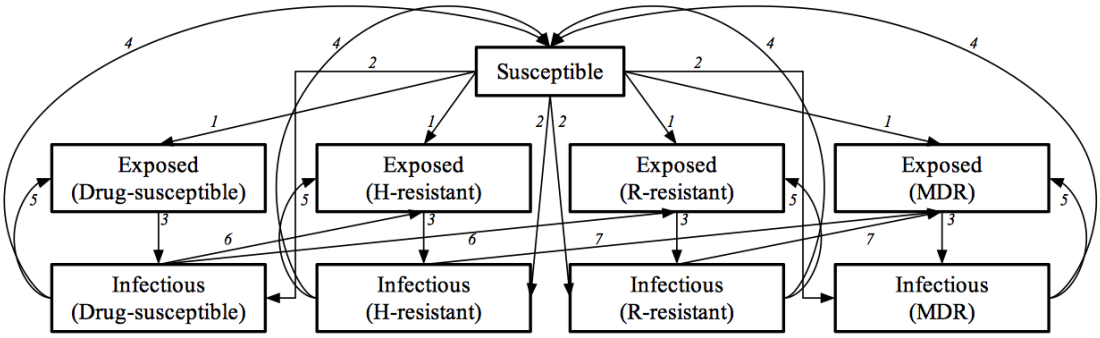

Figure 3.1: Schematic of the four-strain compartmental model

In this schematic, birth and death arrows are omitted for simplicity. As in the previous models, all births enter the Susceptible compartment. Immigrants enter the Susceptible or one of the Exposed compartments, depending on whether they have LTBI, and the strain with which they are infected. Individuals exit all compartments due to unrelated death, and exit the Infectious compartments due to death from TB. *1*: Individuals leave the Susceptible compartment to enter an Exposed compartment when they become latently infected through contact with an Infectious individual. Exogenously infected individuals always enter the compartment of the same strain as the individual who infected them. *2*: Individuals leave the Susceptible compartment to enter an Infectious compartment when they become immediately actively infected through contact with an Infectious individual. Again, these individuals always enter the compartment of the same strain as the individual who infected them. *3*: Individuals with LTBI progress to active disease after some period of time and enter the corresponding Infectious compartment. *4*: If treatment is successful, Infectious individuals return to the Susceptible compartment. *5*: If treatment is unsuccessful, but resistance is not acquired, Infectious individuals return to the corresponding Exposed compartment. *6*: Individuals who fail treatment for active drug-susceptible TB may acquire resistance. In this case, an individual exits the Infectious (Drug-susceptible) compartment and enter the Exposed (H-resistant) or (R-resistant) compartment. *7*: Individuals who fail treatment for active H-resistant or R-resistant TB may acquire additional resistance to become MDR. In this case, they will exit the Infectious (H-resistant) or (R-resistant) compartment and enter the Exposed (MDR) compartment.

with LTBI. They are divided into these four compartments by the level of resistance against the two main anti-TB drugs, Isoniazid and Rifampin.

Similarly, there are four Infectious compartments divided in the same manner.

The dynamics of resistance acquisition are similar to those of the two-strain model, but with more complexity. As before, an individual in the Infectious (Drug-susceptible) compartment is assumed to be receiving treatment, which may be unsuccessful. In some cases, ineffective treatment may select for bacteria that are resistant to one anti-TB drug. Resistance to only one drug may be acquired at a time, so the individual can move from the Infectious (Drug-susceptible) compartment to the Exposed (H-resistant) or Exposed (R-resistant) compartments, but not to the Exposed (MDR) compartment (See Figure 3.1, *6*). However, an individual in either the Infectious (H-resistant) or (R-resistant) compartments may, in the current round of treatment, acquire resistance to a second drug, and progress to the Exposed (MDR) compartment (See Figure 3.1, *7*).

This model differs from the Hill model in several key ways, which will now be discussed.

One deviation from the Hill model is in the progression from the Exposed to Infectious compartments. The Hill model includes two Exposed compartments, one for Chronic LTBI and one for Acute LTBI. After being exogenously infected, an individual may enter the Chronic LTBI compartment, which progresses to active disease very slowly, or they may enter the Acute LTBI compartment, which progresses to active disease more quickly [19]. In my model, there is only one Exposed compartment, and it corresponds roughly to the Hill model's Chronic LTBI compartment, in that it allows only slow progression. An individual whose TB progresses quickly moves directly from the Susceptible compartment to an Infectious compartment in my model, without passing through an Exposed compartment. This change

allows my model to contain four fewer compartments, which is an extremely useful simplification.

Similarly, Hill's model includes two compartments for active disease. Both compartments contain individuals with active TB, but only one is actually infectious; the other contains individuals with non-infectious TB [19]. In my model, I have instead included a parameter q to represent the proportion of active cases that are capable of infecting others, which gives an effectively similar result. This simplification again reduces the number of necessary compartments, which increases the model's efficiency.

Hill's model also includes dynamics of self-cure [19], which my model does not. At low rates, an infected individual's immune system may cure their TB without any outside treatment [31]. However, this occurs at rates that are so small as to be trivial, so their dynamics can be included in other aspects of the model, rather than explicitly stated, while maintaining accuracy.

While my model does not explicitly account for self-cure, much of this behavior is, in fact, included within the scope of the model. For example, individuals who self-cure from the active disease state are counted among those who receive successful treatment.

Individuals may also self-cure from the latent disease state, in which case they will not be counted as ever being latently infected in my model. There is little accurate data for the number of LTBI cases at any given time, since asymptomatic individuals are unlikely to be tested for TB infection. It is probable that more individuals than are accounted for in my model have LTBI, but some of them self-cure before progressing to active disease. In any case, they have a negligible effect

on the dynamics of Infectious TB upon which I am focusing, so they may be safely uncounted.

The Hill model also includes the possibility of re-infection [19], which my model excludes. Re-infection occurs when an individual with LTBI becomes exogenously infected again by an Infectious individual, which causes them to progress to active disease. In the Hill model, this is represented as movement from the Chronic to Acute LTBI compartment [19]. In my model, I do not distinguish between Chronic and Acute LTBI, so the progression from latent to active disease is indistinguishable whether it is due to natural disease progression or a new exogenous infection. Therefore, the possibility of re-infection need not be explicitly taken into account.

My model makes additional assumptions that will now be discussed.

It is assumed that after successful treatment and cure, individuals receive no immunity to TB. There is conflicting evidence for whether such immunity occurs in reality. Therefore, in this model, cured individuals reenter the Susceptible compartment and have an equal likelihood of being exogenously infected as individuals who have never had TB.

Finally, the model assumes that no individuals are vaccinated for TB. The Bacille Calmette-Guérin (BCG) vaccine may be used to prevent TB, but it is generally only administered to children in some countries outside the US [12]. In the US, the CDC does not recommend the use of BCG due to its limited effectiveness [12]. Therefore, it is reasonable to assume that no individuals in the US have been vaccinated against TB. Individuals who immigrate to the US may have been vaccinated, but since the vaccine is largely ineffective in adults, their transmission dynamics can be assumed to be unaffected by the vaccine.

As discussed, these assumptions do not decrease the accuracy of my model, but contribute to simplifying the transmission dynamics expressed in this simulation.

3.2 Fit

Using a similar randomization algorithm in RStudio to those used in the single-strain and two-strain models, I fit this model to four sets of data collected by the CDC for the years 2000-2013 [14]. As shown in Figure 3.2, the total incidence of active cases of TB per year, incidence of active H-resistant cases per year, incidence of active MDR cases per year, and cumulative TB death fit well with the model's prediction.

The process of obtaining this fit will be discussed in more detail in Chapter 4.

3.3 Differential Equations

The equations used to produce the four-strain model are below. The parameters used in these equations are listed in Table 3.1 with their descriptions and ranges. The parameter values for this model were constrained within certain acceptable ranges based on reasonable epidemiological assumptions.

S denotes the Susceptible compartment. E and I denote the Exposed and Infectious compartments, respectively. These compartments are indexed from 1 to 4 to describe the level of drug-resistance of the strain. That is, E_1 and I_1 refer to the Exposed and Infectious compartments of drug-susceptible TB; E_2 and I_2 refer to these compartments for H-resistant TB; E_3 and I_3 refer to R-resistant TB; E_4

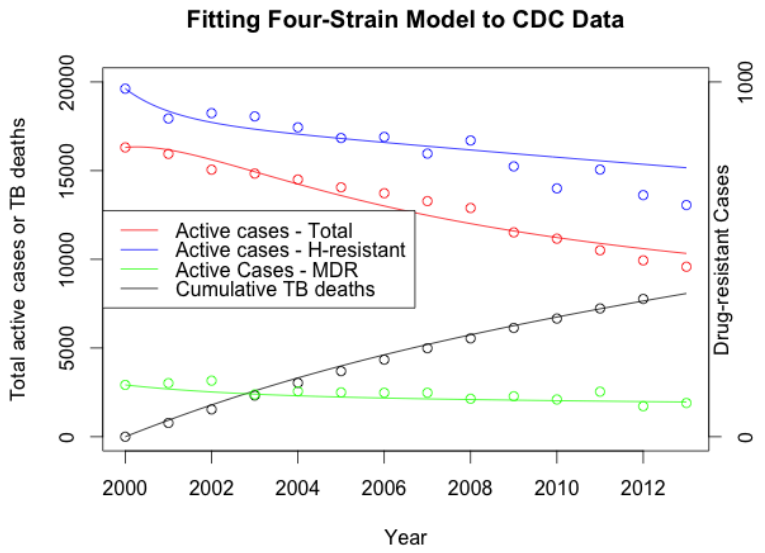

Figure 3.2: Four-strain model fit to CDC data for total incidence of active cases, H-resistant cases, MDR cases, and cumulative TB death for the years 2000-2013 [13].
Open circles represent CDC data points, while solid lines represent the model's prediction. Note that the right-hand axis applies to H-resistant and MDR cases, which occur at much lower quantities than total active cases and TB death do.

and I_4 refer to MDR TB. D represents a compartment for individuals that die of TB. This compartment is cumulative; unlike other compartments, individuals enter but do not leave. Finally, N denotes the total population of the US.

$$S' = \rho N - qt_1\lambda\frac{SI_1}{N} - qt_2\lambda\frac{SI_2}{N} - qt_3\lambda\frac{SI_3}{N} - qt_4\lambda\frac{SI_4}{N} + z_1\phi_1 I_1 + z_2\phi_2 I_2 + z_3\phi_3 I_3$$
$$+ z_4\phi_4 I_4 + (1-l)\,\alpha N - \mu_0 S$$
$$E_1' = (1-p)\,qt_1\lambda\frac{SI_1}{N} - v_L E_1 + (1-y_1)\,(1-z_1)\,\phi_1 I_1 + l\alpha\,(1-r_2-r_3-r_4)\,N - \mu_0 E_1$$
$$I_1' = pqt_1\lambda\frac{SI_1}{N} + v_L E_1 - \phi_1 I_1 - \mu I_1 - \mu_0 I_1$$

$$
\begin{aligned}
E_2' &= (1-p)\,qt_2\lambda\frac{SI_2}{N} - v_L E_2 + (1-y_2)\,(1-z_2)\,\phi_2 I_2 + \gamma\,(1-z_1)\,y_1\phi_1 I_1 + l\alpha r_2 N \\
&\quad - \mu_0 E_2 \\
I_2' &= pqt_2\lambda\frac{SI_2}{N} + v_L E_2 - \phi_2 I_2 - \mu I_2 - \mu_0 I_2 \\
E_3' &= (1-p)\,qt_3\lambda\frac{SI_3}{N} - v_L E_3 + (1-y_2)\,(1-z_3)\,\phi_3 I_3 + (1-\gamma)\,(1-z_1)\,y_1\phi_1 I_1 + l\alpha r_3 N \\
&\quad - \mu_0 E_3 \\
I_3' &= pqt_3\lambda\frac{SI_3}{N} + v_L E_3 - \phi_3 I_3 - \mu I_3 - \mu_0 I_3 \\
E_4' &= (1-p)\,qt_4\lambda\frac{SI_4}{N} - v_L E_4 + y_2\,(1-z_2)\,\phi_2 I_2 + y_2\,(1-z_3)\,\phi_3 I_3 + (1-z_4)\,\phi_4 I_4 \\
&\quad + l\alpha r_4 N - \mu_0 E_4 \\
I_4' &= pqt_4\lambda\frac{SI_4}{N} + v_L E_4 - \phi_4 I_4 - \mu I_4 - \mu_0 I_4 \\
D' &= \mu\,(I_1 + I_2 + I_3 + I_4) \\
N' &= \rho N + \alpha N - \mu\,(I_1 + I_2 + I_3 + I_4) - \mu_0 N
\end{aligned}
\tag{3.1}
$$

Table 3.1: Parameters

Parameter	**Description**	**Range**
a_2	Proportion of initial LTBI cases that are H-resistant	(0, 0.2)
a_3	Proportion of initial LTBI cases that are R-resistant	(0, 0.2)
a_4	Proportion of initial LTBI cases that are MDR	(0, 0.2)
α	Immigration rate	0.00425
b	Of initial active cases that are neither MDR nor H-resistant, the proportion that are drug-susceptible	(0.5, 1)
γ	Proportion of cases of single resistance acquisition where H-resistance is acquired	(0, 1)
l	Proportion of immigrants that have LTBI	(0, 0.3)
λ	Effective contact rate	(0, 50)
μ	TB mortality rate	(0, 0.5)
μ_0	Mortality rate unrelated to TB	0.013
p	Proportion of exogenous infections that are acute	(0, 0.3)
ϕ_1	Rate at which Infectious (Drug-susceptible) individuals end treatment	(0.6, 0.9)
ϕ_2	Rate at which Infectious (H-resistant) individuals end treatment	(0.5, 0.9)
ϕ_3	Rate at which Infectious (R-resistant) individuals end treatment	(0.3, 0.9)
ϕ_4	Rate at which Infectious (MDR) individuals end treatment	(0.3, 0.5)
q	Proportion of active cases that have the potential to be infectious	(0, 1)
r_2	Proportion of immigrant LTBI cases that are H-resistant	(0, 0.2)
r_3	Proportion of immigrant LTBI cases that are R-resistant	(0, 0.2)
r_4	Proportion of immigrant LTBI cases that are MDR	(0, 0.2)
ρ	US birth rate	0.0179
t_1	Proportion of treatment time when individuals are infectious - Drug-susceptible	(0, 0.1)
t_2	Proportion of treatment time when individuals are infectious - H-resistant	(0, 0.1)

t_3	Proportion of treatment time when individuals are infectious - R-resistant	(0, 0.1)
t_4	Proportion of treatment time when individuals are infectious - MDR	(0, 0.1)
v_L	Progression rate from latent to active infection	(0, 0.01)
y_1	Proportion of failed treatments for drug-susceptible TB that result in H- or R-resistance	(0, 1)
y_2	Proportion of failed treatments for H- or R-resistant TB that result in MDR	(0, 1)
z_1	Proportion of treatment courses for Drug-susceptible TB that are successful	(0.6, 0.9)
z_2	Proportion of treatment courses for H-resistant TB that are successful	(0.5, 0.9)
z_3	Proportion of treatment courses for R-resistant TB that are successful	(0.5, 0.9)
z_4	Proportion of treatment courses for MDR TB that are successful	(0.1, 0.8)

Chapter 4

MODEL FITTING

4.1 Algorithm

The randomization algorithm coded in RStudio (see Appendix A for the full code) was used to fit the four-strain model. It takes in epidemiological data from the CDC and acceptable parameter ranges and returns a parameter set that approximately fits the simulation to these data.

To produce a realistic fit, I first assured that the model maintained an accurate total population by setting constant values for ρ (birth rate), μ_0 (death rate unrelated to TB), and α (immigration rate). These are the main parameters that control the differential equation for N', that is, the rate of change of the US population as a whole. These parameters are set to constant, realistic values that produce an accurate total population for the years 2000-2015, as shown in Figure 4.1.

The randomization algorithm fits the model's simulation to four sets of data for the years 2000-2013. These datasets are the total incidence of active cases of

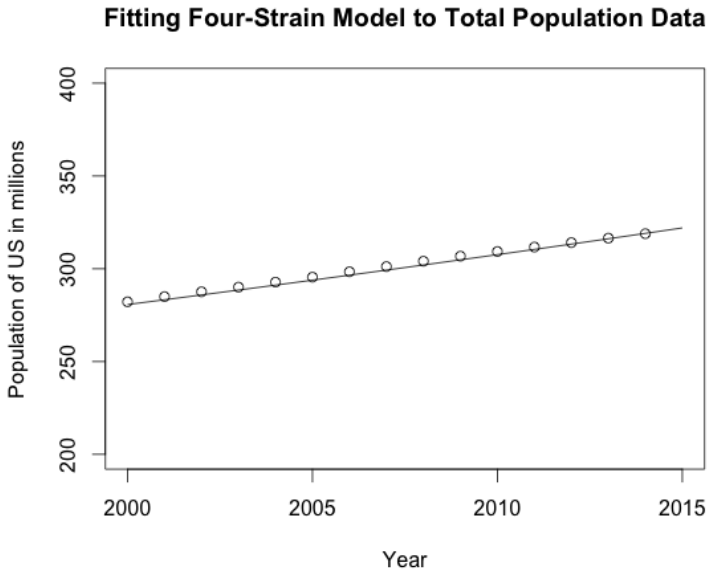

Figure 4.1: Four-strain model fit to census data for total US population, 2000-2015 Open circles represent total population data points [18], while solid lines represent the model's prediction.

TB, active H-resistant cases, active MDR cases, and cumulative TB death in the US. The algorithm generates hundreds of sets of parameters by randomly choosing parameter values within their acceptable ranges.

The fit of the simulation generated by these sets of parameters is quantified using a sum-of-squares method to compare the model's outcome to each of the four datasets from the CDC. As shown in the following equations, for each of the four datasets, the value predicted by the simulation is compared to the actual value from the CDC data at each year. These four difference values are then summed to obtain a measurement of the total fit of the simulation to these datasets.

$$\begin{aligned}
\text{Difference}_j &= \sum_{i=2000}^{2013} \left(\frac{\text{Simulation}_i - \text{CDC}_i}{\text{CDC}_i} \right)^2 \\
&\quad \text{for} j \in \{1, 2, 3, 4\}, \text{ representing the four datasets} \\
\text{Total Difference} &= \sum_{j=1}^{4} \text{Difference}_j
\end{aligned} \tag{4.1}$$

To support my randomization algorithm, I performed control trials to prove that the algorithm was effective. In these trials, I generated a randomized set of parameters and used the simulation produced by these parameters in the years 2000-2013 as the datasets to which the randomization would fit. From this original set of parameters, I generated four datasets corresponding to the CDC data I used in the experimental trials. That is, one value per year for 2000-2013 for total incidence of active cases, H-resistant cases, MDR cases, and cumulative TB death.

Using these as my datasets, I ran the randomization algorithm three times to determine whether it could produce qualitative fit and return the known parameter values. A representative trial is shown in Figure 4.2.

The trials all produced good fit to the original data. However, the three sets of parameter values generated by the randomization algorithm did not match each other, and did not match the original values for many parameters. This shows that fitting qualitatively is not sufficient to ensure accurate parameters, which makes it especially important to establish epidemiologically realistic ranges for parameter values.

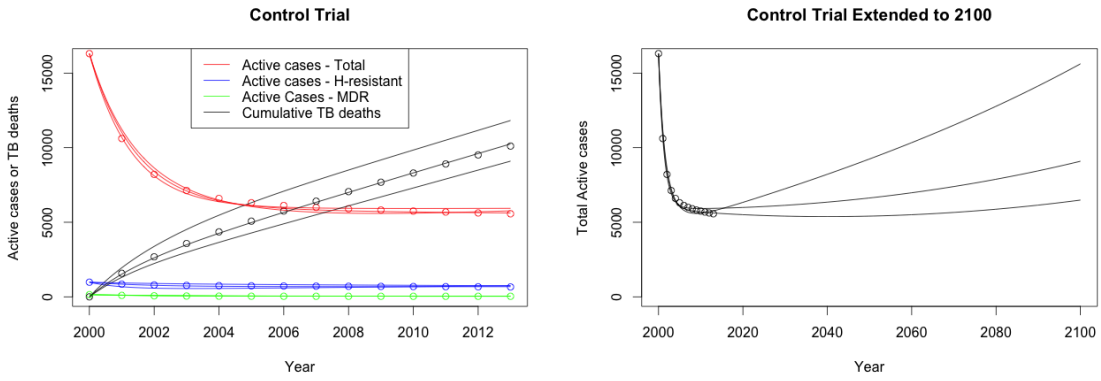

Figure 4.2: Control trial of the randomization algorithm
The fit produced by randomization in four datasets is plotted (left). When the control trials are extended to the year 2100 (right), the total number of active cases is variable. Open circles represent data generated by a random set of parameters. Solid lines represent simulations generated by fitting to these data.

Further, despite the close fit to the data for 2000-2013, the control trials showed notable qualitative variation when their predictions were extended to the year 2100, as shown in Figure 4.2.

While the three trials showed good fit for the interval 2000-2013 to which they were fit, their long-term behavior is quite variable. This reiterates the importance of limiting parameter values to realistic ranges. Since the long-term behavior may vary, when generating simulations for this project, I produced twenty reasonable parameter sets and chose a final simulation based on the consensus of long-term qualitative behavior.

4.2 Parameter Observations

After generating twenty reasonable parameter sets, I compared the values generated for each parameter. For several parameters, an approximately equal value was produced in the majority of trials. These parameters are as follows:

- b, the proportion of unaccounted active cases that are drug-susceptible. This is not a parameter that is used in the differential equations; rather, it is used to set the initial conditions of the simulation. According to data from the CDC [14], there is a known quantity of total active cases in the US, which includes known quantities of H-resistant cases and MDR cases. However, my model also divides total active cases into R-resistant and drug-susceptible cases. Cases reported to the CDC that are neither H-resistant nor MDR may be drug-susceptible or R-resistant, at some proportion b.

- γ, the proportion of cases of acquisition of resistance to a single drug where H-resistance is acquired.

- p, the proportion of exogenous infections that are acute, that is, the proportion of infections that are immediately active.

- ϕ_1, the rate at which Infectious (Drug-susceptible) individuals end treatment for TB.

- ϕ_2, the rate at which Infectious (H-resistant) individuals end treatment for TB.

- ϕ_4, the rate at which Infectious (MDR) individuals end treatment for TB.

- r_2, the proportion of LTBI cases in immigrants that are H-resistant.
- t_3, the proportion of time during the treatment of R-resistant TB during which individuals are actually infectious.
- z_1, the proportion of successful treatments of drug-susceptible TB.
- z_2, the proportion of successful treatments of H-resistant TB.
- z_3, the proportion of successful treatments of R-resistant TB.
- z_4, the proportion of successful treatments of MDR TB.

Note that these parameters are not necessarily the most important in terms of sensitivity analysis (this will be addressed in Section 7.2), but are rather the most consistent over the twenty randomization trials.

Setting these parameters constant to their consensus values and repeating randomization continues to produce good fit, as shown in Figure 4.3, indicating that these values are indeed reasonable.

4.3 Final Parameter Values

The final, consensus set of parameter values that will be used for subsequent analysis are given in Table 4.1. Their fit to CDC data for the years 2000-2013 is shown in Figure 4.4.

These parameters are consistent with epidemiological data or intuition in several respects.

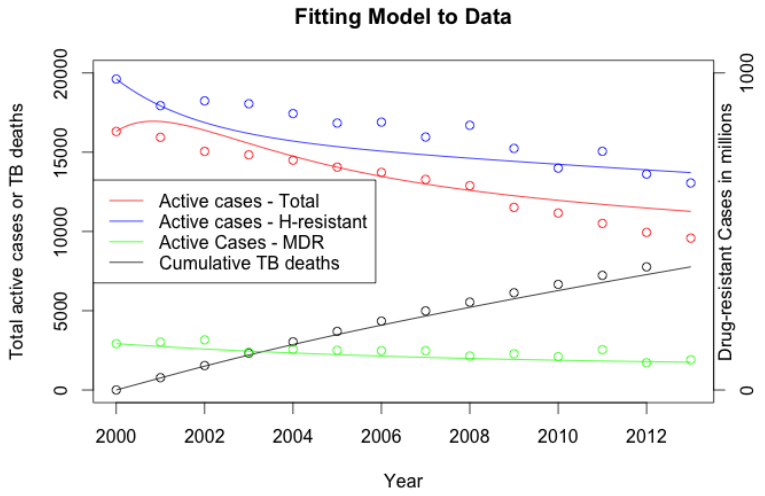

Figure 4.3: Plot of simulation with constant parameters fit to CDC data (open circles) [13]
Note that the right-hand axis applies to drug-resistant cases, which occur at much lower quantities than total active cases and TB death do.

The a parameters represent the proportion of LTBI cases in the initial condition of the model which are H-resistant, R-resistant, and MDR, respectively. Since the presence of LTBI cases and their level of drug-resistance are generally unknown, these values cannot be confirmed. However, it is reasonable that most cases of LTBI are drug-susceptible, which is reflected in the parameter values, since the sum of these terms is much less than 0.5. According to the CDC, there are more active cases of H-resistant TB than MDR TB [14]. It is reasonable to suppose that this trend is reflected in the amounts of LTBI cases. This is true for these parameters, since $a_2 > a_4$.

The constant α represents the rate of immigration into the US, and is calculated directly from immigration statistics [40].

b represents the proportion of initial active cases that are drug-susceptible out

Table 4.1: Parameter Values

Parameter	a_2	a_3	a_4	α	b	γ	l
Value	0.06	0.04	0.005	0.00425	0.95	0.5	0.01
λ	μ	μ_0	p	ϕ_1	ϕ_2	ϕ_3	ϕ_4
30	0.04	0.013	0.3	0.85	0.3	0.3	0.1
q	r_2	r_3	r_4	ρ	t_1	t_2	t_3
0.7	0.06	0.05	0.01	0.0179	0.02	0.01	0.02
t_4	v_L	y_1	y_2	z_1	z_2	z_3	z_4
0.01	0.001	0.5	0.5	0.9	0.9	0.9	0.8

of those cases that are not assigned a compartment by the CDC data. That is, cases that are not described by the CDC as being H-resistant or MDR. These cases may be drug-susceptible at some proportion b, or they may be R-resistant at the proportion $(1 - b)$. Since drug-resistance is still relatively rare, it is reasonable to suppose that most of these remaining cases are drug-susceptible, and only a minority are R-resistant. Therefore the high value of $b = 0.95$ is reasonable.

The constant μ_0 represents the US death rate unrelated to TB. In 2000, the average life expectancy in the US was 76.8 years [39]. The reciprocal of this value is approximately equal to $\mu_0 = 0.013$.

The ϕ parameters represent the rate at which Infectious individuals end treatment for each strain. These are rate parameters, so as treatment time increases, the ϕ parameters will decrease. Treatment of drug-susceptible strains usually lasts approximately six months [24]; this duration gives a parameter of $1 - e^{\frac{-1}{0.5}} = 0.86$, which is consistent with the chosen value of $\phi_1 = 0.85$. Treatment time generally increases with increasing resistance, which is consistent with the parameter values such that $\phi_1 > \phi_2, \phi_3 > \phi_4$.

q represents the proportion of active cases that are able to infect others. Note

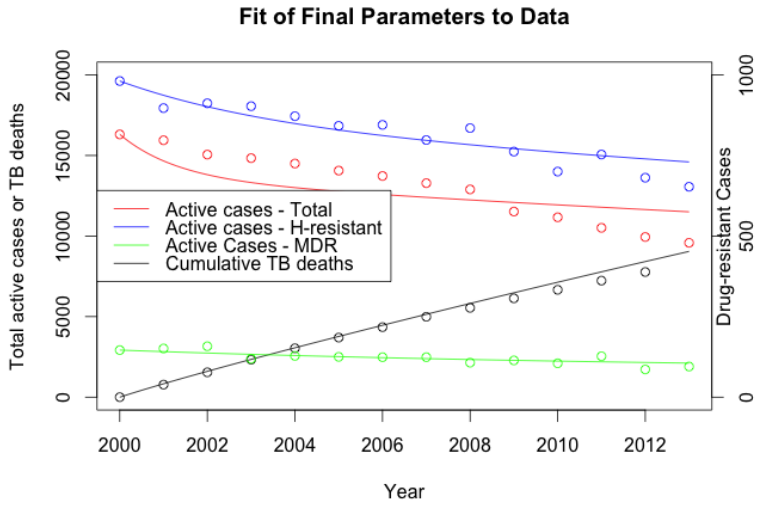

Figure 4.4: Plot of simulation with final parameters fit to CDC data (open circles) [13]
Note that the right-hand axis applies to drug-resistant cases, which occur at much lower quantities than total active cases and TB death do.

that while all active cases are in the Infectious compartment, they are not all infectious; some active cases are noninfectious. This parameter value of $q = 0.7$ is consistent with the Hill model, which used a value of $q = 0.708$ [19].

The r parameters represent the proportion of immigrant LTBI cases that are infected with each strain of TB. As with the a parameters, most LTBI cases should be drug-susceptible, and there should be fewer MDR cases than H-resistant cases. This is consistent with these parameters, since the sum of the r values is much less than 0.5 and $r_2 > r_4$.

The constant ρ represents the US birth rate. This parameter value of $\rho = 0.0179$ is consistent with the Hill model, which used a value of $\rho = 0.018$ [19].

The t parameters represent the proportion of treatment time during which

actively infected individuals are capable of infecting others. The amount of time individuals are infectious remains relatively constant, so the proportion should decrease as total treatment time increases. As treatment time increases, the ϕ rate parameters decrease, and the proportions t should decrease, which is consistent with the parameter values. That is, since $\phi_1 \geq \phi_2, \phi_3 \geq \phi_4$, then $t_1 \geq t_2, t_3 \geq t_4$.

v_L represents the progression rate from latent to active infection. This parameter value of $v_L = 0.001$ is consistent with the Hill model, which used a value of $v_L = 0.0014$ [19].

This epidemiological evidence combined with the qualitative fit of the simulation using these parameters supports these values as accurate.

Chapter 5

PREDICTIONS

5.1 Long-Term Behavior

Figure 5.1 plots the prediction of the simulation through the year 2100 for each of the four strains of active TB.

While the incidence of each category of active TB is expected to decrease, there is a trend towards an increasing proportion of drug-resistant cases. Figure 5.2 illustrates the proportion of all active cases that are expected to be drug-resistant or MDR. These proportions are expected to increase approximately after the year 2020, and continue increasing seemingly indefinitely. Due to the difficulty of treating drug-resistant TB, this predicts great challenges in the future of TB control.

Another important consideration is the possibility of elimination. Elimination of TB is defined as an incidence of fewer than one case per million in population [13]. As shown in Figure 5.3, the simulation does not predict a possibility for elimination before 2100, which is consistent with the predictions made by the Hill model [19].

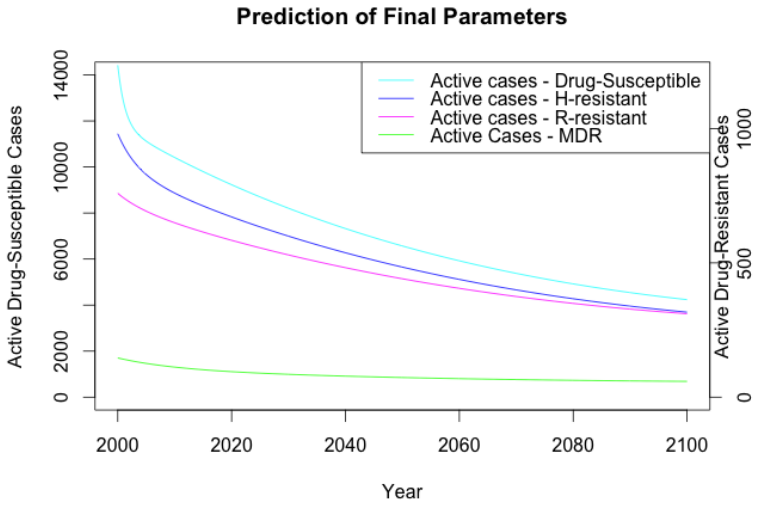

Figure 5.1: Plot of simulation with final parameters through the year 2100
The incidence of each strain of active TB is expected to decrease. Note that the right-hand axis applies to drug-resistant cases, which occur at much lower quantities than drug-susceptible cases do.

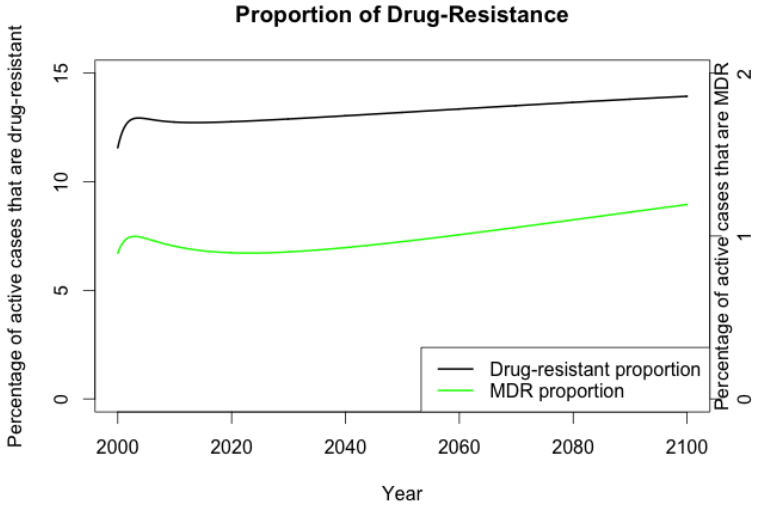

Figure 5.2: Plot of proportion of drug-resistant and MDR active cases through the year 2100
Note that the right-hand axis applies to the proportion of MDR cases, which is much lower than the proportion of all drug-resistant cases.

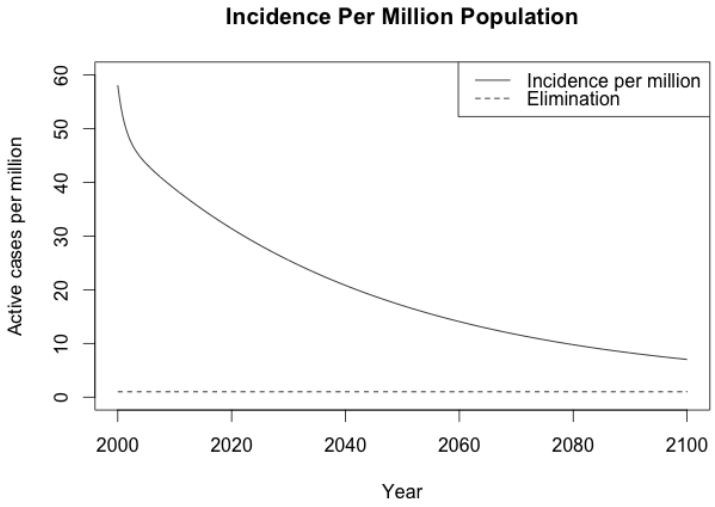

Figure 5.3: Plot of the potential for elimination of TB
The simulation does not predict elimination by 2100.

5.2 Analysis

5.2.1 Immigration

The effects of immigration on the state of TB in the US can be analyzed by comparing the model's simulations with and without immigration. All else being equal, if the parameter α is set to a value of 0, immigration into the US is effectively halted after the year 2000. The consequences of removing immigration from the simulation are illustrated in Figure 5.4.

When immigration is removed, the total incidence of active cases of TB does not show a major reduction until approximately the year 2020. By the year 2100, the total incidence is reduced by approximately 28%. That is, in general terms, 28% of the incidence of cases predicted in the year 2100 by the model are the result of immigration. These removed cases include not only immigrants who were

infected with LTBI upon entering the US, but also immigrants who would have been exogenously infected with TB after arriving in the US and all individuals who would have been exogenously infected with TB by an actively infected immigrant after their arrival.

Removing immigration from the simulation does not show a noticeable effect on the proportion of all cases that are drug-resistant. However, the proportion of cases that are MDR is reduced. This suggests that immigrant populations may contribute more, proportionally, to MDR cases than to other strains.

Perhaps surprisingly, removing immigration increases the incidence per million in population, making the goal of elimination less realistic. While preventing immigration after the year 2000 decreases the rate of TB incidence, it also decreases the total US population. This reduction in total population may explain why the incidence per million increases even as total incidence is reduced.

5.2.2 Reactivation and Exogenous Infection

The new cases predicted by the simulation can be separated into two categories: reactivated cases and exogenous infections. These categories describe the method by which an individual enters an Infectious compartment.

Reactivation cases are all those active cases that are the result of disease progression from the latent stage. Therefore, reactivation cases are represented by movement from an Exposed compartment to an Infectious compartment (See Figure 3.1, *3*).

In contrast, exogenous infections are all those active cases that are the imme-

diate result of infectious contact between an Infectious person and a Susceptible person. Therefore, exogenous cases are represented by movement from the Susceptible compartment to an Infectious compartment (See Figure 3.1, *2*).

These two categories account for all possible entrants into the Infectious compartment. That is, for one of the four Infectious compartments i, where $i \in \{1, 2, 3, 4\}$:

$$\begin{aligned} I_i' &= pqt_i\lambda\frac{SI_i}{N} + v_L E_i - \phi_i I_i - \mu I_4 - \mu_0 I_i \\ \text{Reactivation} &= v_L E_i \\ \text{Exogenous infections} &= pqt_i\lambda\frac{SI_i}{N} \end{aligned} \tag{5.1}$$

Figure 5.5 plots the incidence of active cases in both categories for each of the four strains. Since it is unknown whether the cases that exist at the beginning of the simulation (in the year 2000) are due to exogenous infection or reactivation, all of these categories begin with a value of 0.

By the year 2100, the predictions of the simulation indicate that for all strains, more cases occur as a result of reactivation than exogenous infection. In the year 2100, the ratio of reactivation cases to exogenous infection cases can be computed for each strain. For the drug-susceptible strain, this ratio is approximately $6.1 : 1$. For the H- and R-resistant strains, this ratio is $4.5 : 1$ and $1.8 : 1$, respectively. For the MDR strain, this ratio is $1.4 : 1$.

While reactivation cases outnumber exogenous cases in all strains, drug-resistant

strains (especially R-resistant and MDR strains) show a much lower proportion of reactivation cases to exogenous infections.

This is unexpected given the dynamics of drug-resistant strains. Since cases of acquired resistance enter the Exposed compartments of drug-resistant strains, their Exposed compartments could be expected to be proportionally larger than the Exposed (Drug-susceptible) compartment, which has no equivalent input. In this case, it might be expected that proportionally more reactivation cases would occur in the drug-resistant strains. Since reactivation cases are described by movement from an Exposed compartment to an Infectious compartment, a larger Exposed compartment could be predicted to result in greater potential for reactivation cases. However, the opposite effect is seen: the drug-resistant strains show fewer reactivation cases proportionally than the drug-susceptible strain.

This suggests that exogenous infections are more of a concern for drug-resistant strains than they are for drug-susceptible strains. It may be, therefore, that quarantine for these particular strains would be an effective measure for reducing incidence of the drug-resistant strains. This potential intervention will be explored in more detail in Section 8.6.

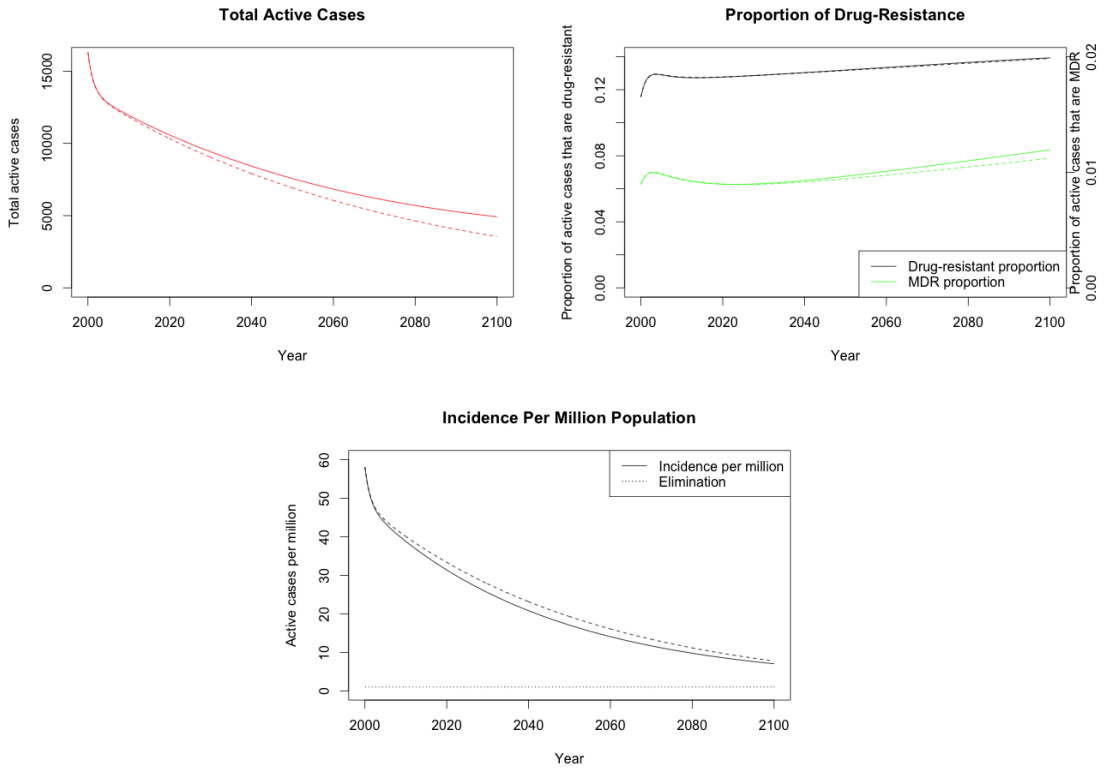

Figure 5.4: Comparison of the model's predictions to simulation without immigration

Dashed lines represent simulations with the immigration rate set to $\alpha = 0$. When immigration is removed, the total incidence of active cases is decreased (top left). The proportion of drug-resistant cases is unaffected, but the proportion of MDR cases is decreased (top right). Incidence per million in population is increased (bottom).

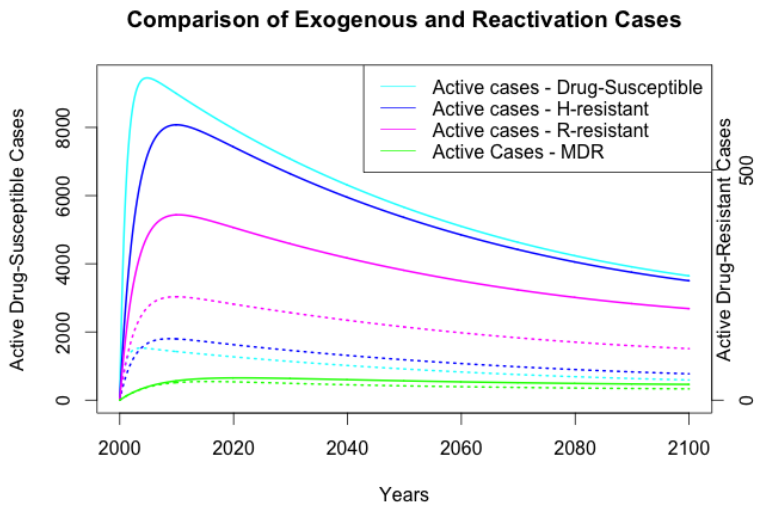

Figure 5.5: Comparison of exogenous and reactivation infections
Solid lines represent incidence of new cases resulting from reactivation. Dotted lines represent incidence of new cases resulting from exogenous infection. Note that the right-hand axis applies to drug-resistant cases, which occur at much lower quantities than drug-susceptible cases do.

Chapter 6

COMPARISON TO HILL

6.1 Extensions of the Hill Model

My four-strain model presented above builds upon the Hill model in several valuable respects.

First, I have included dynamics of drug-resistance. This is an important change, especially regarding parameter values. In models that do not differentiate between drug-susceptible and drug-resistant cases, the increased treatment time and decreased treatment success that is typical of resistant cases cannot be reflected. According to my model, drug-resistant cases are predicted to make up an increasing proportion of all active cases of TB, so these dynamics are of increasing importance.

Second, my model is potentially more accurate to known data collected by the CDC than the Hill model, due to my use of additional data. Hill *et al.* fit their model only to data for total incidence of active cases for 2000-2008 [19], whereas I used four sets of data for the years 2000-2013 (total incidence of TB, incidence of

H-resistant TB, incidence of MDR TB, and cumulative TB death). The Hill model does not include resistance, so the datasets of H-resistant and MDR cases would not have been useful, but their model does include death due to TB, so its prediction can be compared to CDC data for TB death. This plot is shown in Figure 6.1. The TB death rate predicted by the Hill model overestimates these reported TB deaths. In contrast, my model was fit to this dataset, and therefore predicts values that are more accurate to these data.

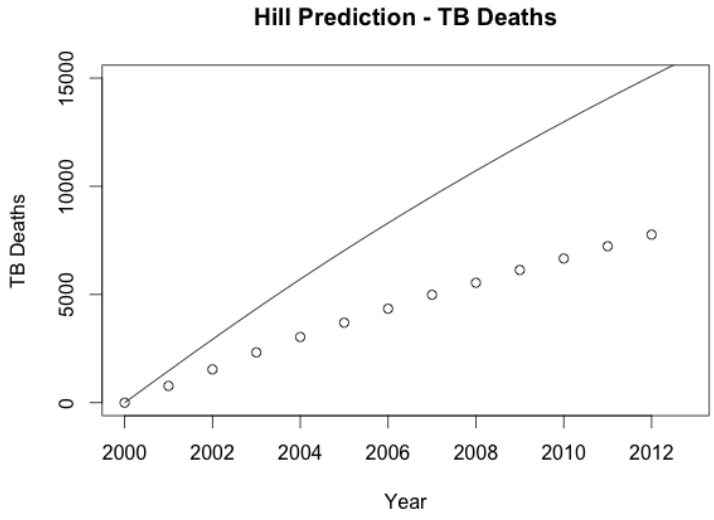

Figure 6.1: Comparison of TB death predicted by the Hill model to CDC data Cumulative deaths from TB disease predicted by the Hill model are shown by the solid line. Open circles represent CDC data for cumulative TB death for the years 2000-2012 [13].

However, it can be argued that the Hill model takes into account TB deaths that go unreported, and that this causes the increased death rate. It is true that TB is more prevalent in at-risk populations that include, among others, undocumented immigrants and the homeless, which may be insufficiently documented by the CDC.

For this reason, it is likely that TB case rates and TB death rates are underreported. However, the Hill model fits accurately to the data on incidence of active cases collected by the CDC; that is, the model does not account for entirely unreported cases. Therefore, the Hill model would be including cases that are reported to the CDC and result in death, but where TB is not recorded as the official cause of death.

Given the discrepancy between the Hill model's prediction and reported TB death, the Hill model suggests that nearly half of all actual TB deaths go unreported. Since such cases are by definition unknown, there cannot be accurate data as to their prevalence. Still, this magnitude of unreported cases seems extreme, and the TB death rate predicted by the Hill model is at least questionable.

Third, I suggest that the Hill model predicts a rate of LTBI in immigrants that is unrealistically high. The parameter the Hill model uses for the proportion of foreign-born arrivals that have LTBI is $f = 0.187$; this value is assumed [19]. In contrast, my parameter for this same proportion is lower by more than an order of magnitude: $l = 0.01$

Indeed, since realistically most cases of LTBI go undetected, the true value cannot be known with certainty, and it may lie somewhere between these two estimates. However, the Hill model's value is potentially too extreme. While it is a widely quoted estimate that one-third of the world's population is infected with LTBI, immigrants to the US are not a random sample of the world's population, and could be expected to show a rate of LTBI much lower than one-third, and even lower than the Hill model's estimate.

One potentially illuminating study was performed by Walter *et al.* on immigrants to California from the Philippines [37]. TB rates are extremely high in the

Philippines; in 2000, Tupasi *et al.* estimated that over 60% of Filipinos have LTBI - nearly twice the worldwide rate [32].

Out of the 123,114 Filipino immigrants included in the study by Walter *et al*, 793 cases of active TB were reported in the years 2001-2010. Of these, the authors estimated that 204 were due to LTBI [37]. Since 5-10% of all LTBI cases progress to active disease, at maximum perhaps 4,080 of these immigrants had undetected LTBI when they entered the US. This would suggest a maximum proportion of immigrants with LTBI of approximately 0.033.

This is much lower than the value predicted by the Hill model, even though it is based on a population already disposed towards high rates of LTBI. Since the base rate of LTBI in the population of the Philippines is twice the rate for the world as a whole, intuitively, the LTBI rate in immigrants from the world population would be expected to be half that of these Filipino immigrants, or 0.0165. This is an interesting result, though it must be noted that this conclusion is uncertain; the available data on LTBI prevalence is too limited for confidence.

If this is true, though, then the Hill model's parameter value of $f = 0.187$ [19] overestimates the true amount of immigrant LTBI cases. It is possible that my parameter value of $l = 0.01$ may underestimate the immigrant LTBI burden, but in light of these data, it does not seem to be unreasonably low.

6.2 Fitting to Hill

Next, I use my randomization algorithm to compare my model to the Hill model's predictions. This will allow me to identify the factors most responsible for the

differences between the two simulations.

First, I removed resistance from the four-strain model. I did this by setting relevant parameters to 0, so that there could be no entry into any of the drug-resistant Exposed or Infectious compartments (effectively returning to the single-strain model). I used the randomization algorithm to produce several sets of parameters. A plot of the resulting simulations is shown in Figure 6.2. From this, it is clear that removing resistance alone is not sufficient for my model to show qualitative agreement with the Hill model. The Hill model predicts an increase in total incidence that is not reflected in most of my model's simulations, and overall the Hill model predicts greater incidence than my model does.

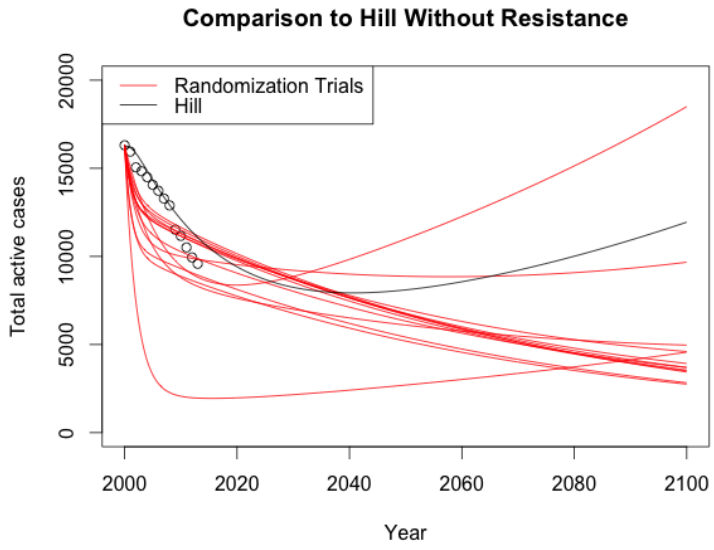

Figure 6.2: Comparison of the Hill model's prediction to my model's prediction, when resistance is removed
Open circles represent CDC data for total active cases [13].

Therefore, I next attempted to match my model to the Hill model by fitting

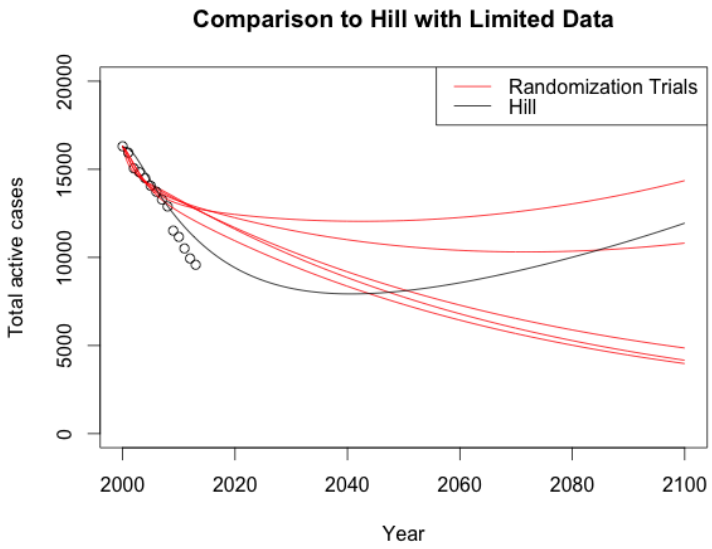

Figure 6.3: Comparison of the Hill model's prediction to my model's prediction, when fit only to the data used to fit the Hill model
Open circles represent CDC data for total incidence of active cases [13].

only to the data to which the Hill model was fit; that is, total incidence of active cases for the years 2000-2008 [19]. To consider the effect of this change alone, drug-resistance was reinstated in these trials. The resulting plot is shown in Figure 6.3. From this, it is clear that limiting the data to which my simulations are fit still does not produce good qualitative fit with the Hill model's predictions.

Since neither of these methods of comparison to the Hill model produced good fit, I combined their approaches. That is, I removed resistance from my model as before, and fit the model only to the data used to fit the Hill model. The resulting plot is shown in Figure 6.4. From this, it is clear that even when these strategies are combined, my model does not produce simulations qualitatively similar to the Hill model's. Further, limiting the model in this way produces simulations that are much

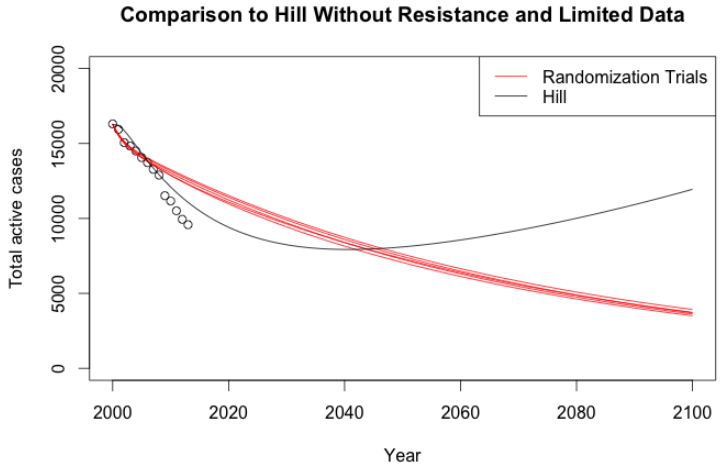

Figure 6.4: Comparison of the Hill model's prediction to my model's prediction, when resistance is removed and fit only to the data used to fit the Hill model Open circles represent CDC data for total incidence of active cases [13].

less variable. Five randomization trials with these limitations all produce extremely similar predictions through the year 2100, indicating that most of the variation in my model's predictions is due to the inclusion of resistance and the use of several datasets.

Finally, to determine whether my model is capable of expressing the qualitative predictions of the Hill model, I fit my model directly to the Hill model. Instead of using CDC data, I generated a dataset using one value of total incidence of active cases per year as predicted by the Hill model from 2000-2100. I then used the randomization algorithm to fit my model to this dataset, which would potentially fit the model directly to the Hill model's prediction. The resulting plot is shown in Figure 6.5. In the short term, most of the trials do not show good qualitative fit with the Hill model. However, in the long term, approximately between the years 2040 and 2100, the randomization trials show fairly good qualitative fit with the Hill

model's prediction. This shows that my model is capable of at least the long-term behavior predicted by the Hill model.

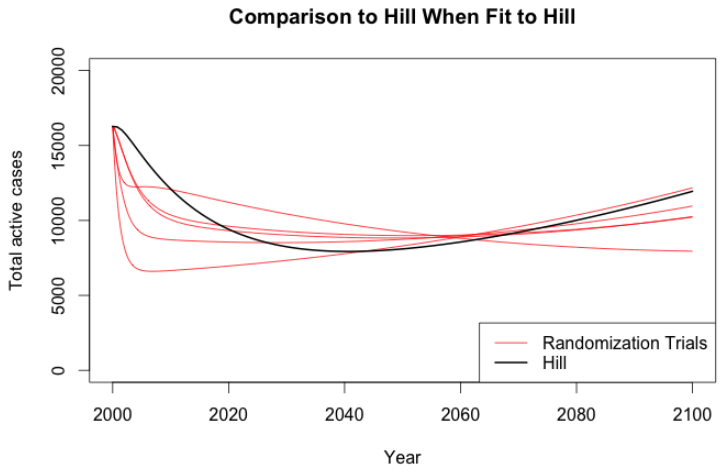

Figure 6.5: Comparison of the Hill model to my model, when fit directly to the Hill model's prediction

Since my model is capable of matching the Hill model's long-term predictions to some extent, I argue that my model's prediction disagrees with the Hill model because it is more accurate, not because my model is limited and unable to express the Hill model's qualitative predictions.

Chapter 7

MATHEMATICAL ANALYSIS

7.1 R_0

The basic reproduction number, R_0, is a measurement of the infectiousness of a disease. It is defined as the average number of new infections produced by one Infectious individual entering an entirely Susceptible population [36]. It is equal to the product of the effective contact rate, the total population, and the duration of Infectiousness [36]. To compute R_0 for this model, I took the average of values for each of the four compartments. Thus for this model, given the final parameter values,

$$R_0 = \frac{q\lambda N}{4} \sum_{i=1}^{4} \frac{t_i}{\mu + \mu_0 + \phi_i} = 254.2233. \tag{7.1}$$

This is a very high value of R_0, seeming to indicate that TB is likely to be in a state of epidemic. However, it is important to note that this figure concerns the total number of infections, not necessarily *active* infections. Very few patients with LTBI will progress to active disease. Under the assumption that 5-10% of LTBI cases progress to active disease, this means that one Infectious patient may result in approximately 13 - 25 new active cases. Additionally, not all of these new cases will themselves be infectious; some proportion of active cases will be non-infectious TB. This is an unusual circumstance that does not usually complicate the use of R_0 to analyze infectious diseases.

This value is still quite high. It can be explained by further considering the challenges facing TB in particular where calculating R_0 is concerned.

R_0 may not be as informative in the case of TB as it is in other infectious diseases. Due to the unusually long latent period of TB when compared to other infectious diseases, the current pattern of incidence depends heavily on transmissions that took place years, or even decades, in the past. Therefore, R_0 does not accurately describe the current incidence of active cases and may not give much meaningful information about the current state of TB epidemiology [36].

The value of R_0 can also be calculated if one considers only MDR TB as a separate disease. Then,

$$R_0 = \frac{q\lambda N t_4}{\mu + \mu_0 + \phi_4} = 385.3103. \tag{7.2}$$

This value is even greater than R_0 for all cases of TB. As with the R_0 value for TB in general, most of these 385 predicted new infections will remain latent and will not progress to active disease. Assuming once more that 5-10% of LTBI cases progress to active disease, this value of R_0 suggests that each Infectious (MDR) patient may result in 19 - 38 new active MDR cases. This again indicates the challenges of applying R_0 values to the analysis of TB [36]. It may also indicate that MDR TB is more infectious than TB in general, but this conclusion is tentative, due to the uncertain nature of this method of analysis.

While these values are higher than would be realistically expected for most infectious diseases, they are not unreasonable for other TB models. Hill *et al.*, for example, used parameters that would result in the following value for R_0, when calculated similarly [19]:

$$R_0 = \frac{\beta N}{2} \sum_{i=0,1} \frac{1}{\mu_d + \mu_i + \phi_i} = 1188.439. \tag{7.3}$$

Other mathematical models of TB give a wide range of high values of R_0. When calculated similarly, Uys *et al.*'s model indicates $R_0 = 5479$, while Cohen and Murray's model indicates $R_0 = 14.1$ [15, 33]. These results continue to indicate that R_0 values are of uncertain use for the study of TB [36].

Therefore, while my model produces a value of R_0 that is unrealistic for infectious diseases, it may not be out of the ordinary for TB modeling.

7.2 Sensitivity Analysis

For a model of this level of complexity, sensitivity analysis is challenging. It would be impossible to sample every region of the parameter space - that is, the complete set of all combinations of all possible values for all the model's parameters - to any level of completeness. Even if I limited each of the 31 parameters to only two values, over two billion trials would be required to sample every possible combination. Therefore, I have limited this project to univariate analysis on the parameters.

For this process, I varied a single parameter value while holding all the other parameters constant at their final values. For the parameter of interest, I examined ten values around its final, consensus value at 10% intervals. That is, a parameter with a final value of x was sampled at $0.5x$, $0.6x$, $0.7x$, $0.8x$, $0.9x$, $1.1x$, $1.2x$, $1.3x$, $1.4x$, and $1.5x$. Note that for many parameters, some of these values fell outside their epidemiologically acceptable range. The summary statistics used to measure the effects of this variation were the total incidence of active cases and the incidence of MDR cases.

The results of this univariate sensitivity analysis are summarized below.

- a_2: No effect on either total incidence or MDR incidence.
- a_3: No effect on either total incidence or MDR incidence.
- a_4: No effect on total incidence. Major effect on MDR incidence for the entire timeframe. Increasing a_4 increases MDR incidence.
- α: Moderate effect on both total incidence and MDR incidence after approximately the year 2040. Increasing α increases total and MDR incidence.

- b: Moderate effect on total incidence only before approximately the year 2020. Increasing b decreases incidence. No effect on MDR incidence.
- γ: No effect on either total incidence or MDR incidence.
- l: Moderate effect on both total incidence and MDR incidence after approximately the year 2040. Increasing l increases total and MDR incidence.
- λ: Major effect on both total incidence and MDR incidence for the entire timeframe. Increasing λ increases total and MDR incidence.
- μ: No effect on total incidence. Moderate effect on MDR incidence for the entire timeframe. Increasing μ decreases MDR incidence.
- μ_0: Major effect on both total incidence and MDR incidence after approximately the year 2005. Increasing μ_0 decreases total and MDR incidence.
- p: Major effect on both total incidence and MDR incidence for the entire timeframe. Increasing p increases total and MDR incidence.
- ϕ_1: Major effect on total incidence for the entire timeframe. Increasing ϕ_1 decreases total incidence. No effect on MDR incidence.
- ϕ_2: No effect on either total incidence or MDR incidence.
- ϕ_3: No effect on either total incidence or MDR incidence.
- ϕ_4: No effect on total incidence. Major effect on MDR incidence for the entire timeframe. Increasing ϕ_4 decreases MDR incidence.

- q: Major effect on both total incidence and MDR incidence for the entire timeframe. Increasing q increases total and MDR incidence.
- r_2: No effect on either total incidence or MDR incidence.
- r_3: No effect on either total incidence or MDR incidence.
- r_4: No effect on total incidence. Moderate effect on MDR incidence after approximately the year 2040. Increasing r_4 increases MDR incidence.
- ρ: Moderate effect on both total incidence and MDR incidence after approximately the year 2050. Increasing ρ increases total and MDR incidence.
- t_1: Major effect on total incidence for the entire timeframe. Increasing t_1 increases total incidence. No effect on MDR incidence.
- t_2: No effect on either total incidence or MDR incidence.
- t_3: No effect on either total incidence or MDR incidence.
- t_4: No effect on total incidence. Major effect on MDR incidence for the entire timeframe. Increasing t_4 increases MDR incidence.
- v_L: Major effect on both total incidence and MDR incidence for the entire timeframe. Increasing v_L increases total and MDR incidence.
- y_1: No effect on either total incidence or MDR incidence.
- y_2: No effect on either total incidence or MDR incidence.
- z_1: No effect on either total incidence or MDR incidence.

- z_2: No effect on total incidence. Moderate effect on MDR incidence after approximately the year 2020. Increasing z_2 decreases MDR incidence.
- z_3: No effect on total incidence. Moderate effect on MDR incidence after approximately the year 2020. Increasing z_3 decreases MDR incidence.
- z_4: No effect on either total incidence or MDR incidence.

According to this univariate analysis, the parameters that have the greatest effect on the total incidence of TB are λ, μ_0, p, ϕ_1, q, t_1, and v_L. The parameters with the greatest effect on MDR incidence are a_4, λ, μ_0, p, ϕ_4, q, t_4, and v_L.

I will return to some of these significant parameters in Chapter 8 when they will be affected by potential interventions.

Chapter 8

POTENTIAL INTERVENTIONS

8.1 LTBI Testing - l

One potential intervention that could be practiced is the implementation of testing and treatment for LTBI in the immigrant population upon entry to the US. Since most cases of TB occur in the foreign-born population, it may be that preventing incoming cases could reduce the incidence of active cases in the US.

To test this, I compared the original prediction to several simulations in which the parameter l, which controls the proportion of immigrants that have LTBI, was decreased. The parameter was progressively decreased from an original value of $l = 0.01$ to $l = 0.009, 0.008, 0.007, 0.006, 0.005, 0.004, 0.003, 0.002, 0.001, 0$.

The results of these simulations are shown in Figure 8.1.

This intervention seems to be effective at reducing the incidence of total active cases as well as the incidence of drug-resistant cases.

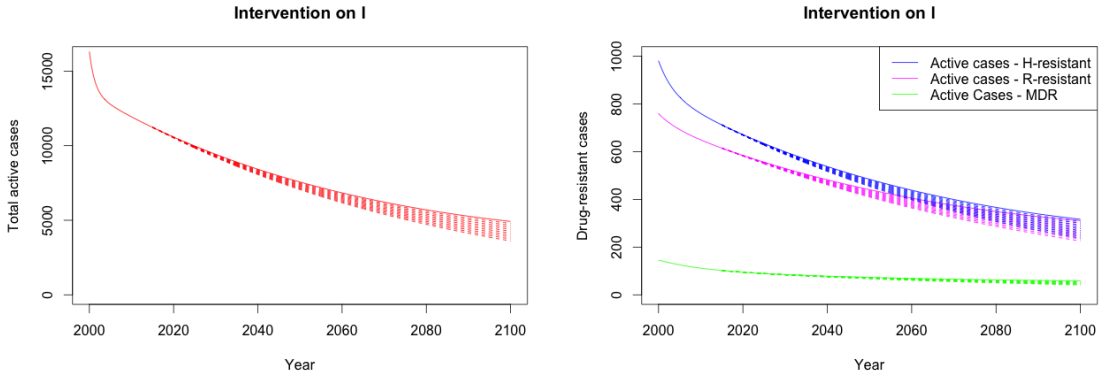

Figure 8.1: Effects of intervention on proportion of immigrants with LTBI
Solid lines represent the original simulation; dashed lines represent trials with interventions starting at the year 2015. Decreasing l decreases the total incidence of active cases (left) and the incidence of drug-resistant cases, including MDR cases (right).

8.2 Treatment Length - ϕ

The next potential intervention concerns the amount of time required to treat active cases of TB. Increasing the parameters ϕ_1, ϕ_2, ϕ_3, and ϕ_4 increases the rate at which individuals leave their respective Infectious compartment. Increasing these rates describes a decrease in treatment time. This decreases the amount of time during which Infectious individuals are able to infect others, so the potential for new exogenous infections should be reduced. Preventing some exogenous infections should result in a decrease in incidence.

Recall that ϕ_1 and ϕ_4 were found to be significant parameters in Section 7.2 affecting total incidence and MDR incidence respectively; this predicts that this intervention will be effective.

To test this, all ϕ parameters were simultaneously increased.

ϕ_1 was progressively increased from an original value of $\phi_1 = 0.85$ to $\phi_1 =$

$0.935, 1.02,\ 1.105, 1.19, 1.275, 1.36, 1.445, 1.53, 1.615, 1.7$, with a maximum of double its original value.

ϕ_2 and ϕ_3 were progressively increased from their original value of $\phi_2 = \phi_3 = 0.3$ to $\phi_2 = \phi_3 = 0.33, 0.36, 0.39, 0.42, 0.45, 0.48, 0.51, 0.54, 0.57, 0.6$, with a maximum of double their original value.

ϕ_4 was progressively increased from an original value of $\phi_4 = 0.1$ to $\phi_4 = 0.11, 0.12, 0.13,\ 0.14, 0.15, 0.16, 0.17, 0.18, 0.19, 0.2$, with a maximum of double its original value.

The results of these simulations are shown in Figure 8.2.

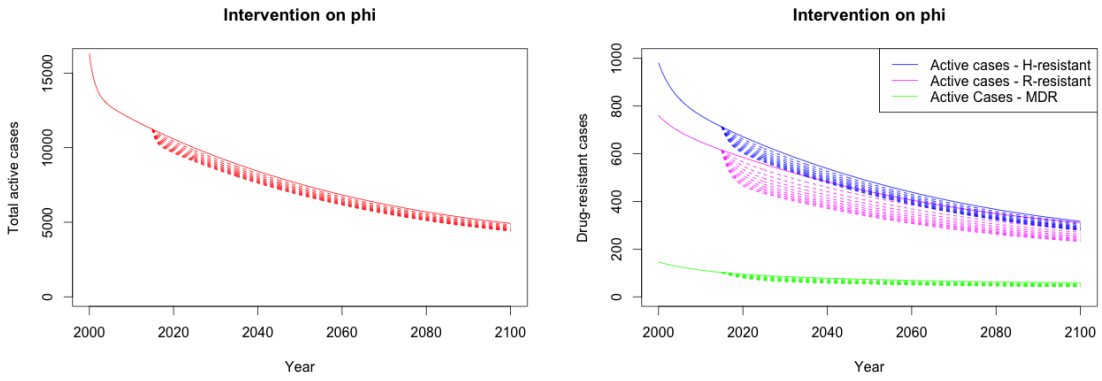

Figure 8.2: Effects of intervention on treatment length for active TB
Solid lines represent the original simulation; dashed lines represent trials with interventions starting at the year 2015. Increasing the four ϕ parameters, to a maximum of doubling their initial value, decreases both the incidence of active cases (left) and incidence of drug-resistant cases, including MDR cases (right).

Intervening on the ϕ parameters does seem to be effective. Further analysis is necessary to determine whether one or all of the parameters are most responsible for this effect. This will be discussed in more detail in Section 9.1.2.

8.3 Treatment Success Rate - z

Next, I modeled interventions on the rate of successful treatment. Increasing the parameters z_1, z_2, z_3, and z_4 increases the proportion of treatments that are successful. If more treatments are successful, fewer overall cases should occur, as individuals will be cured and return to the Susceptible compartment, rather than relapsing to the Exposed compartments. Additionally, fewer drug-resistant cases should occur, since new cases of drug-resistance result from unsuccessful treatment. Increasing the success rate should diminish the opportunity for new cases of resistance to be generated.

To test this, all z parameters were simultaneously increased, to a maximum value of 1.

z_1, z_2, and z_3 were progressively increased from their original value of $z_1 = z_2 = z_3 = 0.9$ to $z_1 = z_2 = z_3 = 0.91, 0.92, 0.93, 0.94, 0.95, 0.96, 0.97, 0.98, 0.99, 1$.

z_4 was progressively increased from an original value of $z_4 = 0.8$ to $z_4 = 0.82, 0.84, 0.86,\ 0.88, 0.9, 0.92, 0.94, 0.96, 0.98, 1$.

The results of these simulations are shown in Figure 8.3.

Intervening on the z parameters does not seem to be effective in reducing the incidence of drug-resistant cases or reducing the incidence of active cases of TB overall.

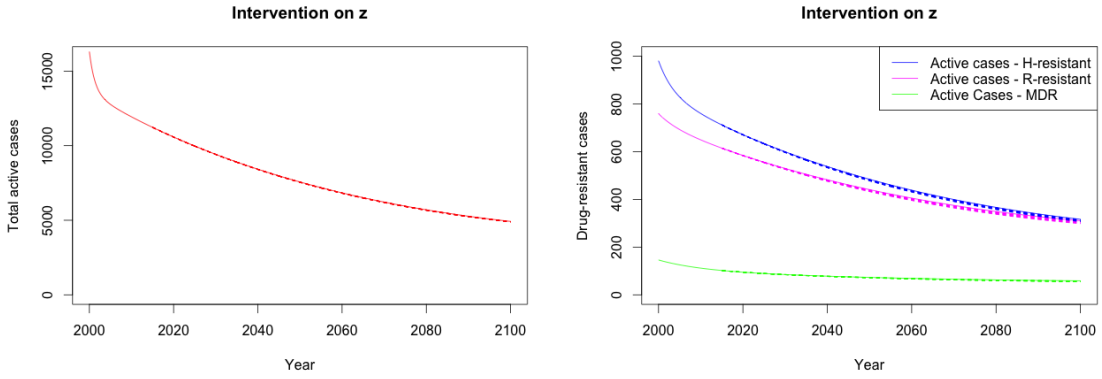

Figure 8.3: Effects of intervention on treatment success rate
Solid lines represent the original simulation; dashed lines represent trials with interventions starting at the year 2015. Increasing the four z parameters, to a maximum value of 1, has only a minor effect on the total incidence of active cases (left) and the incidence of drug-resistant cases (right).

8.4 TB Death Rate - μ

Another way in which the efficacy of treatment can be measured is the death rate due to TB. Decreasing the TB death rate, μ, should be associated with more effective treatment, though this may not necessarily decrease the incidence of active cases.

To test this intervention, I decreased the parameter μ from an original value of $\mu = 0.04$ to $\mu = 0.036, 0.032, 0.028, 0.024, 0.02, 0.016, 0.012, 0.008, 0.004, 0$.

The results of these simulations are shown in Figure 8.4.

Decreasing the death rate μ is not an effective intervention on its own; these simulations show that the incidence of drug-resistant cases actually increases, while the total incidence of active cases is not notably affected. When the death rate is decreased, individuals spend more time on average in the Infectious compartments. This increases the probability that they will infect others. It also increases the probability that individuals will exit the Infectious compartment via treatment failure,

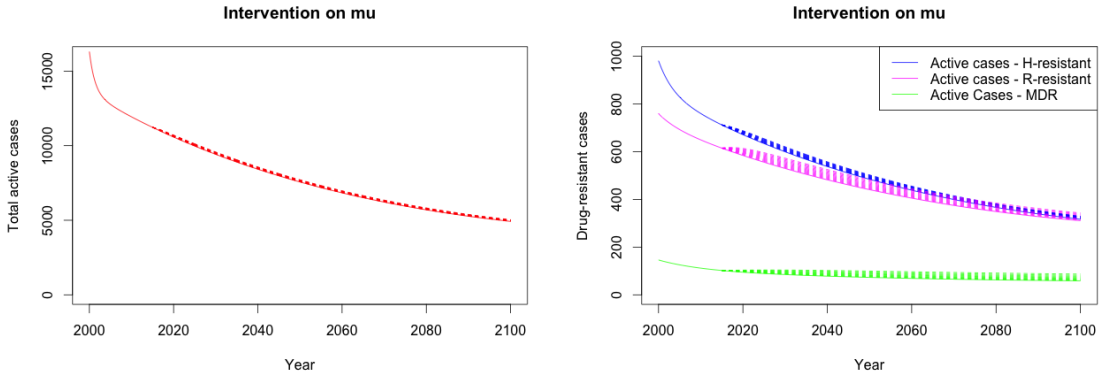

Figure 8.4: Effects of intervention on the death rate from TB
Solid lines represent the original simulation; dashed lines represent trials with interventions starting at the year 2015. Decreasing μ, even to a value of 0, shows no major effect on the total incidence of active cases (left). However, decreasing μ increases the incidence of drug-resistant cases (right).

rather than death, which provides greater opportunity for resistance acquisition. Therefore, a decreased death rate must be paired with other values of treatment efficacy, such as treatment length, in order to be an effective intervention.

8.5 Acquisition of Resistance - y

Another potential intervention is on the rate of acquisition of resistance. Decreasing the parameters y_1 and y_2 decreases the proportion of unsuccessful treatments that will result in drug-resistance. This intervention would be difficult to implement in practice, and would likely have to be paired with other improvements in treatment efficacy. However, it can be modeled mathematically. This intervention is expected to result in decreased incidence of drug-resistant cases.

To test this, both y parameters were simultaneously decreased from their orig-

inal value of $y_1 = y_2 = 0.5$ to $y_1 = y_2 = 0.45, 0.40.35, 0.3, 0.25, 0.2, 0.15, 0.1, 0.05, 0$.

The results of these simulations are shown in Figure 8.5.

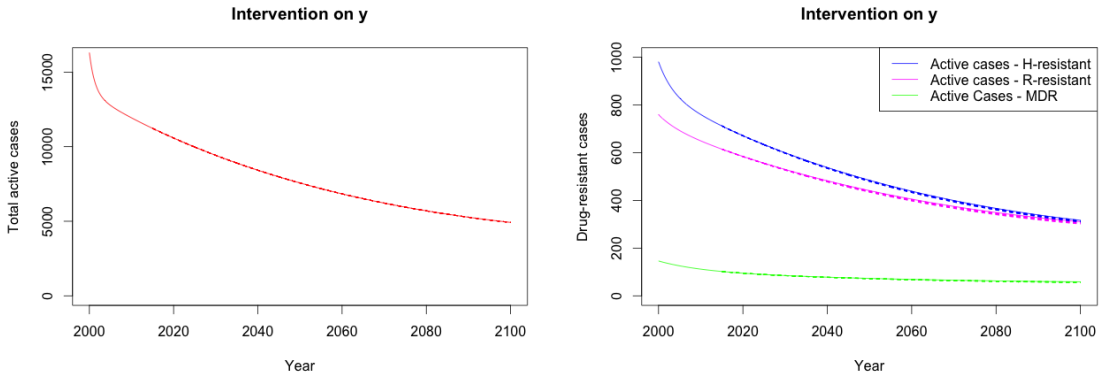

Figure 8.5: Effects of intervention on the proportion of failed treatments that result in resistance

Solid lines represent the original simulation; dashed lines represent trials with interventions starting at the year 2015. Decreasing the two y parameters, even to a value of 0, shows only a minor effect on the total incidence of active cases (left) and the incidence of drug-resistant cases (right).

This intervention does not seem to be effective on the total incidence of active cases or the incidence of drug-resistant cases.

8.6 Quarantine - λ

The final intervention concerns the effective contact rate, λ. This parameter controls the likelihood of contact and resulting exogenous infections between individuals in the Infectious and Susceptible compartments. One potential strategy to lower the effective contact rate (and therefore, the rate of new infections) is to institute quarantine for some or all individuals known to be infected. Quarantining Infectious individuals would reduce their contacts with Susceptible individuals and prevent

exogenous infections from occurring. While such a practice is not officially in effect at this time, it has been suggested for MDR and XDR strains of TB in particular.

Recall that λ was found to be a significant parameter affecting both total incidence and MDR incidence in Section 7.2; this predicts that this intervention will be effective.

To test the effectiveness of this strategy, I first modeled a policy of quarantine that would only be instituted for individuals with active MDR TB. To do so, I modified the value of λ only for individuals in the Infectious (MDR) compartment. For these individuals only, I decreased the parameter λ from an original value of $\lambda = 30$ to $\lambda = 27, 24, 21, 18, 15, 12, 9, 6, 3, 0$.

The results of these simulations are shown in Figure 8.6.

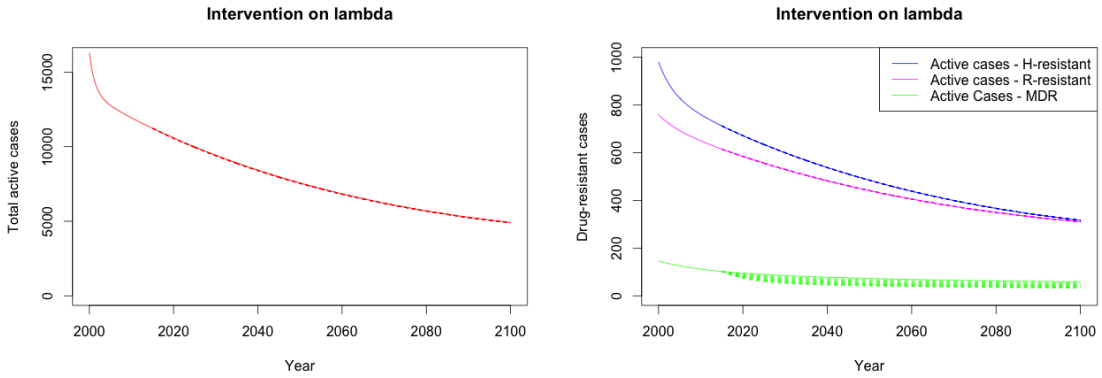

Figure 8.6: Effects of intervention on the effective contact rate for MDR cases only Solid lines represent the original simulation; dashed lines represent trials with interventions starting at the year 2015. Decreasing λ shows little effect on the total incidence of active cases (left), but does decrease the incidence of MDR cases (right).

Reducing the effective contact rate for individuals in the Infectious (MDR) compartment does not affect the total incidence of TB but does decrease incidence

of MDR cases.

Next, I considered a policy of quarantine for all cases of TB. To test this strategy, I decreased the parameter λ as before, but applied the intervention to all four Infectious compartments.

The results of these simulations are shown in Figure 8.7.

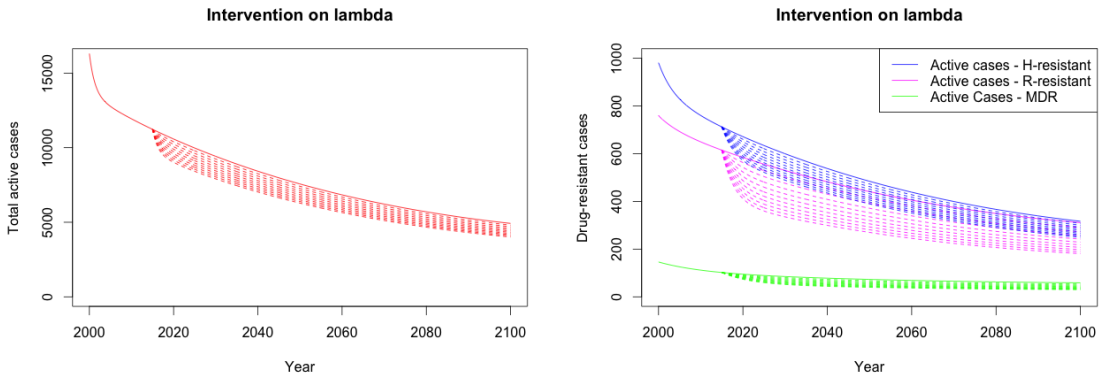

Figure 8.7: Effects of intervention on the effective contact rate for all active cases of TB

Solid lines represent the original simulation; dashed lines represent trials with interventions starting at the year 2015. Decreasing λ decreases the total incidence of active cases (left) and the incidence of drug-resistant cases (right).

Decreasing the effective contact rate by instituting a policy of quarantine for all Infectious individuals predicts a reduction in total incidence of TB as well as reduction in incidence for all drug-resistant strains, including MDR.

Chapter 9

CONCLUSIONS

9.1 Efficacy of Interventions

The goals upon which I will focus in this section are decreasing total TB incidence and decreasing the incidence and proportion of MDR TB. While decreasing incidence of H- and R-resistant cases would also be valuable contributions to public health, the aforementioned priorities are the most pressing.

The three most promising interventions for the goal of decreasing total TB incidence and the incidence of MDR cases are altering l, ϕ, and λ. These interventions will now be explored in more detail.

9.1.1 l

Decreasing l, the proportion of immigrants who have LTBI, decreases the incidence of total cases and drug-resistant cases. For every 10% by which l is decreased, there is a corresponding decrease in the total incidence at the year 2100 of 2.65% and a

decrease in the incidence of MDR cases of 3%.

9.1.2 ϕ

Increasing the four ϕ parameters, which control the rate at which individuals end treatment, decreases the incidence of total cases and drug-resistant cases. The effects of these four parameters can be considered individually.

Increasing ϕ_1 decreases the total incidence of TB, but does not affect the incidence of any of the three drug-resistant strains. Increasing ϕ_1 has a diminishing marginal effect on the total incidence at the year 2100. Initially increasing ϕ_1 by 10% reduces incidence at 2100 by 8.9%. An additional increase of 10% reduces incidence by only 7.4% further. The marginal effect continues to diminish.

Intervening on ϕ_2 and ϕ_3 does not affect the total incidence. Increasing ϕ_2 decreases the incidence of H-resistant cases; increasing ϕ_3 decreases the incidence of R-resistant cases. Neither parameter affects MDR incidence. This is expected, as increasing these parameters lowers the amount of time spent in the relevant Infectious compartment, which decrease the potential for Infectious individuals to infect Susceptible individuals. Preventing these new cases would only have an effect on the strain corresponding to the parameter that was altered, as exogenously infected individuals are infected with the same strain as the individual who infected them.

Intervening on ϕ_4 does not affect the total incidence, but does decrease the incidence of MDR cases. Increasing ϕ_4 has a diminishing marginal effect on MDR incidence at the year 2100. Initially increasing ϕ_4 by 10% reduces incidence at 2100 by 11.2%. An additional increase of 10% reduces incidence by only an additional

8.9%. The marginal effect continues to diminish.

Therefore, the greatest marginal benefit on the incidence of total cases can be achieved by introducing slight decreases in treatment time for drug-susceptible cases, while the greatest marginal benefit on the incidence of MDR cases can be achieved by introducing slight decreases in treatment time for MDR cases.

9.1.3 λ

If quarantine is instituted only for MDR cases, the effective contact rate λ is decreased for individuals in the Infectious (MDR) compartment. This does not affect the total incidence of TB, but does decrease the incidence of MDR cases.

Decreasing λ has a diminishing marginal effect on the incidence of MDR cases at the year 2100. Initially decreasing λ by 10% decreases incidence by 7.3%. An additional decrease of 10% reduces incidence by only 6.3% further. The marginal effect continues to diminish.

I also explore the effects of intervening on λ for all strains. If quarantine is instituted for active cases of all strains of TB, incidence is reduced for total cases and for MDR cases.

For every 10% by which λ is decreased, the total incidence at the year 2100 is reduced by approximately 2%. The diminishing marginal return of altering this parameter is negligible.

Decreasing λ has a diminishing marginal effect on the incidence of MDR cases at the year 2100. Initially decreasing λ by 10% decreases MDR cases by 8.3%. An additional decrease of 10% reduces incidence by only an additional 7.1%. The

marginal effect continues to diminish.

Note that MDR incidence at 2100 is more affected by quarantine in all cases than by quarantine that is limited to MDR cases.

It is important to note that in the US, treatment for TB includes conditions similar to unofficial quarantine. For example, in some dangerous cases of TB when the patient does not comply with doctors' safety recommendations, they may be involuntarily confined to a hospital to reduce their risk of infecting others [26]. Therefore, reducing λ by small amounts may not be an intervention at all; it may actually be more accurate to current practice by accounting for these few cases of confinement.

9.2 Possibility of Elimination

Elimination is defined as incidence of fewer than one case per million in population [13]. Even if the interventions described above were implemented, elimination would not be feasible before 2100.

In Figure 9.1, a simulation is plotted in which the interventions are maximized by setting l and λ to 0 and setting the ϕ parameters to double their original values. This means that no immigrants are infected with LTBI, no exogenous infections occur, and Infectious individuals finish treatment at twice the current rate. Even with these interventions, by the year 2100 there is still an incidence of 4.1 cases per million in population.

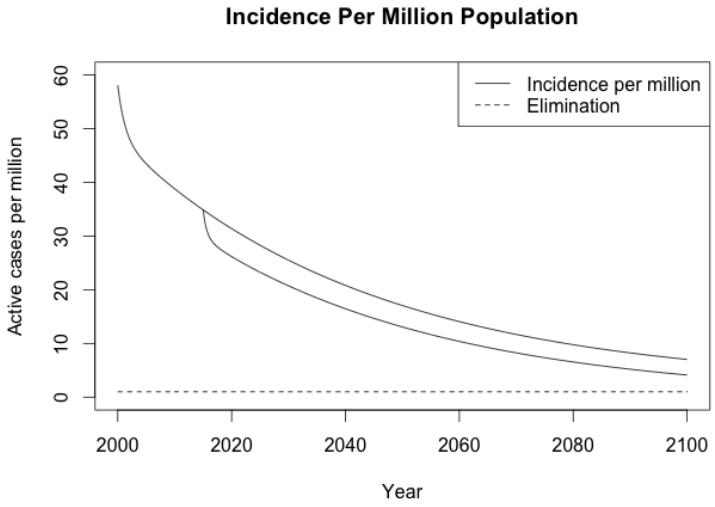

Figure 9.1: Plot of the potential for elimination of TB with interventions maximized after 2015
The simulation still does not predict elimination by 2100.

9.3 Recommendations

To decrease the total incidence of active cases of TB, the most effective intervention is to decrease treatment time for drug-susceptible TB. Instituting quarantine to some extent on all strains of TB and requiring testing and treatment for LTBI in incoming immigrants would also decrease total incidence.

To decrease the incidence of active MDR cases, the most effective intervention is to decrease treatment time for MDR TB. Instituting quarantine on all strains of TB and treating immigrant LTBI would also decrease MDR TB incidence.

To decrease the proportion of all cases that are MDR, the most effective intervention is to decrease treatment time for MDR TB. Under current conditions, MDR cases are expected to make up 1.6% of all active cases of TB in the US by the year

2100. This proportion decreases to less than 1% only after a major increase in ϕ_4 of 50%, but is noticeably decreased with any improvement in MDR TB treatment time, as seen in Figure 9.2.

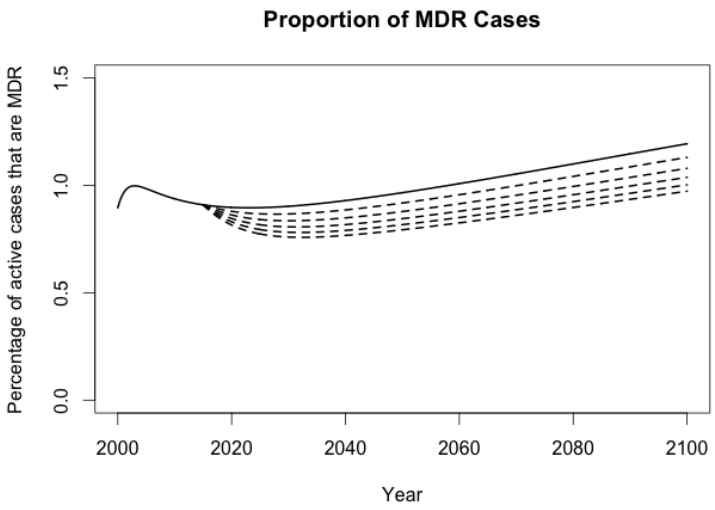

Figure 9.2: Plot of the proportion of MDR cases through the year 2100 when the parameter ϕ_4 is increased
The solid line represents the original simulation; dashed lines represent trials with interventions at 10% intervals, up to an increase of 50%.

These recommendations are based off of mathematical predictions, which are much easier to model than to execute. I recognize that in reality, these interventions could be prohibitively difficult to implement.

Decreasing treatment time for TB may not be practical, since long treatment times are necessary to ensure full eradication of *M. tuberculosis* bacteria. Attempting to speed up the rate of treatment could result in more failed treatments, which would only worsen the problems of TB, especially in drug-resistant strains.

While quarantine may be effective at preventing exogenous infections and the

spread of disease, it would be a highly controversial policy. Confining a larger portion of TB patients during their treatment would be difficult in practice, both for logistical and ethical reasons. Hospitals may not have appropriate space to quarantine TB patients for the long duration of their treatment. This long duration would also make patients less willing to consent to quarantine, which raises ethical concerns regarding the confinement of patients against their will.

Finally, testing for LTBI in incoming immigrants is also a difficult prospect. With the large volume of immigrants entering the US, it would be a major use of resources to test thoroughly for LTBI and treat all those who test positively. For efficiency, such testing could be limited to immigrants from countries with high endemic rates of TB, but this has the potential to encourage discrimination.

Overall, it is clear that while these recommendations seem mathematically effective, it is important to recognize that implementing these policies would not be straightforward.

9.4 Future Work

There are some aspects of the model that, with additional study, could be improved.

The values of the z parameters, which control the proportion of treatments that are successful, could be altered for additional accuracy. In my model, I use the values $z_1 = z_2 = z_3 = 0.9$ and $z_4 = 0.8$, which were generated by the randomization algorithm. While the model as a whole still fits to the data, these values could be inaccurate. One estimate of international treatment success rates indicates that they may be as low as 80% for drug-susceptible strains, 72% for single-drug-resistant

strains, and 15% for MDR strains [7]. The success rates in the US may be greater than in the world as a whole, but this still suggests that my parameter values may be unrealistically high.

These values may also explain why intervening on the z parameters had no beneficial effect on incidence of TB or MDR TB in Section 8.3. These values are already so high that increasing them, even to a value of 1, makes no noticeable difference. More realistic parameter values might show the effects of improvement in success rate more strongly.

Further, the values for the z parameters are suspect because they do not decrease with increasing resistance. Generally, as drug-resistance increases, the cure rates for TB strains decrease considerably [22]. This should dictate that $z_1 > z_2, z_3 > z_4$, but this is not the case in my model, since $z_1 = z_2 = z_3$. Therefore, more examination of the treatment success parameters would be beneficial.

Relatedly, as drug-resistance increases, the survival rates for TB strains may also decrease [22]. Instead of reflecting this, my model uses a single parameter value for the TB death rate, μ. A potential improvement could be to use four different parameters, reflecting the tendency of the death rate to increase for resistant strains of TB.

The final parameters that could be improved are the ϕ parameters, which control the rate at which patients end treatment. In my model, I use the values $\phi_1 = 0.85$, $\phi_2 = \phi_3 = 0.3$, and $\phi_4 = 0.1$, which were generated by the randomization algorithm. The parameter ϕ_1 indicates a treatment length of 0.52 years, which is consistent with the treatment of drug-susceptible strains, which usually lasts approximately six months [24]. However, the parameters ϕ_2 and ϕ_3 indicate

a treatment length of 2.8 years, which is much higher than would be realistic. ϕ_4 suggests that treatment for MDR TB lasts 9.5 years, which is even more extreme. While treatment for MDR TB does last longer that treatment for drug-susceptible TB, current recommendations suggest treatment times of only 20 months, or 1.7 years [24].

Since the ϕ_4 parameter was found to have a significant effect on MDR incidence in Section 7.2, using a more accurate value could have major effects on the model's prediction. Therefore, despite the fit of my model to data overall, the ϕ parameters for drug-resistant strains are unrealistically low. More work on these parameters is necessary to improve their accuracy while maintaining fit.

Potentially, the inaccuracy in the ϕ and z parameters could mediate one another to some extent. That is, if the model indicates that TB treatment is completing too slowly, but at a higher rate of success, the total number of successes might be closer to an accurate figure than it would be if only one of these parameters were inaccurate. However, to make such a claim with any certainty, more parameter analysis is necessary.

With continued study, the model could also be improved by taking other factors into account. For example, my model includes no information about HIV co-infection, which can be an important influence on TB disease dynamics. Individuals with LTBI are more likely to progress to active disease if they are also infected with HIV [12]. Individuals with HIV also have greater rates of death from active TB disease [12]. Therefore, it could be valuable to include specific compartments and rates for these individuals in order to increase accuracy and the information that can be gained from model.

Additionally, my model could be improved with more information about current TB epidemiology. One important weakness of the model is the uncertainty of the number and type of LTBI cases present in the US. LTBI is asymptomatic, so these individuals usually do not realize that they are infected. Therefore, cases of LTBI are rarely reported. Further, the BCG vaccine causes false positive results even in uninfected individuals, further complicated what little LTBI data there is. Since these data are important for setting accurate initial conditions to my model, their uncertainty is problematic. If I were able to find or extrapolate more detailed data to which to fit my model, its accuracy to reality could be improved.

Finally, an interesting extension of this study would be to consider the economic implications of the suggested interventions. While I have studied the efficacy of these interventions mathematically, their practicality could be measured in one respect by introducing cost to the model. Then the cost effectiveness of these interventions could be contemplated. This examination would increase the practical value of these results, as well as further help to guide potential policy.

Appendix A

Randomization Algorithm Code

```
library(deSolve)
deltaT < 0.1 #The length of each time step (in years)
finalYr < 13 #In years   We only have data from the CDC through
#2013
totT < finalYr/deltaT #Time steps
cutoffYr < 8/deltaT
loops < 100 #Number of times a new parameter value will be chosen
data < data.frame(matrix(NA, nrow=loops, ncol=2)) #This empty
#matrix will later contain parameter and difference values
names(data) < c("Parameter", "Difference")
CDCActiveTotal < c(16308, 15945, 15055, 14835, 14499, 14061,
                   13727, 13282, 12893, 11519, 11164, 10509,
                   9940, 9582) #CDC Reported Tuberculosis 2013
#Table 2
ActiveCasesTotal < CDCActiveTotal/1000000 #Calculates total
#active infections from the CDC data
CDCTBDeaths < c(0, 776, 776+764, 776+764+784, 776+764+784+711,
                776+764+784+711+662, 776+764+784+711+662+648,
                776+764+784+711+662+648+644,
                776+764+784+711+662+648+644+554,
                776+764+784+711+662+648+644+554+590,
                776+764+784+711+662+648+644+554+590+529,
```

```
                776+764+784+711+662+648+644+554+590+529+569,
                776+764+784+711+662+648+644+554+590+529+569+536)
#CDC Reported Tuberculosis 2013   Table 1
TotalDeaths < CDCTBDeaths/1000000
CDCActiveHR < c(981, 897, 912, 903, 872, 842, 845, 798, 835,
                762, 700, 753, 681, 653) #CDC Reported
#Tuberculosis 2013   Table 8
ActiveCasesHR < CDCActiveHR/1000000
CDCActiveMDR < c(146, 151, 158, 119, 128, 125, 124, 124, 107,
                 114, 105, 127, 86, 95) #CDC Reported
#Tuberculosis 2013   Table 9
ActiveCasesMDR < CDCActiveMDR/1000000
OldDiff < NewDiff < 0 #Establishes a starting point for the
#difference values
count < 0 #Will count the number of times the algorithm repeats

parameters < vector(mode="numeric", length=27) #Creates a vector
#of parameters with the values set to 0 to start
names(parameters) < c("a2", "a3", "a4", "b", "gamma", "l", "mu",
                      "p", "phi1", "phi2", "phi3", "phi4", "ql",
                      "r2", "r3", "r4", "t1", "t2", "t3", "t4",
                      "vL", "y1", "y2", "z1", "z2", "z3", "z4")

ranges < data.frame(matrix(c(0, 0.2, 0, 0.2, 0, 0.2, 0.5, 1, 0,
                             1, 0, 0.3, 0, 0.5, 0, 0.3, 0.6, 0.9,
                             0.5, 0.9, 0.3, 0.9, 0.3, 0.5, 0, 30,
                             0, 0.2, 0, 0.2, 0, 0.2, 0, 0.1, 0,
                             0.1, 0, 0.1, 0, 0.1, 0, 0.01, 0, 1,
                             0, 1, 0.6, 0.9, 0.5, 0.9, 0.5, 0.9,
                             0.1, 0.8),nrow=2,
                           ncol=length(parameters)))
#Creates a matrix of the minima and maxima for all parameters
names(ranges) < c("a2","a3","a4","b","gamma", "l", "mu", "p",
                  "phi1","phi2","phi3","phi4", "ql","r2","r3",
                  "r4", "t1","t2","t3","t4", "vL","y1","y2","z1",
                  "z2", "z3","z4")
rownames(ranges) < c("min","max")

OldTracking < data.frame(matrix(NA,nrow=length(parameters),
```

```
                                        ncol=3))
#This matrix will contain the parameter and difference values for
#a particular trial
names(OldTracking) < c("Parameter","Value","Difference")

ro = 0.0179
mu0 = 0.013
alpha = 0.00425 #These three parameters are always constant
#They control the overall population size

for (i in 1:length(parameters)){
  parameters[i] < runif(1, ranges[1,i], ranges[2,i])
  #This generates a random value for each parameter, within that
  #parameter's range
}

order = sample(1:length(parameters), length(parameters),
               replace=F)
#This generates a random order in which to fit the parameters

#Functions

Pdot < function(t, v, parms){
  with(as.list(c(parms, v)), {
    #Differential Equations
    dS < ro*N + (1 l)*alpha*N + z1*phi1*I1 + z2*phi2*I2 +
      z3*phi3*I3 + z4*phi4*I4   ql*t1*S*I1/N   ql*t2*S*I2/N
      ql*t3*S*I3/N   ql*t4*S*I4/N   mu0*S
    dE1 < l*alpha*(1 r2 r3 r4)*N + (1 p)*ql*t1*S*I1/N +
      (1 y1)*(1 z1)*phi1*I1   vL*E1   mu0*E1
    dI1 < p*ql*t1*S*I1/N + vL*E1   phi1*I1   (mu0 + mu)*I1
    dE2 < l*alpha*r2*N + (1 p)*ql*t2*S*I2/N +
      gamma*(1 z1)*y1*phi1*I1 + (1 y2)*(1 z2)*phi2*I2   vL*E2
      mu0*E2
    dI2 < p*ql*t2*S*I2/N + vL*E2   phi2*I2   (mu0 + mu)*I2
    dE3 < l*alpha*r3*N + (1 p)*ql*t3*S*I3/N +
      (1 gamma)*(1 z1)*y1*phi1*I1 + (1 y2)*(1 z3)*phi3*I3
      vL*E3   mu0*E3
    dI3 < p*ql*t3*S*I3/N + vL*E3   phi3*I3   (mu0+mu)*I3
```

```
    dE4 < l*alpha*r4*N + (1 p)*ql*t4*S*I4/N + y2*(1 z2)*phi2*I2 +
      y2*(1 z3)*phi3*I3 + (1 z4)*phi4*I4   vL*I4   mu0*E4
    dI4 < p*ql*t4*S*I4/N + vL*E4   phi4*I4   (mu0+mu)*I4
    dN < (ro+alpha)*N   mu0*N   mu*(I1 + I2 + I3 + I4)
    dD < mu*(I1 + I2 + I3 + I4)
    return ( list(c(dS, dE1, dI1, dE2, dI2, dE3, dI3, dE4, dI4,
                    dD, dN)) )
  })
}

hill < function(initial=cutoffYr+1, final=totT+1, dataSet=P) {
  # recursive=TRUE collapses dataframe to labeled vector
  initv < c(dataSet[initial,], recursive=TRUE)
  # times = data points to be calculated
  times < initial:final*deltaT
  # compute master results
  mres < lsoda(initv, times, Pdot, parms)
  # mres[, 1] = mres without 1st column
  dataSet[initial:final,] < c(mres[, 1])
  return(dataSet)
}

generateResults < function(mres) {
  with(as.list(parms), {
    Susceptible < mres$S
    Exposed1 < mres$E1
    Infectious1 < mres$I1
    Exposed2 < mres$E2
    Infectious2 < mres$I2
    Exposed3 < mres$E3
    Infectious3 < mres$I3
    Exposed4 < mres$E4
    Infectious4 < mres$I4
    Dead < mres$D
    Total < mres$S + mres$E1 + mres$I1 + mres$E2 + mres$I2 +
      mres$E3 + mres$I3 + mres$E4 + mres$I4
    InfectiousTotal < mres$I1 + mres$I2 + mres$I3 + mres$I4
    return(data.frame(Susceptible, Exposed1, Infectious1,
                      Exposed2, Infectious2, Exposed3,
```

```
                        Infectious3, Exposed4, Infectious4, Total,
                        Dead, InfectiousTotal))
  })
}

#Initial randomization loop

for (i in 1:length(parameters)){
  #This runs the randomization code once for each parameter
  for (j in 1:loops){ #For each parameter, the code will generate
    #100 random values and quantify their fit

    parameters[order[i]] < runif(1, ranges[1,order[i]],
                                 ranges[2,order[i]])
    #This generates a random value for the parameter of interest,
    #dictated by the "order" vector

    a2 < as.numeric(parameters[1])
    a3 < as.numeric(parameters[2])
    a4 < as.numeric(parameters[3])
    b < as.numeric(parameters[4])
    gamma < as.numeric(parameters[5])
    l < as.numeric(parameters[6])
    mu < as.numeric(parameters[7])
    p < as.numeric(parameters[8])
    phi1 < as.numeric(parameters[9])
    phi2 < as.numeric(parameters[10])
    phi3 < as.numeric(parameters[11])
    phi4 < as.numeric(parameters[12])
    ql < as.numeric(parameters[13])
    r2 < as.numeric(parameters[14])
    r3 < as.numeric(parameters[15])
    r4 < as.numeric(parameters[16])
    t1 < as.numeric(parameters[17])
    t2 < as.numeric(parameters[18])
    t3 < as.numeric(parameters[19])
    t4 < as.numeric(parameters[20])
    vL < as.numeric(parameters[21])
    y1 < as.numeric(parameters[22])
```

```
y2 <- as.numeric(parameters[23])
z1 <- as.numeric(parameters[24])
z2 <- as.numeric(parameters[25])
z3 <- as.numeric(parameters[26])
z4 <- as.numeric(parameters[27])
#These lines set the parameter values so that they are usable
#as numeric characters

S <- E1 <- I1 <- E2 <- I2 <- E3 <- I3 <- E4 <- I4 <- D <-
  N <- rep(0,totT) #Sets compartment values to 0
P <- data.frame(S, E1, I1, E2, I2, E3, I3, E4, I4, D, N)
#Creates a matrix of compartment values

#Total Population
P$N[1] <- 280.726081 #From census data
#LTBI
P$E1[1] <- 11.213*(1 a2 a3 a4) #Data from Hill
P$E2[1] <- 11.213*a2
P$E3[1] <- 11.213*a3
P$E4[1] <- 11.213*a4
#Active TB
P$I1[1] <- (b*(CDCActiveTotal[1]   CDCActiveHR[1]
                  CDCActiveMDR[1])/(mu0 + mu + phi1))/1000000
#Method from Hill; The CDC tracks H Resistant and MDR cases.
#Those leftover are either drug susceptible or R resistant
#(scaled by b and (b 1), respectively)
P$I2[1] <- (CDCActiveHR[1]/(mu0 + mu + phi2))/1000000
P$I3[1] <- ((1 b)*(CDCActiveTotal[1]   CDCActiveHR[1]
                      CDCActiveMDR[1])/(mu0 + mu + phi3))/1000000
P$I4[1] <- (CDCActiveMDR[1]/(mu0 + mu + phi4))/1000000
#Susceptible Population
P$S[1] <- P$N[1]   P$E1[1]   P$I1[1]   P$E2[1]   P$I2[1]
  P$E3[1]   P$I3[1]   P$E4[1]   P$I4[1]

parms <- c(ro=ro, mu0=mu0, alpha=alpha, a2=a2, a3=a3, a4=a4,
           b=b, gamma=gamma, l=l, mu=mu, p=p, phi1=phi1,
           phi2=phi2, phi3=phi3, phi4=phi4, ql=ql, r2=r2,
           r3=r3, r4=r4, t1=t1, t2=t2, t3=t3, t4=t4, vL=vL,
           y1=y1, y2=y2, z1=z1, z2=z2, z3=z3, z4=z4)
```

```
#This 'parms' vector is redundant, but needed for some of the
#ODE functions

yrs <- seq(2000, 2000+finalYr, deltaT)

P <- hill(1, totT+1)
Results <- generateResults(P)

percdiff1 <- 0
#percdiff will be used to quantify how well the model fits to
#the actual data by taking the percent difference between the
#data point for that year and the model value for that year
#and add it to our total measure of difference
for (k in 1:14) { #For 14 years (2000 2013)
  percdiff1 <- percdiff1 +
    (100*(((Results[(1/deltaT)*(k-1)+1,3])*(mu0+mu+phi1) +
             (Results[(1/deltaT)*(k-1)+1,5])*(mu0+mu+phi2) +
             (Results[(1/deltaT)*(k-1)+1,7])*(mu0+mu+phi3) +
             (Results[(1/deltaT)*(k-1)+1,9])*(mu0+mu+phi4)
             ActiveCasesTotal[k])/ActiveCasesTotal[k]))^2
  #Total active cases: The sum of the compartments I1, I2, I3,
  #and I4
}

percdiff2 <- 0
for (k in 1:14) {
  percdiff2 <- percdiff2 +
    (100*((Results[(1/deltaT)*(k-1)+1,5]*(mu0+mu+phi2)
             ActiveCasesHR[k])/ActiveCasesHR[k]))^2
  #H resistant cases
}

percdiff3 <- 0
for (k in 1:14) {
  percdiff3 <- percdiff3 +
    (100*((Results[(1/deltaT)*(k-1)+1,9]*(mu0+mu+phi4)
             ActiveCasesMDR[k])/ActiveCasesMDR[k]))^2
  #MDR cases
}
```

```
    percdiff4 < 0
    for (k in 2:13) { #For 12 years (2000 2012)
      percdiff4 < percdiff4 +
        (100*((Results[(1/deltaT)*(k 1)+1,11]
                TotalDeaths[k])/TotalDeaths[k]))^2 #TB deaths
    }

    data[j,1] < parameters[order[i]] #The first column contains
    #the parameter value
    data[j,2] < percdiff1^2 + percdiff2 + percdiff3 + percdiff4
    #The second column contains the total difference between the
    #model and the data. Since the H resistant and MDR data depend
    #on total active cases, this value is most important and
    #therefore squared

  } #The 100 rounds of generating random parameter values now end

  m < min(data$Difference) #This finds the minimum difference value
  parameters[order[i]] < data[data[,2]==m, 1]
  #This changes the parameter of interest to the value that
  #produced the minimum difference value
  OldTracking[i,1] < names(parameters)[order[i]]
  OldTracking[i,2] < parameters[order[i]]
  OldTracking[i,3] < OldDiff < m
  #These lines fill in the appropriate row of the tracking matrix
  #with the name, value, and difference value of the parameter of
  #interest.

} #This ends the initial round of randomization; a new value has
#been generated for each parameter

NewTracking < OldTracking
NewDiff < OldDiff
OldDiff < OldDiff+1 #This makes sure the subsequent while loop is
#active


while(OldDiff    NewDiff > 0){ #As long as the fit is being
```

```
#improved, the algorithm continues to run.
OldDiff < NewDiff #The values that were newly generated in the
#previous loop are now old
OldTracking < NewTracking
count < count + 1 #Keeps track of how many times the algorithm
#cycles through

for (i in 1:length(parameters)){ #This runs the randomization
  #code once for each parameter
  for (j in 1:loops){ #For each parameter, the code will
    #generate 100 random values and quantify their fit

    parameters[order[i]] < runif(1, ranges[1,order[i]],
                                 ranges[2,order[i]])
    #This generates a random value for the parameter of
    #interest, dictated by the "order" vector

    a2 < as.numeric(parameters[1])
    a3 < as.numeric(parameters[2])
    a4 < as.numeric(parameters[3])
    b < as.numeric(parameters[4])
    gamma < as.numeric(parameters[5])
    l < as.numeric(parameters[6])
    mu < as.numeric(parameters[7])
    p < as.numeric(parameters[8])
    phi1 < as.numeric(parameters[9])
    phi2 < as.numeric(parameters[10])
    phi3 < as.numeric(parameters[11])
    phi4 < as.numeric(parameters[12])
    ql < as.numeric(parameters[13])
    r2 < as.numeric(parameters[14])
    r3 < as.numeric(parameters[15])
    r4 < as.numeric(parameters[16])
    t1 < as.numeric(parameters[17])
    t2 < as.numeric(parameters[18])
    t3 < as.numeric(parameters[19])
    t4 < as.numeric(parameters[20])
    vL < as.numeric(parameters[21])
    y1 < as.numeric(parameters[22])
```

```
y2 < as.numeric(parameters[23])
z1 < as.numeric(parameters[24])
z2 < as.numeric(parameters[25])
z3 < as.numeric(parameters[26])
z4 < as.numeric(parameters[27])
#These lines set the parameter values so that they are
#usable as numeric characters

S < E1 < I1 < E2 < I2 < E3 < I3 < E4 < I4 <
  D < N < rep(0,totT) #Sets compartment values to 0
P < data.frame(S, E1, I1, E2, I2, E3, I3, E4, I4, D, N)
#Creates a matrix of compartment values

#Total Population
P$N[1] < 280.726081 #From census data
#LTBI
P$E1[1] < 11.213*(1 a2 a3 a4) #Data from Hill
P$E2[1] < 11.213*a2
P$E3[1] < 11.213*a3
P$E4[1] < 11.213*a4
#Active TB
P$I1[1] < (b*(CDCActiveTotal[1]    CDCActiveHR[1]
                  CDCActiveMDR[1])/(mu0 + mu + phi1))/1000000
#Method from Hill; The CDC tracks H Resistant and MDR cases.
#Those leftover are either drug susceptible or R resistant
#(scaled by b and (b 1), respectively)
P$I2[1] < (CDCActiveHR[1]/(mu0 + mu + phi2))/1000000
P$I3[1] < ((1 b)*(CDCActiveTotal[1]    CDCActiveHR[1]
                      CDCActiveMDR[1])/
             (mu0 + mu + phi3))/1000000
P$I4[1] < (CDCActiveMDR[1]/(mu0 + mu + phi4))/1000000
#Susceptible Population
P$S[1] < P$N[1]    P$E1[1]    P$I1[1]    P$E2[1]    P$I2[1]
  P$E3[1]    P$I3[1]    P$E4[1]    P$I4[1]

parms < c(ro=ro, mu0=mu0, alpha=alpha, a2=a2, a3=a3, a4=a4,
          b=b, gamma=gamma, l=l, mu=mu, p=p, phi1=phi1,
          phi2=phi2, phi3=phi3, phi4=phi4, ql=ql, r2=r2,
          r3=r3, r4=r4, t1=t1, t2=t2, t3=t3, t4=t4, vL=vL,
```

```
            y1=y1, y2=y2, z1=z1, z2=z2, z3=z3, z4=z4)
#This 'parms' vector is redundant, but needed for some of
#the ODE functions

yrs <- seq(2000, 2000+finalYr, deltaT)

P <- hill(1, totT+1)
Results <- generateResults(P)

percdiff1 <- 0
#percdiff will be used to quantify how well the model fits
#to the actual data by taking the percent difference between
#the data point for that year and the model value for that
#year and add it to our total measure of difference
for (k in 1:14) { #For 14 years (2000 2013)
  percdiff1 <- percdiff1 +
    (100*(((Results[(1/deltaT)*(k-1)+1,3])*(mu0+mu+phi1) +
             (Results[(1/deltaT)*(k-1)+1,5])*(mu0+mu+phi2) +
             (Results[(1/deltaT)*(k-1)+1,7])*(mu0+mu+phi3) +
             (Results[(1/deltaT)*(k-1)+1,9])*(mu0+mu+phi4)
             ActiveCasesTotal[k])/ActiveCasesTotal[k]))^2
  #Total active cases: The sum of the compartments I1, I2,
  #I3, and I4
}

percdiff2 <- 0
for (k in 1:14) {
  percdiff2 <- percdiff2 +
    (100*((Results[(1/deltaT)*(k-1)+1,5]*(mu0+mu+phi2)
             ActiveCasesHR[k])/ActiveCasesHR[k]))^2
  #H resistant cases
}

percdiff3 <- 0
for (k in 1:14) {
  percdiff3 <- percdiff3 +
    (100*((Results[(1/deltaT)*(k-1)+1,9]*(mu0+mu+phi4)
             ActiveCasesMDR[k])/ActiveCasesMDR[k]))^2
  #MDR cases
```

```
    }

    percdiff4 < 0
    for (k in 2:13) { #For 12 years (2000 2012)
      percdiff4 < percdiff4 +
        (100*((Results[(1/deltaT)*(k 1)+1,11]
                TotalDeaths[k])/TotalDeaths[k]))^2 #TB deaths
    }

    data[j,1] < parameters[order[i]] #The first column contains
    #the parameter value
    data[j,2] < percdiff1^2 + percdiff2 + percdiff3 + percdiff4
    #The second column contains the total difference between the
    #model and the data. Since the H resistant and MDR data
    #depend on total active cases, this value is most important
    #and therefore squared

  } #The 100 rounds of generating random parameter values now end

  m < min(data$Difference) #This finds the minimum difference value
  parameters[order[i]] < data[data[,2]==m, 1]
  #This changes the parameter of interest to the value that
  #produced the minimum difference value
  NewTracking[i,1] < names(parameters)[order[i]]
  NewTracking[i,2] < parameters[order[i]]
  NewTracking[i,3] < m
  #These lines fill in the appropriate row of the tracking
  #matrix with the name, value, and difference value of the
  #parameter of interest.

}#A new, more accurate value has been generated for each parameter

NewDiff=NewTracking[length(parameters),3]
#This sets the NewDiff to the final difference value generated
#by these parameters

}

a2 < OldTracking[OldTracking[,1]=="a2",2]
```

```
a3 < OldTracking[OldTracking[,1]=="a3",2]
a4 < OldTracking[OldTracking[,1]=="a4",2]
b < OldTracking[OldTracking[,1]=="b",2]
gamma < OldTracking[OldTracking[,1]=="gamma",2]
l < OldTracking[OldTracking[,1]=="l",2]
mu < OldTracking[OldTracking[,1]=="mu",2]
p < OldTracking[OldTracking[,1]=="p",2]
phi1 < OldTracking[OldTracking[,1]=="phi1",2]
phi2 < OldTracking[OldTracking[,1]=="phi2",2]
phi3 < OldTracking[OldTracking[,1]=="phi3",2]
phi4 < OldTracking[OldTracking[,1]=="phi4",2]
ql < OldTracking[OldTracking[,1]=="ql",2]
r2 < OldTracking[OldTracking[,1]=="r2",2]
r3 < OldTracking[OldTracking[,1]=="r3",2]
r4 < OldTracking[OldTracking[,1]=="r4",2]
t1 < OldTracking[OldTracking[,1]=="t1",2]
t2 < OldTracking[OldTracking[,1]=="t2",2]
t3 < OldTracking[OldTracking[,1]=="t3",2]
t4 < OldTracking[OldTracking[,1]=="t4",2]
vL < OldTracking[OldTracking[,1]=="vL",2]
y1 < OldTracking[OldTracking[,1]=="y1",2]
y2 < OldTracking[OldTracking[,1]=="y2",2]
z1 < OldTracking[OldTracking[,1]=="z1",2]
z2 < OldTracking[OldTracking[,1]=="z2",2]
z3 < OldTracking[OldTracking[,1]=="z3",2]
z4 < OldTracking[OldTracking[,1]=="z4",2]
#The while loop ended because the "NewDiff" value did not improve,
#so the best parameters are found in the OldTracking matrix, so we
#set to those values

#Now, we run the model one final time to plot the simulation
#generated by these parameters

S < E1 < I1 < E2 < I2 < E3 < I3 < E4 < I4 < D < N <
  rep(0,totT) #Sets compartment values to 0
P < data.frame(S, E1, I1, E2, I2, E3, I3, E4, I4, D, N)
#Creates a matrix of compartment values

#Total Population
```

```
P$N[1] < 280.726081 #From census data
#LTBI
P$E1[1] < 11.213*(1 a2 a3 a4) #Data from Hill
P$E2[1] < 11.213*a2
P$E3[1] < 11.213*a3
P$E4[1] < 11.213*a4
#Active TB
P$I1[1] < (b*(CDCActiveTotal[1]  CDCActiveHR[1]
                CDCActiveMDR[1])/(mu0 + mu + phi1))/1000000
#Method from Hill; The CDC tracks H Resistant and MDR cases.
#Those leftover are either drug susceptible or R resistant
#(scaled by b and (b 1), respectively)
P$I2[1] < (CDCActiveHR[1]/(mu0 + mu + phi2))/1000000
P$I3[1] < ((1 b)*(CDCActiveTotal[1]  CDCActiveHR[1]
                  CDCActiveMDR[1])/(mu0 + mu + phi3))/1000000
P$I4[1] < (CDCActiveMDR[1]/(mu0 + mu + phi4))/1000000
#Susceptible Population
P$S[1] < P$N[1]  P$E1[1]  P$I1[1]  P$E2[1]  P$I2[1]
  P$E3[1]  P$I3[1]  P$E4[1]  P$I4[1]

parms < c(ro=ro, mu0=mu0, alpha=alpha, a2=a2, a3=a3, a4=a4, b=b,
          gamma=gamma, l=l, mu=mu, p=p, phi1=phi1, phi2=phi2,
          phi3=phi3, phi4=phi4, ql=ql, r2=r2, r3=r3, r4=r4, t1=t1,
          t2=t2, t3=t3, t4=t4, vL=vL, y1=y1, y2=y2, z1=z1, z2=z2,
          z3=z3, z4=z4)
#This 'parms' vector is redundant, but needed for some of the ODE
#functions

yrs < seq(2000, 2000+finalYr, deltaT)

P < hill(1, totT+1)
Results < generateResults(P)

#Plot model and CDC data for total active cases, HR cases, MDR
#cases, and TB deaths on the same plot, using two sets of axes
years = 2000:2013 #For CDC data where we have through 2013
plot(yrs, (Results$Infectious1*(mu0+mu+phi1)) +
       (Results$Infectious2*(mu0+mu+phi2)) +
       (Results$Infectious3*(mu0+mu+phi3)) +
```

```
        (Results$Infectious4*(mu0+mu+phi4)),
      main='Fitting Model to Data', xlab='Year',
      ylab='Total active  cases or TB deaths in millions',
      type='l',col='red', ylim=
        c(0,.02)) #Total active cases in the model is the sum of
#each I compartment
points(years,ActiveCasesTotal,col='red')
years=2000:2012 #For CDC data on TB deaths, where we have through
#2012 only
lines(yrs, Results$Dead, col='black')
points(years, TotalDeaths,col='black')
par(new = TRUE) #This uses the right hand side axis, since HR and
#MDR cases exist on a much smaller scale
years = 2000:2013
plot(yrs, Results$Infectious2*(mu0+mu+phi2), axes = FALSE,
      bty = "n", xlab = "", ylab = "", col='blue', ylim=c(0,0.001),
      type='l')
points(years,ActiveCasesHR, col='blue')
lines(yrs,Results$Infectious4*(mu0+mu+phi4),col='green')
points(years,ActiveCasesMDR,col='green')
mtext("Drug resistant Cases in millions",side=4)
axis(side=4,at=c(0,0.001))
legend('left', legend=c('Active cases  Total',
                        'Active cases  H resistant',
                        'Active Cases  MDR',
                        'Cumulative TB deaths'),
       col=c('red', 'blue','green', 'black'), lty=c(1,1, 1, 1))

OldTracking #Prints the parameter and difference values used to
#generate this plot
count #Prints how many times the randomization was repeated
```

Bibliography

[1] ATS. Targeted tuberculin testing and treatment of latent tuberculosis infection. *American Journal of Respiratory and Critical Care Medicine*, 161:S221–S247, 2000.

[2] S. Basu, J.R. Andrews, E.M. Poolman, N.R. Gandhi, N.S. Shah, A. Moll, P. Moodley, A.P. Galvani, and G.H. Friedland. Prevention of nosocomial transmission of extensively drug-resistant tuberculosis in rural south african district hospitals: an epidemiological modelling study. *The Lancet*, 370:1500–1507, October 2007.

[3] C.P. Bhunu. Mathematical analysis of a three-strain tuberculosis transmission model. *Applied Mathematical Modelling*, 35:4647–4660, 2011.

[4] P. Brewin, A. Jones, M. Kelley, M. McDonald, E. Beasley, P. Sturdy, G. Bothamley, and C. Griffiths. Is screening for tuberculosis acceptable to immigrants? a qualitative study. *Journal of Public Health*, 28(3):253–260, July 2006.

[5] Kevin P. Cain, Stephen R. Benoit, Carla A. Winston, and William R. Mac Kenzie. Tuberculosis among foreign-born persons in the united states. *The Journal of the American Medical Association*, 300(4):405–412, July 2008.

[6] K.P. Cain, C.A. Haley, L.R. Armstrong, K.N. Garman, C.D. Wells, M.F. Iademarco, K.G. Castro, and K.F. Laserson. Tuberculosis among foreign-born persons in the united states: Achieving tuberculosis elimination. *American Journal of Respiratory and Critical Care Medicine*, 175:75–79, 2007.

[7] J.A. Caminero, editor. *Guidelines for Clinical and Operational Management of Drug-Resistant Tuberculosis*. International Union Against Tuberculosis and Lung Disease (The Union), 2003.

[8] CDC. Essential components of a tuberculosis prevention and control program; and screening for tuberculosis and tuberculosis infection in high-risk popula-

tions; recommendations of the advisory council for the elimination of tuberculosis. *Morbidity and Mortality Weekly Report*, 44(RR-11), 1995.

[9] CDC. Tb elimination: Multidrug-resistant tuberculosis (mdr tb). Atlanta, GA, May 2012.

[10] CDC. Tuberculosis (tb) facts: Tb can be treated. Atlanta, GA, 2012.

[11] CDC. Latent tuberculosis infection: A guide for primary health care providers. Atlanta, GA, 2013.

[12] CDC. Tb elimination: Extensively drug-resistant tuberculosis (xdr tb). Atlanta, GA, January 2013.

[13] CDC. Reported tuberculosis in the united states, 2013. Atlanta, GA, October 2014.

[14] CDC. Trends in tuberculosis, 2014. Atlanta, GA, September 2015.

[15] Ted Cohen and Megan Murray. Modeling epidemics of multidrug-resistant m. tuberculosis of heterogeneous fitness. *Nature Medicine*, 10(10):1117–1121, October 2004.

[16] H.L. David. Probability distribution of drug-resistant mutants in unselected populations of mycobacterium tuberculosis. *Applied Microbiology*, 20(5):810–814, November 1970.

[17] Keertan Dheda, Tawanda Gumbo, Neel R. Gandhi, Megan Murray, Grant Theron, Zarir Udwadia, G.B. Migliori, and Robin Warren. Global control of tuberculosis: from extensively drug-resistant to untreatable tuberculosis. *Lancet Respiratory Medicine*, 2:321–338, 2014.

[18] Trading Economics. United states population, 2016.

[19] A.N. Hill, J.E. Becerra, and K.G. Castro. Modelling tuberculosis trends in the usa. *Epidemiological Infections*, 2012 2012.

[20] Yee Hyung Kim, Gee Young Suh, Man Pyo Chung, Hojoong Kim, O. Jung Kwon, Seong Yong Lim, Si Young Lim, and Won-Jung Koh. Treatment of isoniazid-resistant pulmonary tuberculosis. *BMC Infectious Diseases*, 8(6), January 2008.

[21] G.B. Migliori, R. Centis, L. D'Ambrosio, A. Spanevello, E. Borroni, D.M. Cirillo, and G. Sotgiu. Totally drug-resistant and extremely drug-resistant tuberculosis: The same disease? *Clinical Infectious Diseases*, 54(9):1379–1380, May 2012.

[22] G.B. Migliori, G. Sotgiu, N.R. Gandhi, D. Falzon, K. DeRiemer, R. Centis, M.G. Hollm-Delgado, D. Palmero, C. Pérez-Guzmán, M.H. Vargas, L. D'Ambrosio, A. Spanevello, M. Bauer, E.D. Chan, H.S. Schaaf, S. Keshavjee, T.H. Holtz, D. Menzies, and "The Collaborative Group for Meta-Analysis of Individual Patient Data in MDR-TB". Drug resistance beyond xdr-tb: results from a large individual patient data meta-analysis. *European Respiratory Journal*, 42(1):169–179, July 2013.

[23] U.S. Department of Health Human Services, Centers for Disease Control, Prevention, National Center for Emerging, and Zoonotic Infectious Diseases. Cdc immigration requirements: Technical instructions for tuberculosis screening and treatment using cultures and directly observed therapy. http://www.cdc.gov/immigrantrefugeehealth/pdf/tuberculosis-ti-2009.pdf, October 2009.

[24] World Health Organization. Global tuberculosis report 2015, 2015.

[25] A. Pablos-Méndez, M.C. Raviglione, A. Laszlo, N. Binkin, H.L. Rieder, F. Bustreo, D.L. Cohn, C.S.B. Lambregts van Weezenbeek, S.J. Kim, P. Chaulet, and P. Nunn. Global surveillance for antituberculosis-drug resistance, 1994-1997. *The New England Journal of Medicine*, 338(23):1641–1649, June 1998.

[26] L.B. Reichman and J.H. Tanne. *Timebomb: The Global Epidemic of Multi-Drug-Resistant Tuberculosis*. McGraw-Hill, 2002.

[27] Philip M. Ricks, Kevin P. Cain, John E. Oeltmann, J. Steve Kammerer, and Patrick K. Moonan. Estimating the burden of tuberculosis among foreign-born persons acquired prior to entering the u.s., 2005-2009. *PLoS ONE*, 6(11), November 2011.

[28] Kimberly M. Shea, J. Steve Kammerer, Carla A. Winston, Thomas R. Navin, and Jr. C. Robert Horsburgh. Estimated rate of reactivation of latent tuberculosis infection in the united states, overall and by population subgroup. *Am J Epidemiol.*, 179(2):216–225, October 2013.

[29] Shashikant Srivastava, Jotam G. Pasipanodya, Claudia Meek, Richard Leff, and Tawanda Gumbo. Multidrug-resistant tuberculosis not due to noncompliance

but to between-patient pharmacokinetic variability. *The Journal of Infectious Diseases*, 204:1951–1959, 2011.

[30] Timothy R. Sterling, Harold P. Lehmann, and Thomas R. Frieden. Impact of dots compared with dots-plus on multidrug resistant tuberculosis and tuberculosis deaths: decision analysis. *British Medical Journal*, 326:574–577, March 2003.

[31] E.W. Tiemersma, M.J. van der Werf, M.W. Borgdorff, B.G. Williams, and N.J.D Nagelkerke. Natural history of tuberculosis: Duration and fatality of untreated pulmonary tuberculosis in hiv negative patients: A systematic review. *PLoS ONE*, 6(4), April 2011.

[32] T.E. Tupasi, S. Radhakrishna, M.L. Pascual, M.I.D Quelapio, M.L. Villa, V.M. Co, J. Sarrol, N. Mangubat, A.C. Reyes, A. Sarmiento, M. Solon, F. Solon, L. Burton, V.S. Lofranco, M.P.C. Rostrata, and M.J. Mantala. Bcg coverage and the annual risk of tuberculosis infection over a 14-year period in the philippines assessed from the nationwide prevalence surveys. *The International Journal of Tuberculosis and Lung Disease*, 4(3):216–222, 2000.

[33] P.W. Uys, R. Warren, P.D. van Helden, M. Murray, and T.C. Victor. Potential of rapid diagnosis for controlling drug-susceptible and drug-resistant tuberculosis in communities where mycobacterium tuberculosis infections are highly prevalent. *Journal of Clinical Microbiology*, 47(5):1484–1490, May 2009.

[34] A.A. Velayati, M.R. Masjedi, P. Farnia, P. Tabarsi, J. Ghanavi, A.H. ZiaZarifi, and S.E. Hoffner. Emergence of new forms of totally drug-resistant tuberculosis bacilli: Super extensively drug-resistant tuberculosis or totally drug-resistant strains in iran. *Chest*, 136:420–425, 2009.

[35] Ana Sofia Vilariça, Carlos Gomes, and Jaime Pina. Comparative analysis of multidrug-resistant tuberculosis and extensively drug-resistant tuberculosis - epidemiology and predictive factors. *Revista Portuguesa de Pneumologia*, 14(6):829–842, 2008.

[36] E. Vynnycky and R.G. White. *An Introduction to Infectious Disease Modelling*. Oxford University Press, 2010.

[37] N.D. Walter, J. Painter, M. Parker, P. Lowenthal, J. Flood, Y. Fu, R. Asis, and R. Reves. Persistent latent tuberculosis reactivation risk in united states immigrants. *American Journal of Respiratory and Critical Care Medicine*, 189(1):88–95, January 2014.

[38] J.J. Wilkins, G. Langdon, H. McIlleron, G.C. Pillai, P.J. Smith, and U.S.H. Simonsson. Variability in the population pharmacokinetics of pyrazinamide in south african tuberculosis patients. *European Journal of Clinical Pharmacology*, 62:727–735, May 2006.

[39] J. Xu, S.L. Murphy, K.D. Kochanek, and B.A. Bastian. Deaths: Final data for 2013. *National Vital Statistics Reports*, 64(2), February 2016.

[40] J. Zong and J. Batalova. Frequently requested statistics on immigrants and immigration in the united states. *Migration Policy Institute*, February 2015.