

Distribution Agreement

In presenting this thesis as a partial fulfillment of the requirements for a degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis in whole or in part in all forms of media, now or hereafter now, including display on the World Wide Web. I understand that I may select some access restrictions as part of the online submission of this thesis. I retain all ownership rights to the copyright of the thesis. I also retain the right to use in future works (such as articles or books) all or part of this thesis.

Erik Herslebs

April 10, 2018

Modeling the Susceptibility of Treated Individuals to Further Tuberculosis Episodes

by

Erik Herslebs

Dr. Rustom Antia
Adviser

Department of Biology

Dr. Rustom Antia
Adviser

Dr. Cheryl L. Day
Committee Member

Dr. Lance A. Waller
Committee Member

2018

Modeling the Susceptibility of Treated Individuals to Further Tuberculosis Episodes

By

Erik Herslebs

Dr. Rustom Antia

Adviser

An abstract of
a thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Sciences with Honors

Department of Biology

2018

Abstract

Modeling the Susceptibility of Treated Individuals to Further Tuberculosis Episodes

By Erik Herslebs

The risk of relapse and reinfection among individuals treated for tuberculosis (TB) is an important area of study. It has been found that treated individuals are four times more likely to become infected with another episode of tuberculosis than those with no previous infection. In order to explore possible causes for this, six different mathematical models are created to describe different hypotheses of the post-treatment behavior of individuals in a hyperendemic community. These hypotheses include heterogeneity in susceptibility among hosts and initially high, but waning susceptibility to subsequent TB. It was found by fitting the models to relapse and reinfection incidence data that host heterogeneity and waning susceptibility can both adequately explain trends in the data, while other, simpler models cannot.

Modeling the Susceptibility of Treated Individuals to Further Tuberculosis Episodes

By

Erik Herslebs

Dr. Rustom Antia

Adviser

A thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Sciences with Honors

Department of Biology

2018

Acknowledgements

I would like to thank Dr. James Moore and all members of the Antia lab for their help with this project.

Table of Contents

1. Introduction.....	1
2. Materials and Methods.....	5
2.1 Experimental Data.....	5
2.2 Mathematical Models.....	7
Instantaneous Transition.....	7
Simple Treatment Variation.....	9
Exposed Class.....	11
Immune Stabilization.....	13
Waning Susceptibility.....	16
Host Heterogeneity.....	20
2.3 Fitting Procedure.....	25
3. Results.....	27
3.1 Results from Literature.....	27
3.2 Determining Within-Host Parameters.....	28
3.3 Model Discrimination.....	30
4. Discussion.....	40
References.....	42

Figures and Tables

Figure 1: Flow Diagram of Instantaneous Transition Model.....	8
Table 1: Parameters for Instantaneous Transition Model.....	9
Figure 2: Flow Diagram of Simple Treatment Variation Model.....	10
Table 2: Parameters for Simple Treatment Variation Model.....	11
Figure 3: Flow Diagram of Exposed Model.....	11
Table 3: Parameters for Exposed Model.....	12
Figure 4: Flow Diagram of Immune Stabilization Model.....	13
Table 4: Parameters for Immune Stabilization Model.....	14
Table 5: Parameters for Immune Stabilization Model With Treatment Variation.....	15
Figure 5: Flow Diagram of Waning Susceptibility Model.....	17
Table 6: Parameters for Waning Susceptibility Model.....	18
Table 7: Parameters for Waning Susceptibility Model With Treatment Variation.....	19
Figure 6: Flow Diagram of Host Heterogeneity Model.....	21
Table 8: Parameters for Host Heterogeneity Model.....	22
Table 9: Parameters for Host Heterogeneity Model With Treatment Variation.....	24
Figure 7: Flow Diagram for Within-Host Disease Progression.....	25
Figure 8: Reinfection Case Percentages After Treatment According to Uys et al. (2015).....	27
Figure 9: Active Disease Progression Over Time Following Initial Infection.....	28
Figure 10: Plotted Model Fitting Results.....	33
Table 10: Quantified Model Fitting Results.....	34
Figure 11: Plotted Model Fitting Results with Treatment Variation.....	35
Figure 12: Plotted Model Fitting Results with Rapid Progression Rate Fit.....	35

Figure 13: Plotted Model Fitting Results with both Treatment Variation and Rapid Progression Rate Fit.....	36
Figure 14: Plotted Model Fitting Results with Background Reinfection Rate Fit.....	37
Table 11: Quantified Model Fitting Results with Rapid Progression Rate Fit.....	38
Table 12: Quantified Model Fitting Results with Background Reinfection Rate Fit.....	39

1. Introduction

Tuberculosis (TB) is a disease caused by the *Mycobacterium tuberculosis* complex (MTBC), a group of related *Mycobacterium* species (Pai et al. 2016). Currently, tuberculosis is one of the top 10 causes of death throughout the world, with 1.7 million people succumbing to the disease in 2016 alone. Additionally, 10.4 million people developed symptomatic TB in 2016, of which 1 million were children (“Fact Sheet on Tuberculosis” 2018). This makes TB a high-priority issue for world health.

TB primarily infects the lungs, although infections can occur elsewhere in the body (Pai et al. 2016), and it spreads when those with symptomatic pulmonary TB release the bacteria into the air through coughing, speaking, or sneezing (“Global tuberculosis report 2017”). TB infections are also characterized by their long duration due to latency. During latency, the bacteria have infected an individual, but symptoms have not arisen and a risk is present of progressing to symptomatic TB (Esmail et al. 2014). The latency of TB is complex and can be best understood as a spectrum of possible outcomes to infection, ranging from elimination of the bacteria to maintenance of the infection to subclinical symptomatic disease (Barry et al. 2009, Pai et al. (2016)).

Currently, the most widely used test for active, symptomatic TB in middle and low-income countries is sputum smear microscopy (Pai et al. 2016, Kik et al. (2014)). This test involves the direct observation of respiratory sputum samples (from coughing) to identify the presence of bacteria. However, sputum smear microscopy can only identify 50-60% of active TB cases (Siddiqi, Lambert, and Walley 2003), and extrapulmonary TB could also occur, which sputum smears would not identify. This method is even more problematic when it comes to identifying tuberculosis in children because they are less able to produce sputum (Pai et al. 2016, Swaminathan and Ramachandran (2015)). Culture samples can also be used to identify TB (Pai et al. 2016), which can be better for lower bacterial loads in sputum or for testing samples of suspected extrapulmonary TB.

Treatment and vaccination are two strategies that are currently employed to counter TB ("Global tuberculosis report 2017", Pai et al. (2016)). Those that are infected with the disease are treated for about 6 months with four antibiotics: isoniazid, rifampicin, ethambutol, and pyrazinamide ("Global tuberculosis report 2017"). For prevention, the current widely used vaccine is the bacille Calmette-Guérin (BCG) vaccine (Fine 1995), which can induce protection against severe TB ("Global tuberculosis report 2017"). However, several problems have complicated the elimination of TB. One problem is the spread of multidrug-resistant TB (MDR-TB) from the use of antibiotics. In the Russian Federation, for example, the overall TB incidence and TB-related death is decreasing, but the MDR-TB incidence is increasing ("Global tuberculosis report 2017"). Additionally, the HIV epidemic has become an enormous problem for the elimination of TB, since those that are infected with HIV are 20-30 times more likely to develop symptomatic TB ("Fact Sheet on Tuberculosis" 2018). Unfortunately, the BCG vaccine is, at best, of limited efficacy as well ("Global tuberculosis report 2017", Fine (1995)). Estimations of the efficacy of the vaccine vary considerably (Fine 1995, Colditz et al. (1994)), with an average TB risk reduction of 50% (Colditz et al. 1994). In 2017, 12 new vaccines were in trials ("Global tuberculosis report 2017"). However, the complexity of tuberculosis means that the likelihood of failure is high (Graham, Ledgerwood, and Nabel 2009, Fletcher and Schragar (2016)), and during the phase 2b clinical trial of one vaccine, MVA85A, there was only a 17.3% protection against incident tuberculosis, and a -3.8% protection against TB infection (Tameris et al. 2013).

In order to meet the goal of the WHO to reduce TB incidence by 90% and TB-related death by 95% before 2035 ("The End Tb Strategy: A Global Rally" 2014), new strategies would be beneficial (Esmail et al. 2014). A greater understanding of the epidemiology of the disease may be advantageous in improving these prevention methods. For example, the nature of latency is poorly understood. Since it is difficult to know the exact time of infection, we do not know with great certainty how long the latent period typically is. Thus, the current TB incidence could be explained by a small population of infected individuals that rapidly progress to disease or a large population that progress slowly. Also, not much is

known about the activity of the disease after treatment. While most infections induce an immune response after recovery that protects from reinfection, the opposite has been seen in tuberculosis. Those that are successfully treated are actually four times more likely to be infected than those with no previous infection (Verver et al. 2005). This phenomenon will be referred to as “supersusceptibility” of treated individuals, and it severely complicates our ability to control tuberculosis in areas of high prevalence. Additionally, on the individual level, supersusceptibility limits our ability to properly treat patients, since they have an unusually high rate of returning to active TB.

The exact nature of supersusceptibility is not known, which makes prevention difficult. It could be that all individuals who are treated show elevated rates of subsequent disease, or it may be more complex. It has been seen that tuberculosis causes suppression of the immune system (Sahiratmadja et al. 2007, Hirsch et al. (1999)), so if the antibiotic treatment regimen or bacteria activity cause immunosuppression, one would expect a higher risk of reinfection after treatment that would decrease over time as the immune system recovers. This will be referred to as a “waning susceptibility” event. Additionally, genetic factors have been shown to play a role in tuberculosis susceptibility (Abel et al. 2018, Puffer (1944), Kallmann and Reisner (1943), Cobat et al. (2012)). If some people are more susceptible because of these genetic factors or because they live with others who have TB in an area of high prevalence, then these individuals would compose a large fraction of the active cases we see, and thus supersusceptibility would be the result of treated individuals being mainly a higher-risk group to begin with. In this case, what we will define as “host heterogeneity” is the underlying cause.

The behavior of this increased susceptibility after recovery could have great effects on the strategies needed to eradicate the disease, and has the potential to change the current understanding of the activity of tuberculosis in hyperendemic communities. To this end, population dynamic models can be instructive (D. W. Dowdy, Dye, and Cohen 2013). Predictive mathematical models are especially useful for TB, as the disease’s long latent periods and slow recovery times makes acquiring accurate, long-term data difficult (Ozcaglar et al. 2012). As a result, these models can be used to test hypotheses that, as of

now, data cannot. Current models serve as good representations of the global dynamics of the disease. However, these models do not account for supersusceptibility, and this reduces the accuracy of predictions regarding post-treatment behavior using these models. The objective of this study is to improve these models while simultaneously gaining an understanding of the cause of supersusceptibility.

In accomplish this, nine differential equation models were constructed to describe distinct biologically plausible scenarios regarding the post-treatment behavior of TB patients to subsequent disease (i.e. reinfection, infection with a new strain of tuberculosis, or relapse, the reactivation of bacteria that remained in the body after the index case). These models were then used to determine the likelihood of each scenario being true when faced with data. These models included three of the most simplistic representations of this biology, which began with treated individuals simply experiencing disease due to relapse and reinfection at two separate rates. This was followed by the addition of “treatment variation” to this model. For this study, this was defined as two possible outcomes of the index case treatment. Either the initial infection was completely eliminated, leaving no risk of relapse, or the infection remained to some degree, allowing a risk for relapse. Finally, an disease exposure condition was added for reinfected individuals in order to incorporate latency of the disease.

The fourth model included two latent classes of individuals, depending on the time since infection, and each progressed to active disease at different rates. The next two models adjusted the previous model to include waning susceptibility of treated individuals to subsequent disease or heterogeneity of susceptibility among treated individuals. Finally, the last three models studied were modified versions of the previous three to include treatment variation.

The last six models contained identical biological parameters that were assumed to be constant across the entire population: the rate of progressing to active disease soon after infection, the rate of progressing to disease long after infection, and the rate determining the line between “soon” and “long” after infection. These parameters were estimated using data on the percentage of infected individuals that

had progressed to active disease each year after infection (extrapolated from Diel et al. (2008) and Borgdorff et al. (2011)). The portion of the models containing these parameters were fit to this data and held constant through the study, but these parameter assumptions were also relaxed to determine their impact on the final results.

Finally, each of the nine models was then fit to data on the cumulative relapse and reinfection incidence for every year after treatment among those in a hyperendemic community (Marx et al. 2014). The ability for these models to produce cumulative relapse and reinfection incidences that matched the data, as calculated using the programming language R, was used to deduce the validity of each of the nine hypotheses. These results were quantified using the Akaike information criterion (AIC) equation.

2. Materials and Methods

2.1 Experimental Data

We use data from three studies. The first study by Borgdorff et al. (2011) presents the Kaplan-Meier probabilities of developing symptomatic TB within 15 years of infection for those individuals who do progress in this time period. The authors took Netherlands patient data from 1993 to 2007 to determine case connections. If two patients had genetically identical TB fingerprints, as determined by standard restriction fragment length polymorphism (RFLP) typing and, if necessary, sub-typing on patient cultures, then they were interviewed for a possible contact. If a contact was found, then transmission was said to have occurred between the two patients. The time of infection for the secondary case was then approximated as the halfway point between the onset of symptoms in the source case and either the onset of symptoms or diagnosis of the secondary case. Data from the Kaplan-Meier probability curve was extracted for this study using Engauge Digitizer (digitizer.sourceforge.net).

The second study by Diel et al. (2008) provided data on the proportion of individuals who progress to symptomatic TB within two years of infection. In this study from May 2005 until April 2006,

those in close contact with smear-positive and culture-confirmed TB cases were tested for latent infection within 8 weeks of identification of the index case. These individuals were tested using both the QuantiFERON-TB Gold In-Tube assay (QFT) and the tuberculin skin test (TST), and were observed until September 2007 for the development of TB. 601 close contacts were found, 243 of which were TST positive (at 5 mm) and 66 were QFT positive. 6 of these 601 individuals progressed to disease after declining treatment. All 6 were identified by QFT but only 5 were identified with TST, leading to progression rates of 14.6% and 2.3%, respectively, within the 2-year follow-up. For our study, the estimate provided by the QFT test was used because, while both tests have limitations, the TST test has been known to produce false positives in those vaccinated with BCG or infected with non-tuberculosis causing strains of *mycobacteria*. The QFT test, on the other hand, is more specific to MTB and does not produce these false positives or increase the likelihood of false positives in subsequent tests (Centers for Disease Control and Prevention 2010, Diel et al. (2008)).

For use in this study, the data extracted from Borgdorff et al. (2011) was adjusted using the progression rate from Diel et al. (2008). The initial probability curve included only individuals who eventually progressed, but an estimation including all infected individuals regardless of outcome could be created by scaling the Borgdorff et al. (2011) data to match a 14.6% risk of progression after 2 years. This results in about 25% of infected individuals progressing to symptomatic TB after 15 years (Figure 9). This is slightly above the current estimates of 5-15% lifetime risk of progressing to active TB (“Fact Sheet on Tuberculosis” 2018). However, those at higher risk of TB, including those with HIV, those with previously cured TB episodes, those who suffer from malnutrition or diabetes, or those who use tobacco products, can have lifetime risks of over 20% (“Fact Sheet on Tuberculosis” 2018, Horsburgh (2004)).

Finally, the third study by Marx et al. (2014) provided the incidences of relapse and reinfection each year following index case treatment. The study included individuals in Cape Town, South Africa who were treated for TB between 1996 and 2008. Sputum smear-positive index and subsequent cases were considered, and sputum samples were used for fingerprinting. Individuals were classified as relapsed

if the DNA fingerprint of the new TB case was identical to the initial case or differed by up to 2 bands. Otherwise, the patient was classified as reinfected. The reinfection and relapse incidences were reported as the number of cases per 100,000 treated index cases, but for our study, the cumulative incidences of reinfection and relapse were used. Additionally, since the incidences for every year after the fourth were combined, this data could only be used for the maximum time point possible. It is for this reason that the final data point is at year 13, which is the longest period of time until secondary disease that would be documented in this study, if the initial treatment was completed in 1996 and secondary disease occurred in 2008. Thus, this is an underestimation of the cumulative number of secondary cases at year 13, but this estimate will still be useful in describing the general trend of secondary disease risk over time.

2.2 Mathematical Models

Instantaneous Transition

The first model to describe the behavior of individuals after treatment for tuberculosis is the simplest possible representation. Every individual after treatment has the potential to either relapse or become reinfected. This model assumes that no treatment for TB successfully eliminates the bacteria within the host, leaving all individuals capable of relapsing with the same disease. Additionally, the model assumes no latency of the disease. This is known to be biologically inaccurate and this accuracy is sacrificed for simplicity. Finally, it is assumed for this model and all subsequent models that the background prevalence is constant for the time period simulated. The rate of infection for diseases is dependent on both the number of infections and the number of susceptible people to become infected. For the sake of this study, it is assumed that the number of infected individuals in the environment remains constant, and thus the force of infection is a constant rate. The resulting compartmental model can be visualized as follows:

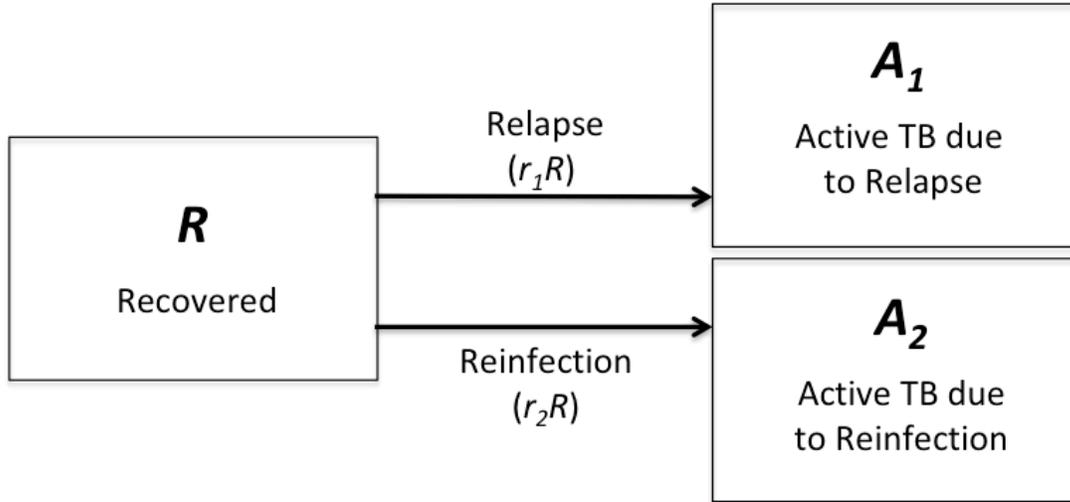


Figure 1: Illustration of the flow rates for a model with instantaneous reinfection. This is the simplest representation of recurrent disease. The treated and recovered population can move into the active TB population due to relapse and reinfection at two separate rates.

The equations for this model, where r_1 is the relapse rate and r_2 is the reinfection rate, are:

$$\frac{dR}{dt} = -(r_1 + r_2)R$$

$$\frac{dA_1}{dt} = r_1R$$

$$\frac{dA_2}{dt} = r_2R$$

Table 1: Parameters used in the instantaneous model as well as the origin of the parameters. For this model, all parameters were fit to yearly reinfection and relapse risk data. These parameters describe the flow rates in terms of the fraction of the population.

<u>Parameter</u>	<u>Symbol</u>	<u>Value</u>	<u>Source</u>
Reinfection Rate	r_2	0.01 per year	Fit in this Study
Relapse Rate	r_1	0.01 per year	Fit in this Study

Simple Treatment Variation

The second model incorporates individuals both with and without elimination of the initial infection (R_2 and R_1 classes, respectively). In both classes, reinfection may occur, causing active disease. However, in only the latter treated class may relapse occur. This is based on the assumption that, while exogenous reinfection may occur in any recovered individual, relapse due to the same infection cannot occur unless the infection remains after treatment.

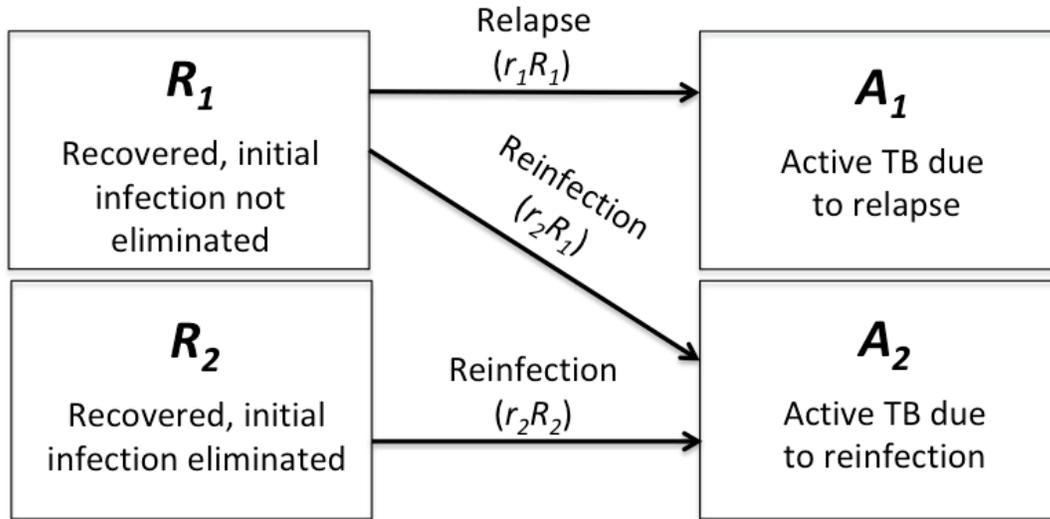


Figure 2: This illustration describes the flow rates of the instantaneous transition model with the addition of two treated classes, with and without the possibility of relapse. They may still become reinfected at the same rate, however.

This model can be represented by:

$$\frac{dR_1}{dt} = -(r_1 + r_2)R_1$$

$$\frac{dR_2}{dt} = -r_2 R_2$$

$$\frac{dA_1}{dt} = r_1 R_1$$

$$\frac{dA_2}{dt} = r_2 (R_1 + R_2)$$

Table 2: Parameter values used in the simple treatment variation model. All values were fit in this study to yearly relapse and reinfection incidence data.

<u>Parameter</u>	<u>Symbol</u>	<u>Value</u>	<u>Source</u>
Fraction in R_1	$\frac{R_1}{R_1 + R_2}$	0.05	Fit in this Study
Reinfection Rate	r_2	0.01 per year	Fit in this Study
Relapse Rate	r_1	0.63 per year	Fit in this Study

Exposed Class

The third model modifies the second by applying latently reinfected individuals in the form of a single exposed class (E), which may progress at a rate p back to active disease. This model increases the biological accuracy of the model. However, it is known that those with recent TB infection have a higher risk of progression to disease than those with a more remote TB infection (Horsburgh et al. 2010).

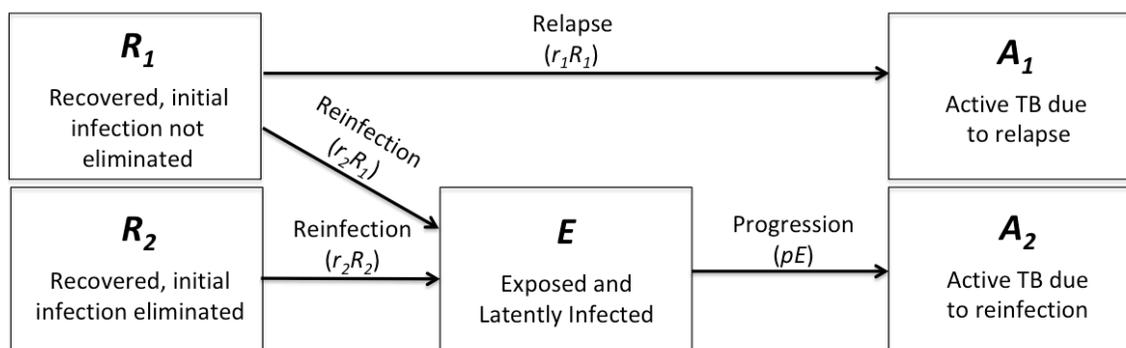


Figure 3: Illustration of the flow between populations in a model implementing latent infection. This latent infection is added through a single exposed class through which reinfected individuals must pass before experiencing active disease. Treatment variation remains in this model as well.

This model's equations are:

$$\frac{dR_1}{dt} = -(r_1 + r_2)R_1$$

$$\frac{dR_2}{dt} = -r_2R_2$$

$$\frac{dE}{dt} = r_2(R_1 + R_2) - pE$$

$$\frac{dA_1}{dt} = r_1R_1$$

$$\frac{dA_2}{dt} = pE$$

Table 3: Parameters used in the model with a latent, or exposed, class. All parameters were fit in this paper to yearly reinfection and relapse risk data. Flow rates show movement in terms of population fractions.

<u>Parameter</u>	<u>Symbol</u>	<u>Value</u>	<u>Source</u>
Fraction in R_1	$\frac{R_1}{R_1 + R_2}$	1	Fit in this Study
Reinfection Rate	r_2	1.71 per year	Fit in this Study
Relapse Rate	r_1	0.08 per year	Fit in this Study
Progression Rate	p	0.006 per year	Fit in this Study

Immune Stabilization

Next, the fourth model represents the behavior described by Dowdy, Dye, and Cohen (D. W. Dowdy, Dye, and Cohen 2013), in which a treated individual enters a recovered class after treatment that can allow for relapse back to active disease (A_1 class) or reinfection back to latency. Latently infected individuals may rapidly progress to active disease (in the recently infected, or L_1 class) or stabilize and have a lower chance of becoming actively infected (in the remotely infected, or L_2 class).

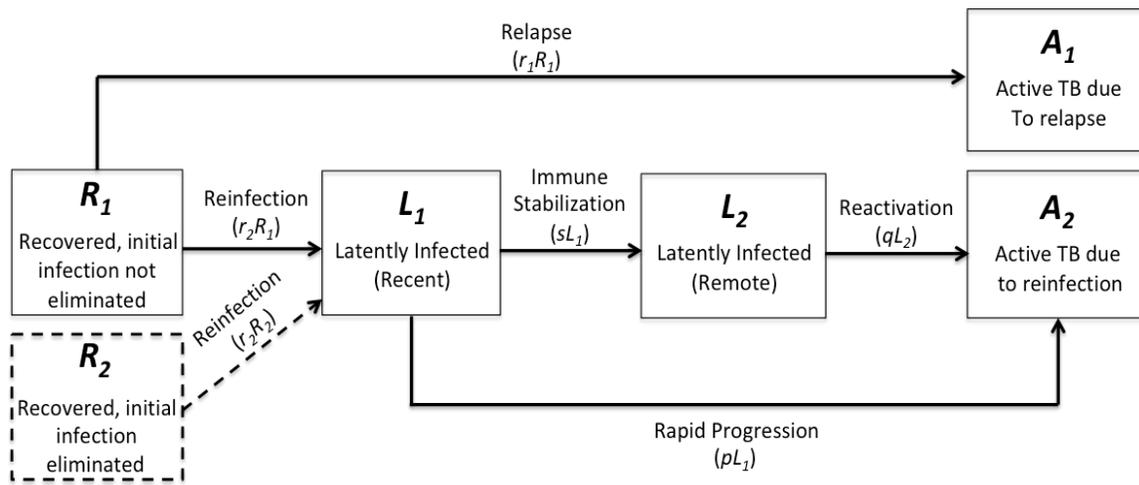


Figure 4: Flow diagram of the exposed model with immune stabilization. This diagram adds a second latent class, a remotely infected class that infected individuals can move into after a certain amount of time. This class has a lower rate of progressing into active disease, referred to as reactivation.

Additionally, either one or two treatment classes may be used with this model, as indicated by the dotted lines.

The equations for this model are as follows, where s is the rate of latent stabilization, and p and q are the rates of progression to active disease for the recently and remotely latent classes, respectively:

$$\frac{dR_1}{dt} = -(r_1 + r_2)R_1$$

$$\frac{dL_1}{dt} = r_2 R_1 - (s + p)L_1$$

$$\frac{dL_2}{dt} = sL_1 - qL_2$$

$$\frac{dA_1}{dt} = r_1 R_1$$

$$\frac{dA_2}{dt} = pL_1 + qL_2$$

Table 4: Parameters used in the model with immune stabilization. The reinfection and relapse rates are fit to yearly reinfection and relapse risk data. The remaining parameters are extrapolated using data from both Borgdorff et al. (2011) and Dye et al. (2008)

<u>Parameter</u>	<u>Symbol</u>	<u>Value</u>	<u>Source</u>
Reinfection Rate	r_2	0.05 per year	Fit in this Study
Relapse Rate	r_1	0.08 per year	Fit in this Study
Rapid Progression Rate	p	0.16 per year	(Diel et al. 2008, Borgdorff et al. 2011)
Immune Stabilization Rate	s	0.78 per year	(Diel et al. 2008, Borgdorff et al. 2011)
Reactivation Rate	q	0.008 per year	(Diel et al. 2008, Borgdorff et al. 2011)

Treatment variation can also be inserted into this model to produce the following equations, with the R_2 and R_1 classes representing those with and without elimination of the initial infection, respectively:

$$\frac{dR_1}{dt} = -(r_1 + r_2)R_1$$

$$\frac{dR_2}{dt} = -r_2R_2$$

$$\frac{dL_1}{dt} = r_2(R_1 + R_2) - (s + p)L_1$$

$$\frac{dL_2}{dt} = sL_1 - qL_2$$

$$\frac{dA_1}{dt} = r_1R_1$$

$$\frac{dA_2}{dt} = pL_1 + qL_2$$

Table 5: Parameters used in the immune stabilization model with treatment variation. This model differs from the immune stabilization model by the addition of a single parameter for the fraction of treated individuals without elimination of the initial infection. This parameter was fit using the yearly reinfection and relapse risk data.

<u>Parameter</u>	<u>Symbol</u>	<u>Value</u>	<u>Source</u>
Fraction in R_1	$\frac{R_1}{R_1 + R_2}$	0.05	Fit in this Study
Reinfection Rate	r_2	0.03 per year	Fit in this Study
Relapse Rate	r_1	0.68 per year	Fit in this Study

Rapid Progression Rate	p	0.16 per year	(Diel et al. 2008, Borgdorff et al. 2011)
Immune Stabilization Rate	s	0.78 per year	(Diel et al. 2008, Borgdorff et al. 2011)
Reactivation Rate	q	0.008 per year	(Diel et al. 2008, Borgdorff et al. 2011)

Waning Susceptibility

Individuals after treatment may have high susceptibility to recurrent disease, which decreases over time, due to immunosuppression from disease or treatment. We call this phenomenon waning susceptibility.

In the model, we modify the immune stabilization model to include waning susceptibility. All treated individuals begin in the high susceptibility class (R_1) and this population is transferred to the low susceptibility class (R_3) at a waning susceptibility rate w . In this case, it is assumed that the low susceptibility class has no susceptibility to relapse and a background risk of reinfection. This background risk was determined as the average risk of infection in areas of high endemicity. Cobelens et al. (2000) provides this information in the form of the monthly infection incidence in areas of high TB prevalence. In this study by Cobelens et al., individuals from the Netherlands who travelled to areas of high TB prevalence for 3-12 months were investigated. A TST was performed on the individuals before and after the trip to verify that the infection occurred during this time. It was found that 12 out of the 656 tested individuals contracted TB, 2 of which had active TB. The resulting monthly incidence that was calculated

in the study was 2.8 per 1,000 individuals. This was expanded to a yearly incidence of 33.6 per 1,000 for the current study, and was implemented in this model.

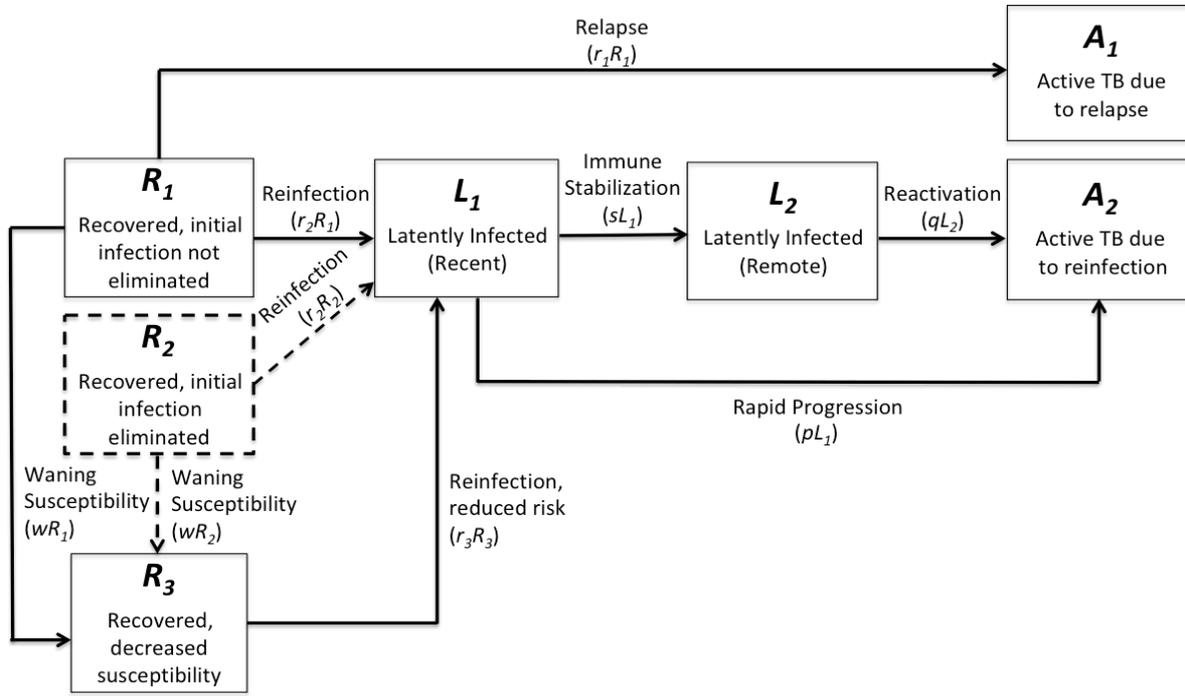


Figure 5: Flow diagram for the adjusted immune stabilization model with waning susceptibility of treated individuals to both reinfection and relapse. A compartment representing the population of treated individuals who cannot relapse may be added as well, as indicated by the dotted lines.

The resulting equations for this model are:

$$\frac{dR_1}{dt} = -(r_1 + r_2 + w)R_1$$

$$\frac{dR_3}{dt} = wR_1 - r_3R_3$$

$$\frac{dL_1}{dt} = r_2R_1 + r_3R_3 - sL_1 - pL_1$$

$$\frac{dL_2}{dt} = sL_1 - qL_2$$

$$\frac{dA_1}{dt} = r_1 R_1$$

$$\frac{dA_2}{dt} = pL_1 + qL_2$$

Table 6: Parameters used in the waning susceptibility model. The flow rates refer to the fraction of the population that is transitioning. The rates of reinfection, relapse, and waning susceptibility are determined in this study by fitting the models to yearly reinfection and relapse risk data, while the remaining parameters are extrapolated from Borgdorff et al. (2011) and Diel et al. (2008).

<u>Parameter</u>	<u>Symbol</u>	<u>Value</u>	<u>Source</u>
Reinfection Rate	r_2	0.1 per year	Fit in this Study
Relapse Rate	r_1	0.05 per year	Fit in this Study
Rapid Progression Rate	p	0.16 per year	(Diel et al. 2008, Borgdorff et al. 2011)
Immune Stabilization Rate	s	0.78 per year	(Diel et al. 2008, Borgdorff et al. 2011)
Reactivation Rate	q	0.008 per year	(Diel et al. 2008, Borgdorff et al. 2011)
Waning Susceptibility Rate	w	0.85 per year	Fit in this Study

Next, treatment variation can be inserted into this model as well. With the R_1 (initial infection not eliminated, high susceptibility) and R_2 (initial infection eliminated, high susceptibility) classes added, the equations are as follows:

$$\frac{dR_1}{dt} = -(r_1 + r_2 + w)R_1$$

$$\frac{dR_2}{dt} = -(r_2 + w)R_2$$

$$\frac{dR_3}{dt} = w(R_1 + R_2) - r_3R_3$$

$$\frac{dL_1}{dt} = r_2(R_1 + R_2) + r_3R_3 - sL_1 - pL_1$$

$$\frac{dL_2}{dt} = sL_1 - qL_2$$

$$\frac{dA_1}{dt} = r_1R_1$$

$$\frac{dA_2}{dt} = pL_1 + qL_2$$

Table 7: Parameters used in the waning susceptibility model with treatment variation. This means that the fraction of treated individuals who have not eliminated the initial infection must also be fit using yearly reinfection and relapse risk data.

<u>Parameter</u>	<u>Symbol</u>	<u>Value</u>	<u>Source</u>
Fraction in R_1	$\frac{R_1}{R_1 + R_2}$	1	Fit in this Study

Reinfection Rate	r_2	0.1 per year	Fit in this Study
Relapse Rate	r_1	0.05 per year	Fit in this Study
Rapid Progression Rate	p	0.16 per year	(Diel et al. 2008, Borgdorff et al. 2011)
Immune Stabilization Rate	s	0.78 per year	(Diel et al. 2008, Borgdorff et al. 2011)
Reactivation Rate	q	0.008 per year	(Diel et al. 2008, Borgdorff et al. 2011)
Waning Susceptibility Rate	w	0.85 per year	Fit in this Study

Host Heterogeneity

The last hypothesis tested is that of host heterogeneity. This assumes that supersusceptibility is the result of differences among individuals, either genetically or socioeconomically, which leads to higher susceptibility to disease in one population than another. Treated individuals would have higher susceptibility to subsequent disease because they are most likely the population that is more predisposed to the disease to begin with. This is accomplished in a model by separating the population into high (R_1) and low (R_3) susceptibility groups. In this case, the high susceptibility group will have high rates of recurrent disease while the low susceptibility group will only have the background risk of reinfection, 33.6 per 1,000 treated individuals.

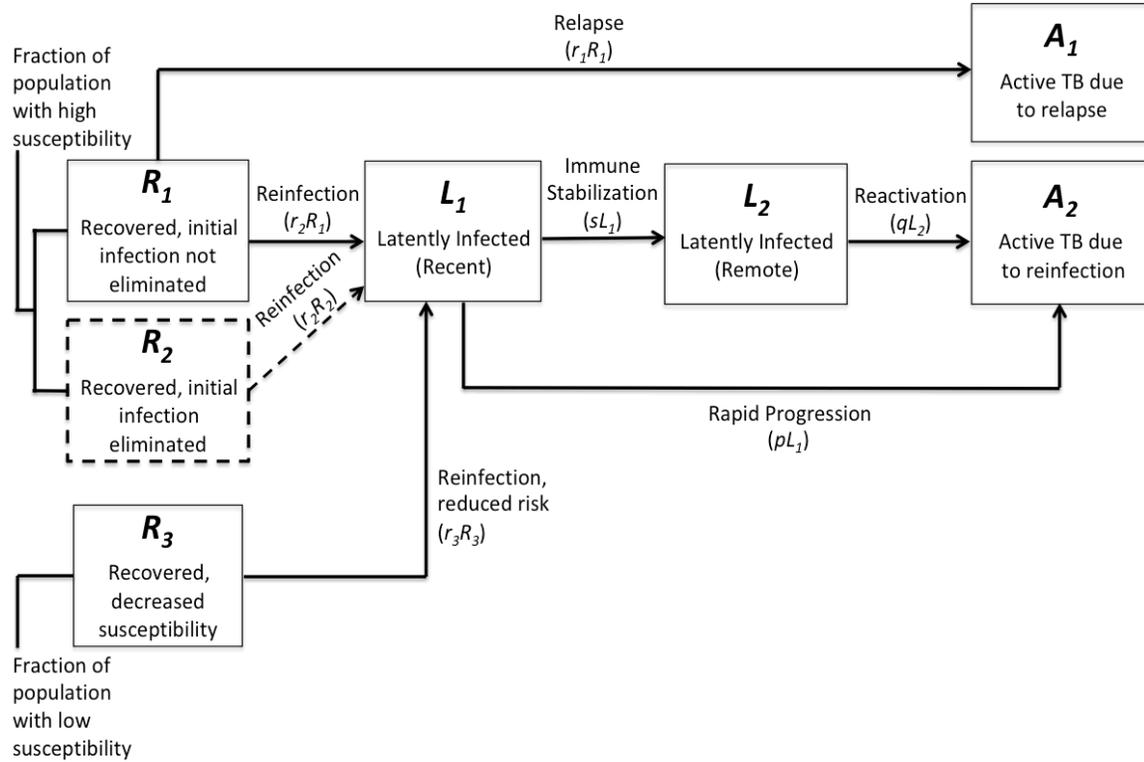


Figure 6: Flow diagram of an adjusted immune stabilization model representing the addition of host heterogeneity. Two types of hosts are added, one with high susceptibility to both relapse and reinfection, and one with no susceptibility to relapse and a background susceptibility to reinfection. The entire population of treated individuals is divided between these two groups initially and there is no movement between them. A compartment representing the population without any risk of relapse can be inserted as indicated by the dotted lines. In this case, the treated population is first divided between the two high susceptibility groups and the low susceptibility group, then subsequently divided between the two high susceptibility groups.

The resulting equations are:

$$\frac{dR_1}{dt} = -(r_1 + r_2)R_1$$

$$\frac{dR_3}{dt} = -r_3R_3$$

$$\frac{dL_1}{dt} = r_2R_1 + r_3R_3 - sL_1 - pL_1$$

$$\frac{dL_2}{dt} = sL_1 - qL_2$$

$$\frac{dA_1}{dt} = r_1R_1$$

$$\frac{dA_2}{dt} = pL_1 + qL_2$$

Table 8: Parameters used in the host heterogeneity model. The total treated population is divided between the highly susceptible group and the low susceptibility group, and the fraction in each was fit, along with the reinfection and relapse rates, to yearly reinfection and relapse risk data. The remaining parameters are extrapolated from Borgdorff et al. (2011) and Diel et al. (2008) The rates that are shown refer to the fraction of the population that is being transitioned in the model.

<u>Parameter</u>	<u>Symbol</u>	<u>Value</u>	<u>Source</u>
Reinfection Rate	r_2	0.6 per year	Fit in this Study
Relapse Rate	r_1	0.39 per year	Fit in this Study
Rapid Progression Rate	p	0.16 per year	(Diel et al. 2008, Borgdorff et al. 2011)
Immune Stabilization Rate	s	0.78 per year	(Diel et al. 2008, Borgdorff et al. 2011)
Reactivation Rate	q	0.008 per year	(Diel et al. 2008,

			Borgdorff et al. 2011)
Fraction in High Susceptibility Group	$\frac{R_1}{R_1 + R_3}$	0.12	Fit in this Study

Finally, we can incorporate treatment variation once again to produce the following equations. R_1 is the highly susceptible class of individuals in which elimination of the initial infection did not occur, while R_2 is the highly susceptible class of individuals in which elimination did occur.

$$\frac{dR_1}{dt} = -(r_1 + r_2)R_1$$

$$\frac{dR_2}{dt} = -r_2R_2$$

$$\frac{dR_3}{dt} = -r_3R_3$$

$$\frac{dL_1}{dt} = r_2(R_1 + R_2) + r_3R_3 - sL_1 - pL_1$$

$$\frac{dL_2}{dt} = sL_1 - qL_2$$

$$\frac{dA_1}{dt} = r_1R_1$$

$$\frac{dA_2}{dt} = pL_1 + qL_2$$

Table 9: Parameters used in the host heterogeneity model with treatment variation. The fraction of the population that is in the R_1 class is fit to the yearly reinfection and relapse risk data along with the reinfection and relapse rates and the fraction in the highly susceptible group.

<u>Parameter</u>	<u>Symbol</u>	<u>Value</u>	<u>Source</u>
Fraction in R_1	$\frac{R_1}{R_1 + R_2}$	1	Fit in this Study
Reinfection Rate	r_2	0.6 per year	Fit in this Study
Relapse Rate	r_1	0.39 per year	Fit in this Study
Rapid Progression Rate	p	0.16 per year	(Diel et al. 2008, Borgdorff et al. 2011)
Immune Stabilization Rate	s	0.78 per year	(Diel et al. 2008, Borgdorff et al. 2011)
Reactivation Rate	q	0.008 per year	(Diel et al. 2008, Borgdorff et al. 2011)
Fraction in High Susceptibility Group	$\frac{R_1 + R_2}{R_1 + R_2 + R_3}$	0.12	Fit in this Study

2.3 Fitting Procedure

Two fitting procedures are utilized in this study. The first uses the adjusted Borgdorff et al. (2011) data to determine three parameter values that are used in the immune stabilization model, the host heterogeneity model, and the waning susceptibility model. These are the within-host biological rates of rapid progression (p), immune stabilization (s), and reactivation (q). Although it is possible that these rates vary between hosts like the rates of reinfection and relapse, it is assumed for this paper that they are constant. It would likely improve the fits if they were allowed to be variable as well but it is assumed that they would do so consistently across the models and thus fitting the rates of relapse and reinfection will be sufficient to discriminate between the hypotheses. This was later verified using the rapid progression rate, the parameter most associated with variation in the model output of these three (D. W. Dowdy, Dye, and Cohen 2013).

The flow diagram for the section of these models being fit, along with the corresponding equations are as follows:

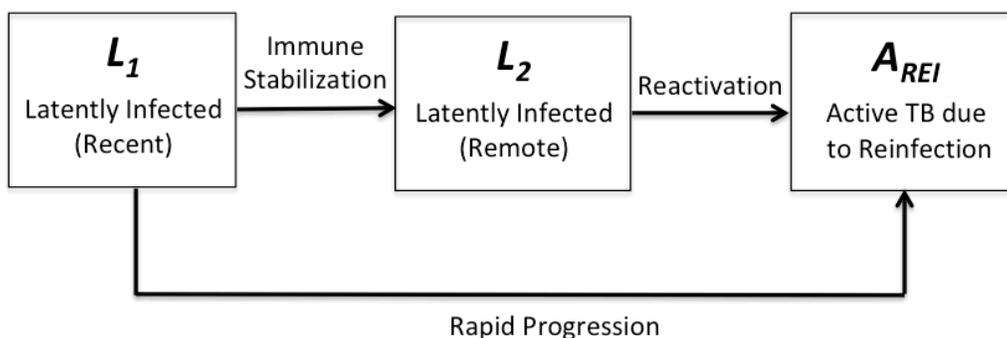


Figure 7: Flow diagram of the portion of the immune stabilization model that shows the progression of hosts to active disease after infection. This is assumed to be a biological constant regardless of the individual, so values determined for each transition rate will be used in all models containing this immune stabilization.

$$\frac{dL_1}{dt} = -sL_1 - pL_1$$

$$\frac{dL_2}{dt} = sL_1 - qL_2$$

$$\frac{dA_{REI}}{dt} = pL_1 + qL_2$$

In the second fitting procedure, the remaining parameters for each model are fit to the Marx et al. (2014) data. The relative success of each fit is used to discriminate between the models. In this procedure, the reinfection rates for both classes of treated individuals (i.e. those with and without elimination of the initial infection) are treated as the same, despite the potential for heterogeneity. This is based on the assumption that the only difference among these individuals is the elimination of the previous disease that could have led to relapse. Otherwise they behave the same way in the model.

Using the programming language R, all models were fit with the function `nloptr`. The method of least squares was used in this fitting procedure by minimizing the equation $\sum_{i=1}^n (y_i - y_i^m)^2$ where y_i is the value of the data and y_i^m is the corresponding output of the model. The negative log-likelihood of each case was calculated and used to find the Akaike information criterion (AIC) with a correction for small sample sizes. These equations are as follows, where N is the number of data points used for fitting, RSS is the residual sum of squares determined from the optimization function, and k is the number of parameters in the model, including the standard deviation of the likelihood function.

$$AIC = (2 * -\log L) + (2 * k) + \frac{2 * k * (k + 1)}{N - k - 1}$$

$$-\log L = \frac{N}{2} [1 + \log(2\pi) + \log(\frac{RSS}{N})]$$

3. Results

3.1 Results from Literature

A similar study by Uys et al. (2015) has sought to determine the underlying cause of supersusceptibility. In this paper, a mathematical approach was used to determine the time-dependent rates of reinfection and progression to disease. However, using these rates to predict the annual risk of active disease due to reinfection produces an overestimate compared to the data used in the study (Uys et al. 2015). This could be attributable to the mathematical factor of assuming constant rates during the formulae derivation but fitting time-dependent rates, or it could be due to the assumption of a constant proportion of reinfection to relapse cases over time. The data from Marx et al. (2014) show that this proportion actually changes a great deal every year following treatment. Regardless, the goal of the present study is to utilize an alternative method to extend and improve this work.

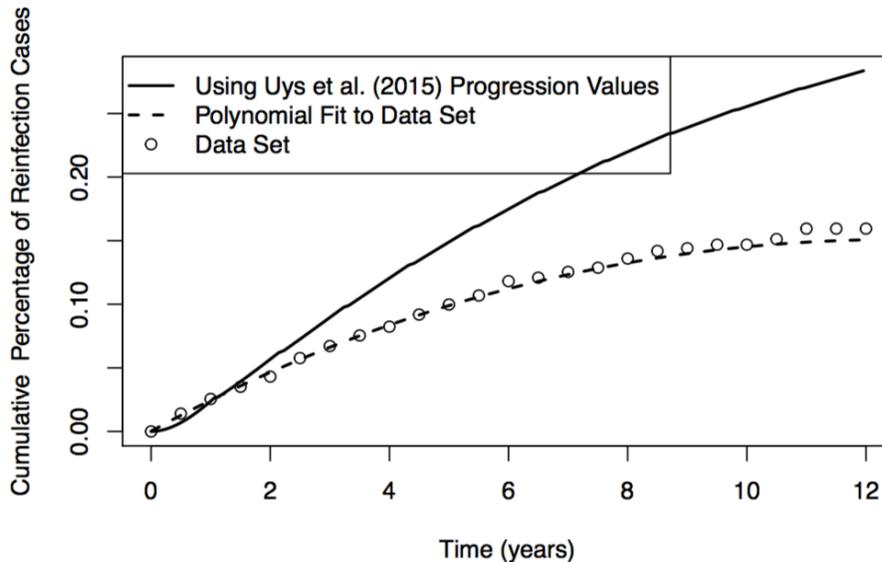


Figure 8: Cumulative percentage of reinfection cases in the population over time after treatment for the initial episode. Here we show the data values used in Uys et al. (2015) and the polynomial fit to these values. Additionally, we show the resulting curve if the progression and reinfection rates determined from the paper were used to model reinfection cases over time.

3.2 Determining Within-Host Parameters

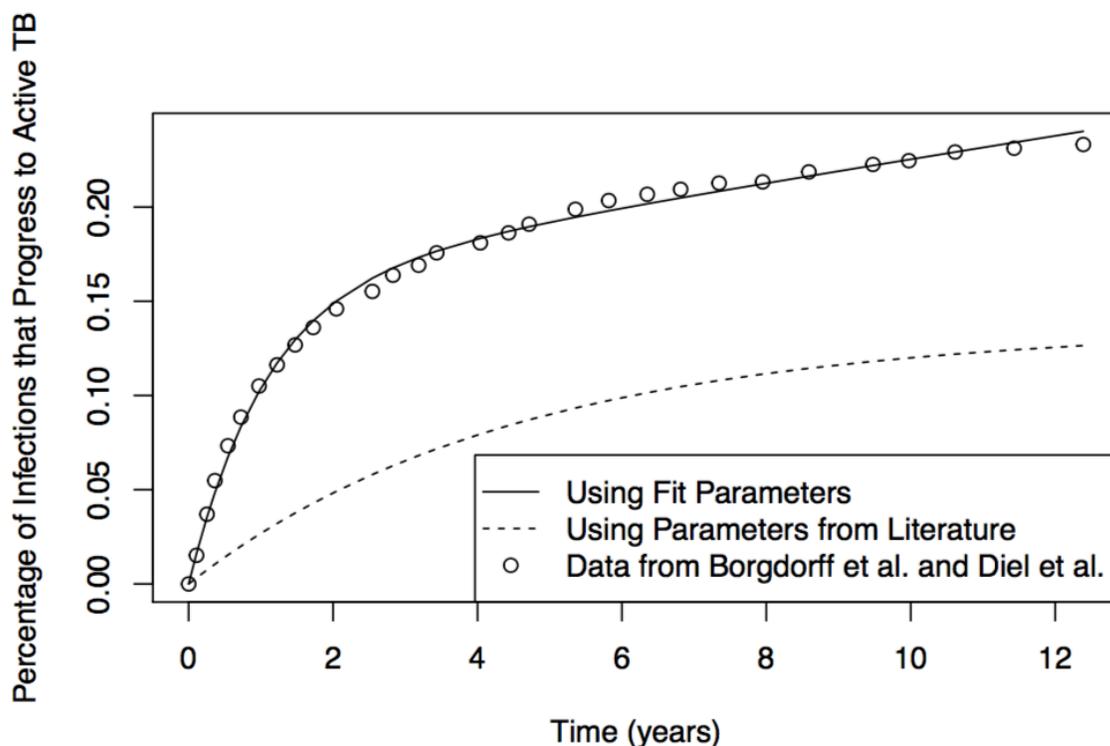


Figure 9: The percentage of infections that have progressed to active disease each year following initial infection. The data is extrapolated from Borgdorff et al. (2011) and Diel et al. (2008). The model representing the transition of individuals to active disease after reinfection (Figure 7) is fit to this data, and the resulting parameters are used for this study. This assumes that the parameters included in this model are constant between initial infection and reinfection. The resulting fit is used to plot the percentage of active cases of reinfection that should occur each year after reinfection. Additionally, parameters taken from the literature (Dowdy et al., 2013) are used to plot the results and the difference is shown.

The results after fitting the three within-host parameters to the adjusted Borgdorff et al. (2011) data (Figure 9) are very different from those given by the current literature (which are as follows: immune stabilization rate = 0.2 per year, rapid progression rate = 0.03 per year, reactivation rate = 0.0005 per

year) (D. W. Dowdy, Dye, and Cohen 2013). The immune stabilization rate is almost four times as high, correlating to an individual stabilizing after about 1.29 years rather than 5 years. The rapid progression rate is over 5 times as high, and the reactivation rate is 16 times as high. This could be attributable to the higher lifetime risk given by this data compared to the conventional estimates of 5-15%, which matches the lifetime risk attributable to the parameters from the literature. The high fit rapid progression rate may not be unreasonably high, however. Cobelens, et al. (2000) found that within 1 year of becoming infected with TB, approximately 17% of these individuals progressed to disease. Compared to the 15.6% risk of annual progression to disease for recently infected individuals (within 1.289 years) as determined with the fit, these parameters do not appear unreasonable.

When these parameter assumptions are relaxed and the rapid progression rate is allowed to be fit, there does not appear to be any substantial improvement of the fits, except in the case of the immune stabilization model (Figure 12, Table 11). In this case, there is an improvement in the fit of relapse cases but not that of reinfection cases. This may be due to the high reinfection rate in the entire population but low rapid progression rate, which would effectively reduce the pool of treated individuals so that relapse cases would plateau quickly (due to exhaustion of the source population). This would not impact reinfection cases significantly because the low progression rates would lead to a large latent population, and no new behavior in active cases would be seen.

The fact that fitting the rapid progression rate does not change the resulting fits lends credence to the parameterization used, which means that despite the high sensitivity of the models to changes in rapid progression rates (infection rate and rapid progression rates are by far the most sensitive parameters) (D. W. Dowdy, Dye, and Cohen 2013), the results in this study are not greatly dependent on this. Thus, the estimates used are not detrimental to this study. In fact, the fitted rapid progression rates calculated for the waning susceptibility model and host heterogeneity model without treatment variation are identical, 0.19 per year, which is also close to the estimated parameter from the Borgdorff et al. (2011) data, 0.16 per

year. Both of these are much larger than the 0.03 per year rate from the current literature (D. W. Dowdy, Dye, and Cohen 2013), which also lends credence to the unusually high rates that were estimated.

3.3 Model Discrimination

When comparing each model's ability to match the shape of the data, it appears as if the eventual plateau is the most difficult to aspect to match (Figure 10). No model adequately represents the reduction in new cases of both relapse and reinfection, but several models do match the shape of the relapse curve. Interestingly, this is not a matter of complexity, since the immune stabilization model does not show a plateau while the simple treatment variation model does. This is accomplished in the latter by allowing only a small number of individuals to relapse, with a fraction of 0.05 in the R_1 class (Table 2). In this way, the population is exhausted and no further relapses may occur. This is evidenced by the lack of any reduction in the corresponding reinfection incidence over time.

Both the waning susceptibility and host heterogeneity models fit equally well to the data (Table 10). The likelihoods associated with both of these hypotheses are much lower than the other models, as are the AICs despite the increase in parameters. However, these values are identical, so distinguishing between the validity of the two hypotheses is impossible given the current results. Both models also demonstrate the same shape when fit to the data (Figure 10). This may be indicative of the two models being treated similarly in the fitting process, meaning that despite the different modes of action behind the hypotheses, the compartmental models used to describe them act the same. The parameter values fit in both models do not demonstrate any identical behavior (Tables 6 and 8) however. The reinfection rate in the host heterogeneity model is even 3.5 times that of the waning susceptibility model, and the relapse rate is 3.25 times as high. This is balanced by the lower proportion of individuals that are in the high susceptibility group in the host heterogeneity model (0.29), while all treated individuals are in the high susceptibility group for about 2 years in the waning susceptibility model.

The same behavior is seen when allowing treatment variation (Figure 11). In fact, in the fits for these two hypotheses, the fraction of the population in the R_1 class was 1 (Tables 6 and 8), so the entire population was in this class. Despite any biological meaning behind this assumption, this did not correspond with any increased accuracy of the models. This is likely due to the nature of the compartmental model, which contains simplified representations of groups within the population. In this case, a single treated group could explain the entire population just as well as two treated groups since the relapse rate can be scaled according to the population, and a lower rate in the overall population is just as accurate as a higher rate for a smaller population. Interestingly, this indicates that more freedom between the reinfection and relapse rates (by separating the population depending on the likelihood of each) does not improve the fits. This means that drawing from the same treated population does not complicate the dynamics, which may be because of the size compared the number of people who actually experience secondary disease. If this proportion were larger, then the competition between the two would be more extreme, especially as time goes on and the population is exhausted. However, in this case the pool is sufficiently large to negate competition as much as two separate populations. The only case in which treatment variation improved the fit was with the immune stabilization model. However, the improvement was not present in reinfections (Figure 11). In the relapse cases, this addition actually allowed for the plateau. This is because, similarly to the case of the simple treatment variation model, a very small fraction of the population was in the R_1 class (Table 5), so this group could be exhausted quickly with relapse and cause a quick decrease in the apparent rate. The biological validity of this cannot be discussed without information on treatment outcomes, a likely invasive investigation.

In the case of the immune stabilization model, either fitting rapid progression or adding this new R_1 group are sufficient to allow the relapse cases to be fit fairly well (Figures 14 and 15). When either addition is used, both the AIC and likelihood is improved (Tables 10 and 11). However, there is only a marginal improvement in the reinfection cases fit when both additions are used, as shown by the difference in likelihood, but the addition of another parameter required for this actually worsens the AIC

(Figure 13, Tables 10 and 11). This also produces the most unusual parameter set, with an extremely high relapse and rapid progression rate as well as a very low reinfection rate (Table 5). The lack of significant improvement in the fit associated with this enormous change in parameter estimations indicates that these parameters are unlikely to be biologically valid.

Interestingly, as was the case without fitting the rapid progression rate, there does not appear to be significant improvement by implementing treatment variation in any models with the rapid progression rate fit (Figure 13, Table 11). However, this is the only case in which the host heterogeneity model and waning susceptibility model did not match in both shape and likelihood. The host heterogeneity model fits slightly worse in the relapse cases, reaching its plateau within the first year and not changing at all afterward. This indicates a small fraction in the R_1 class that is immediately exhausted, which is especially likely given the high rates of reinfection and relapse previously seen in the host heterogeneity model. Indeed, the fraction in the R_1 class is 0.458317. The difference is minor in the likelihood output when compared to the waning susceptibility model (Table 11), but this may indicate that, if given more freedom, these two models may differ more in their abilities to match the data.

The biggest improvement to the fit and the likelihood values of the waning susceptibility and host heterogeneity models occurs when the background reinfection rate is allowed to be fit (Figure 14, Table 12). By allowing this, both the reinfection and relapse plots follow the shape of the data. This is because the reinfection rate is set to zero for the low-susceptibility groups of these models, which means that there can be a limit on the long-term trend. In the previous cases, this was not possible because there was always at least a background risk of reinfection. This would mean that those in the low-susceptibility groups become immune to further tuberculosis, which seems biologically unlikely but interesting nonetheless.

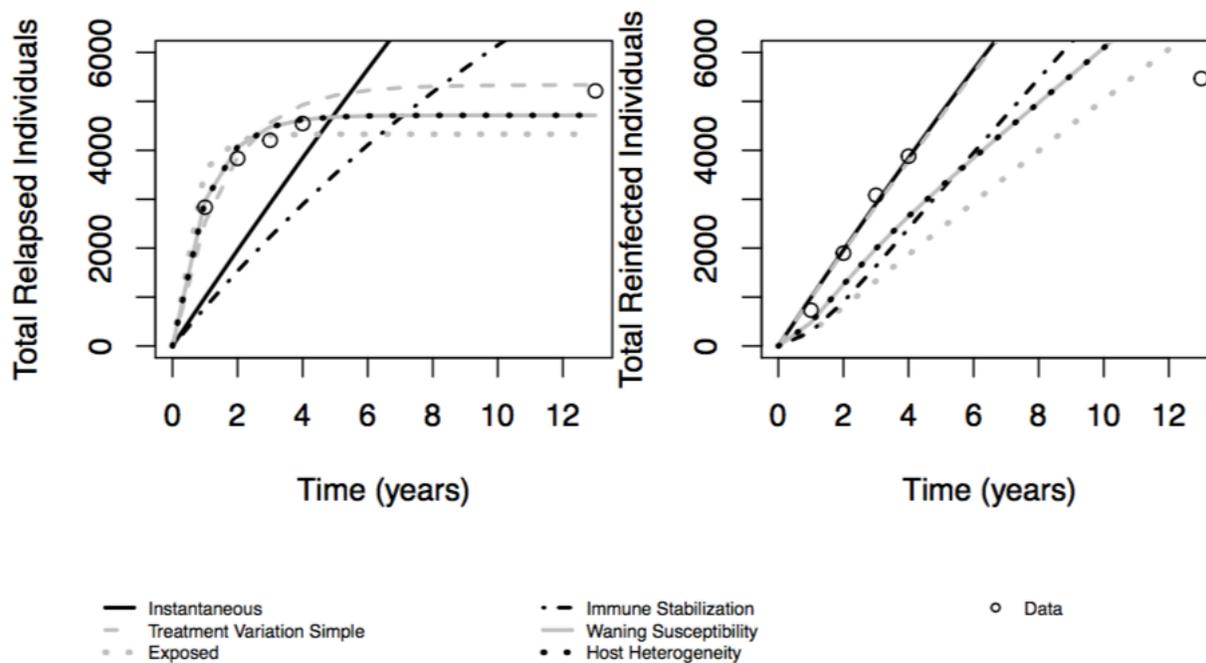


Figure 10: These two plots show the results of the fits for each proposed model. The cumulative number of relapses (right) and reinfections (left) for every year in a population of 100,000 treated individuals are shown based on the best fits of these models.

Table 10: The results of the fitting procedure for each model. The likelihood that the difference between the fitted results and the data is due to random chance is shown, as well as the AIC values, which takes the number of parameters used to reach such a likelihood into account. The bolded models show the best fits in terms of AIC.

<u>Model</u>	<u>Likelihood</u>	<u>AIC</u>	<u>Parameters</u>
Instantaneous	93.9	197.79	2
Treatment Variation Simple	89.98	195.96	3
Exposed Class	83.9	192.8	4
Immune Stabilization	90.1	190.21	2
Immune Stabilization Treatment Variation	84.35	184.7	3
Waning Susceptibility	82.7	181.4	3
Waning Susceptibility Treatment Variation	82.7	190.4	4
Host Heterogeneity	82.7	181.4	3
Host Heterogeneity Treatment Variation	82.7	190.4	4

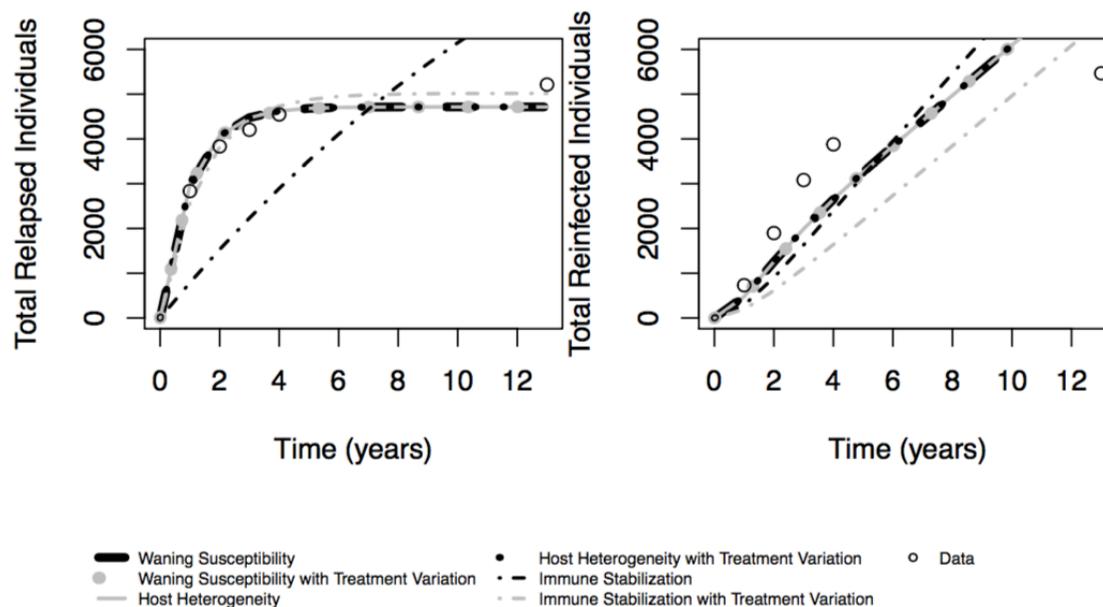


Figure 11: The benefit that adding treatment variation provides to each model in terms of ability to fit the data is shown. Each model is fit to the cumulative number of relapses (right) and reinfections (left) each year following treatment, in a population of 100,000 treated individuals.

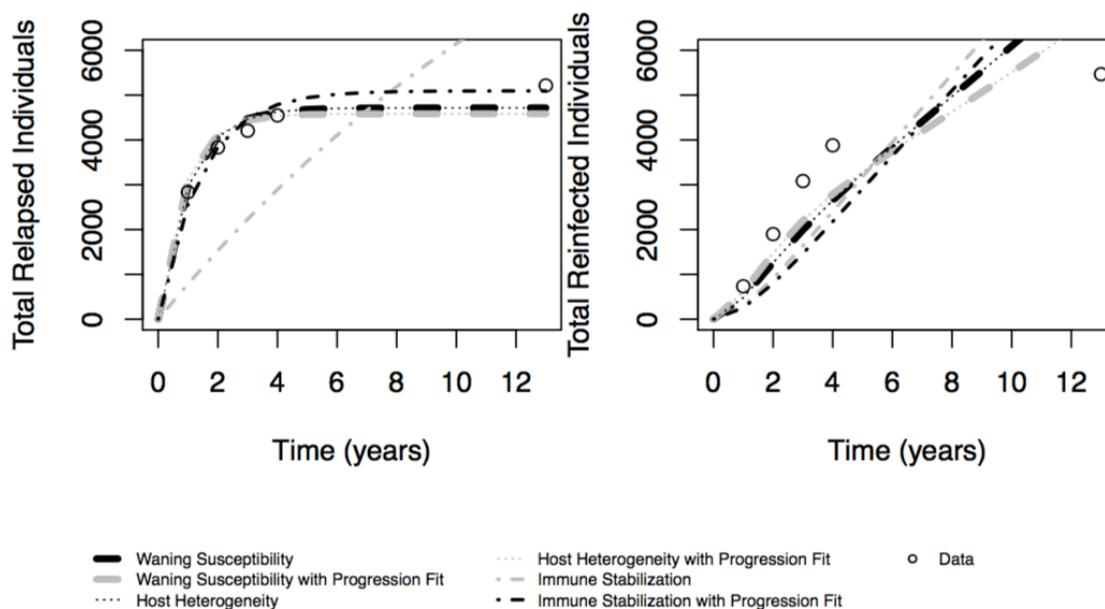


Figure 12: The assumptions made to the rapid progression rate were relaxed to determine whether or not the parameterization assumptions were valid. The ability for each model shown, with and without the

rapid progression rate fit, to match the observed data for the cumulative number of relapses (right) and reinfections (left) each year following treatment in a population of 100,000 people is shown.

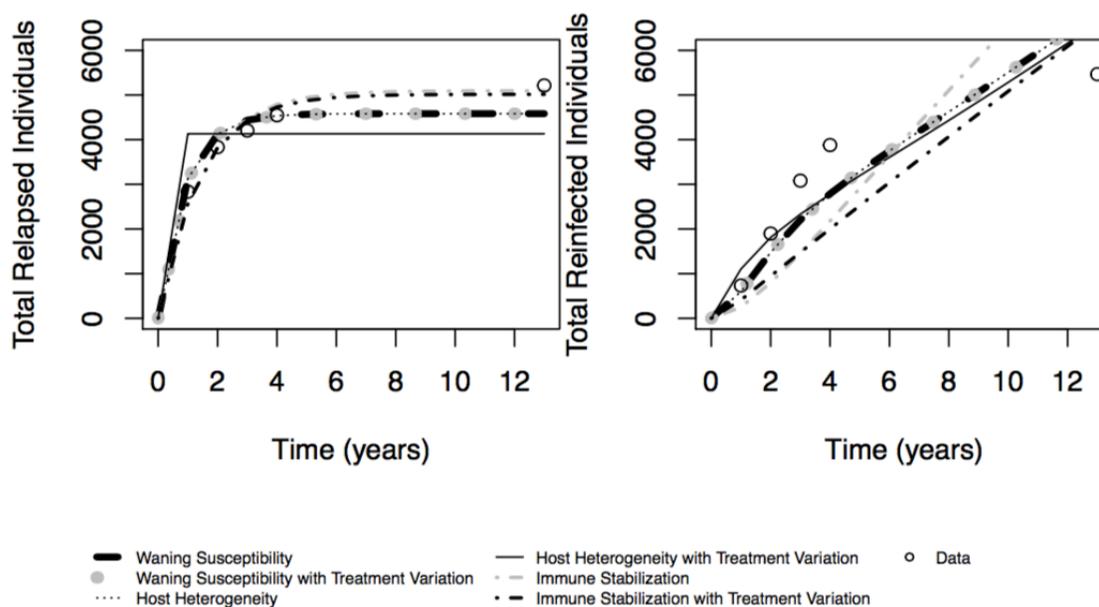


Figure 13: With rapid progression rate no longer fixed, these models are fit to the data on cumulative number of relapses (right) and reinfection (left) each year following treatment in a population of 100,000 treated individuals. The waning susceptibility, host heterogeneity, and immune stabilization models are compared to their counterparts with treatment variation to determine the effect.

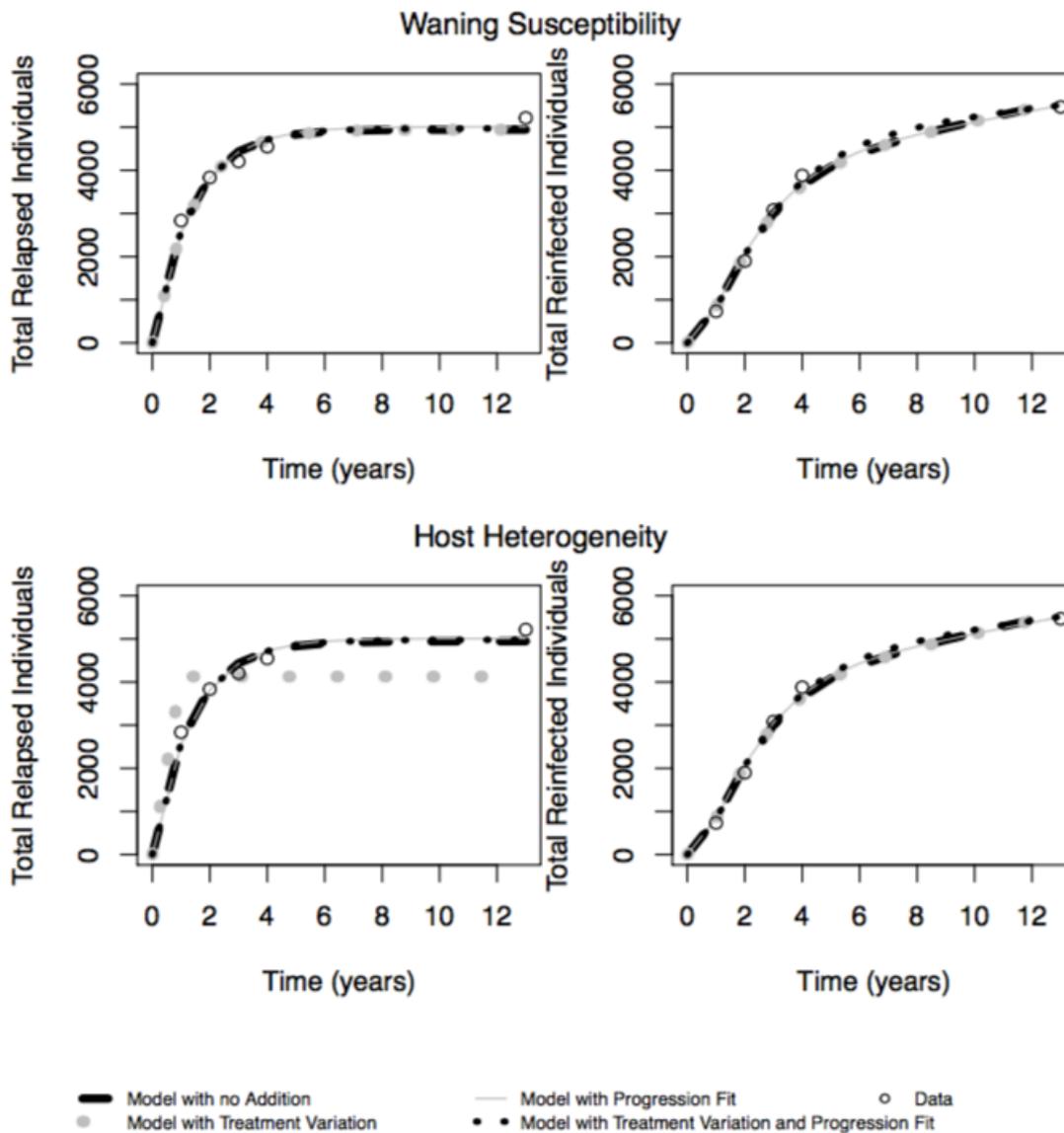


Figure 14: Allowing the background reinfection rate to be fit and its effect on the previous plots of host heterogeneity (below) and waning susceptibility (above).

Table 11: The results of the fitting procedure for each model with the rapid progression rate no longer fixed. The likelihood that the difference between the fitted results and the data is due to random chance is shown, as well as the corresponding AIC values. The bolded models show the best fits in terms of AIC, demonstrating the best balance of model complexity and accuracy.

<u>Model (With Progression Fit)</u>	<u>Likelihood</u>	<u>AIC</u>	<u>Parameters</u>
Immune Stabilization	86.04	188.07	2
Immune Stabilization Treatment Variation	82.5	190.01	3
Waning Susceptibility	79.54	184.07	3
Waning Susceptibility Treatment Variation	79.54	199.07	4
Host Heterogeneity	79.54	184.07	3
Host Heterogeneity Treatment Variation	80.99	201.98	4

Table 12: The results of the fitting procedure while allowing the background reinfection rate to be fit. The effect this has on the host heterogeneity and waning susceptibility models with and without treatment variation and the rapid progression rate fit is shown. These effects are quantified by the likelihood that the difference between the fitted results and the data is due to random chance and the AIC values that correspond this. The bolded models show the best fits in terms of AIC, demonstrating the best balance of model complexity and accuracy.

<u>Model (With Background Reinfection Rate Fit)</u>	<u>Likelihood</u>	<u>AIC</u>	<u>Parameters</u>	<u>Background Reinfection Rate</u>
Waning Susceptibility	65.77	156.54	4	0 per year
Waning Susceptibility- Treatment Variation	65.77	171.54	5	0 per year
Waning Susceptibility- Progression Fit	65.34	170.67	5	0 per year
Waning Susceptibility- Treatment Variation and Progression Fit	64.64	199.28	6	0 per year
Host Heterogeneity	65.77	156.54	4	0 per year
Host Heterogeneity- Treatment Variation	77.61	195.22	5	0 per year

Host Heterogeneity- Progression Fit	65.34	170.67	5	0 per year
Host Heterogeneity- Treatment Variation and Progression Fit	64.92	199.83	6	0 per year

4. Discussion

By comparing the ability for these models to fit the observed data of relapse and reinfection after treatment for TB, it is clear that both waning susceptibility and host heterogeneity can explain the temporal behavior of secondary disease equally well. While the simplifications made to the first few models could explain the inability for them to match the curve in cumulative recurrent disease cases, even the more complex and biologically accurate immune stabilization model cannot fit the data appropriately. This indicates that there is some unique behavior present in TB infections subsequent to the index case. At the global level, the absence of these behaviors in models is likely not impactful, but given the large spatial heterogeneity of tuberculosis cases, this can be incredibly impactful in some predictions made from models. With 64% of TB-related deaths coming from only 7 countries (“Fact Sheet on Tuberculosis” 2018), hyperendemic communities are vital to consider in modeling tuberculosis, and it is in these areas that recurrent disease is most likely, due to the high prevalence. Thus, current models are less useful in the areas where they most need to be used. Improved models must incorporate some unique behavior in post-treatment recurrent disease, and determining what this behavior is is a vital first step. In this case, both hypotheses tested are equally likely and distinguishing between each was impossible.

It is worth noting that all studies used in this investigation used sputum smears or cultures for determination of active tuberculosis. This means that the data used is likely an underestimate, since these

tests are not completely accurate (Pai et al. 2016, Siddiqi, Lambert, and Walley (2003)). Smear tests, for example, are about 50-60% accurate (Siddiqi, Lambert, and Walley 2003). The studies including culture tests do not have the same limitations, but still are not accurate for children who cannot produce sputum or for extrapulmonary TB if not specifically tested for it.

Additionally, this study is limited by the variety of different papers that must be used in order to attain data. This is a limitation created from necessity, as the tuberculosis data is difficult to find for these specific parameters and questions asked. In order to account for this, freedom in the parameter values was allowed whenever possible.

It is possible that both host heterogeneity and waning susceptibility impact population dynamics for tuberculosis. Many studies have indicated that genetic factors play a role in tuberculosis susceptibility (Abel et al. 2018, Puffer (1944), Kallmann and Reisner (1943), Cobat et al. (2012)). One study has shown that about 65% of TST variability can be explained with a codominant gene (Cobat et al. 2012). Additionally, familial studies (Puffer 1944) and twin studies (Kallmann and Reisner 1943) have shown evidence for host genetic factors playing a role in progression to active TB. Additionally, it has been found that the immunosuppression does occur in TB-infected individuals (Sahiratmadja et al. 2007, Hirsch et al. (1999)). One study showed that IFN- γ production is reduced in patients with active TB and this reduction is proportional to the severity of the disease, but this effect is reversed with cure (Sahiratmadja et al. 2007). Another study showed that this IFN- γ depression remained for at least 12 months (Hirsch et al. 1999). Both of these studies could support a form of waning susceptibility in the form of a reversible immunosuppression after active TB.

This current study was not comprehensive. There are many other possibilities for the behavior seen in recurrent disease and investigations into these may help with future efforts to improve tuberculosis modeling. Additionally, in the future, a model incorporating both host heterogeneity and waning susceptibility may help to determine the impact of both. Furthermore, more comprehensive

studies on the characteristics of those who experience recurrent TB episodes may indicate which factors (e.g. location, genes, immunosuppression, infection of a family member, etc.) put people at the most risk. Spatial heterogeneity is important in tuberculosis prevalence (“Fact Sheet on Tuberculosis” 2018), so efforts to alleviate the prevalence in areas with high recurrence may make the most impact. If this is accomplished, the elimination of TB may become much more likely, by focusing efforts where it matters most.

References

- Abel, L., J. Fellay, D. W. Haas, E. Schurr, G. Srikrishna, M. Urbanowski, N. Chaturvedi, S. Srinivasan, D. H. Johnson, and W. R. Bishai. 2018. “Genetics of Human Susceptibility to Active and Latent Tuberculosis: Present Knowledge and Future Perspectives.” *Lancet Infectious Diseases* 18 (3): E64–E75.
- Barry, 3rd, C. E., H. I. Boshoff, V. Dartois, T. Dick, S. Ehrt, J. Flynn, D. Schnappinger, R. J. Wilkinson, and D. Young. 2009. “The Spectrum of Latent Tuberculosis: Rethinking the Biology and Intervention Strategies.” *Nat Rev Microbiol* 7 (12): 845–55.
- Borgdorff, M. W., M. Sebek, R. B. Gesskus, K. Kremer, N. Kalisvaart, and D. van Soolingen. 2011. “The Incubation Period Distribution of Tuberculosis Estimated with a Molecular Epidemiological Approach.” *Int J Epidemiol* 40 (4): 964–70.
- Centers for Disease Control and Prevention. 2010. “Updated Guidelines for Using Interferon Gamma Release Assays to Detect Mycobacterium Tuberculosis Infection.” United States.
- Cobat, A., L. F. Barrera, H. Henao, P. Arbelaez, L. Abel, L. F. Garcia, E. Schurr, and A. Alcais. 2012. “Tuberculin Skin Test Reactivity Is Dependent on Host Genetic Background in Colombian Tuberculosis Household Contacts.” *Clinical Infectious Diseases* 54 (7): 968–71.

Cobelens, F. G., H. van Deutekom, I. W. Draayer-Jansen, A. C. Schepp-Beelen, P. J. van Gerven, R. P. van Kessel, and M. E. Mensen. 2000. "Risk of Infection with Mycobacterium Tuberculosis in Travellers to Areas of High Tuberculosis Endemicity." *Lancet* 356 (9228): 461–5.

Colditz, G. A., T. F. Brewer, C. S. Berkey, M. E. Wilson, E. Burdick, H. V. Fineberg, and F. Mosteller. 1994. "Efficacy of Bcg Vaccine in the Prevention of Tuberculosis. Meta-Analysis of the Published Literature." *JAMA* 271 (9): 698–702.

Diel, R., R. Loddenkemper, K. Meywald-Walter, S. Niemann, and A. Nienhaus. 2008. "Predictive Value of a Whole Blood Ifn-Gamma Assay for the Development of Active Tuberculosis Disease After Recent Infection with Mycobacterium Tuberculosis." *Am J Respir Crit Care Med* 177 (10): 1164–70.

Dowdy, D. W., C. Dye, and T. Cohen. 2013. "Data Needs for Evidence-Based Decisions: A Tuberculosis Modeler's 'Wish List'." *Int J Tuberc Lung Dis* 17 (7): 866–77.

Esmail, H., 3rd Barry C. E., D. B. Young, and R. J. Wilkinson. 2014. "The Ongoing Challenge of Latent Tuberculosis." *Philos Trans R Soc Lond B Biol Sci* 369 (1645): 20130437.

"Fact Sheet on Tuberculosis." 2018. *Wkly Epidemiol Rec* 93 (4-5): 39–43.

Fine, P. E. 1995. "Variation in Protection by Bcg: Implications of and for Heterologous Immunity." *Lancet* 346 (8986): 1339–45.

Fletcher, H. A., and L. Schrager. 2016. "TB Vaccine Development and the End Tb Strategy: Importance and Current Status." *Trans R Soc Trop Med Hyg* 110 (4): 212–8.

Graham, B. S., J. E. Ledgerwood, and G. J. Nabel. 2009. "Vaccine Development in the Twenty-First Century: Changing Paradigms for Elusive Viruses." *Clin Pharmacol Ther* 86 (3): 234–6.

- Hirsch, C. S., Z. Toossi, C. Othieno, J. L. Johnson, S. K. Schwander, S. Robertson, R. S. Wallis, et al. 1999. "Depressed T-Cell Interferon-Gamma Responses in Pulmonary Tuberculosis: Analysis of Underlying Mechanisms and Modulation with Therapy." *J Infect Dis* 180 (6): 2069–73.
- Horsburgh, Jr., C. R. 2004. "Priorities for the Treatment of Latent Tuberculosis Infection in the United States." *N Engl J Med* 350 (20): 2060–7.
- Horsburgh, Jr., C. R., M. O'Donnell, S. Chamblee, J. L. Moreland, J. Johnson, B. J. Marsh, M. Narita, L. S. Johnson, and C. F. von Reyn. 2010. "Revisiting Rates of Reactivation Tuberculosis: A Population-Based Approach." *Am J Respir Crit Care Med* 182 (3): 420–5.
- Kallmann, F. J., and D. Reisner. 1943. "Twin Studies on Genetic Variations in Resistance to Tuberculosis." *Journal of Heredity* 34 (10): 293–301.
- Kik, S. V., C. M. Denkinger, P. Chedore, and M. Pai. 2014. "Replacing Smear Microscopy for the Diagnosis of Tuberculosis: What Is the Market Potential?" *European Respiratory Journal* 43 (6): 1793–6.
- Marx, F. M., R. Dunbar, D. A. Enarson, B. G. Williams, R. M. Warren, G. D. van der Spuy, P. D. van Helden, and N. Beyers. 2014. "The Temporal Dynamics of Relapse and Reinfection Tuberculosis After Successful Treatment: A Retrospective Cohort Study." *Clin Infect Dis* 58 (12): 1676–83.
- Ozcaglar, C., A. Shabbeer, S. L. Vandenberg, B. Yener, and K. P. Bennett. 2012. "Epidemiological Models of Mycobacterium Tuberculosis Complex Infections." *Math Biosci* 236 (2): 77–96.
- Pai, M., M. A. Behr, D. Dowdy, K. Dheda, M. Divangahi, C. C. Boehme, A. Ginsberg, et al. 2016. "Tuberculosis." *Nat Rev Dis Primers* 2: 16076.
- Puffer, Ruth Rice. 1944. *Familial Susceptibility to Tuberculosis*. Cambridge, Mass., Harvard university press.

Sahiratmadja, E., B. Alisjahbana, T. de Boer, I. Adnan, A. Maya, H. Danusantoso, R. H. Nelwan, et al. 2007. "Dynamic Changes in Pro- and Anti-Inflammatory Cytokine Profiles and Gamma Interferon Receptor Signaling Integrity Correlate with Tuberculosis Disease Activity and Response to Curative Treatment." *Infect Immun* 75 (2): 820–9.

Siddiqi, K., M. L. Lambert, and J. Walley. 2003. "Clinical Diagnosis of Smear-Negative Pulmonary Tuberculosis in Low-Income Countries: The Current Evidence." *Lancet Infectious Diseases* 3 (5): 288–96.

Swaminathan, S., and G. Ramachandran. 2015. "Challenges in Childhood Tuberculosis." *Clinical Pharmacology & Therapeutics* 98 (3): 240–44.

Tameris, M. D., M. Hatherill, B. S. Landry, T. J. Scriba, M. A. Snowden, S. Lockhart, J. E. Shea, et al. 2013. "Safety and Efficacy of Mva85a, a New Tuberculosis Vaccine, in Infants Previously Vaccinated with Bcg: A Randomised, Placebo-Controlled Phase 2b Trial." *Lancet* 381 (9871): 1021–8.

"The End Tb Strategy: A Global Rally." 2014. *Lancet Respir Med* 2 (12): 943.

Uys, P., H. Brand, R. Warren, G. van der Spuy, E. G. Hoal, and P. D. van Helden. 2015. "The Risk of Tuberculosis Reinfection Soon After Cure of a First Disease Episode Is Extremely High in a Hyperendemic Community." *PLoS One* 10 (12): e0144487.

Verver, S., R. M. Warren, N. Beyers, M. Richardson, G. D. van der Spuy, M. W. Borgdorff, D. A. Enarson, M. A. Behr, and P. D. van Helden. 2005. "Rate of Reinfection Tuberculosis After Successful Treatment Is Higher Than Rate of New Tuberculosis." *Am J Respir Crit Care Med* 171 (12): 1430–5.

"Global tuberculosis report 2017". Geneva: World Health Organization; 2017, pp. 4 & 41. Licence: CC BY-NC-SA 3.0 IGO.