Modelling Tuberculosis in South Africa

by

Riley Lawson Wishart

B.Sc., University of Victoria, 2015

Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of
Master of Science

in the
Department of Mathematics
Faculty of Science

SIMON FRASER UNIVERSITY
Spring 2018

Approval

Name:	**Riley Lawson Wishart**
Degree:	**Master of Science (Mathematics)**
Title:	***Modelling Tuberculosis in South Africa***
Examining Committee:	**Chair:** Tom Archibald Professor

JF Williams
Senior Supervisor
Associate Professor

Alexander Rutherford
Co-Supervisor
Adjunct Professor

Paul Tupper
External Examiner
Professor

Date Defended: April 16, 2018

Abstract

Tuberculosis ranks alongside HIV/AIDS as one of the deadliest infectious diseases in the world, killing 1.4 million people in 2015. The World Health Organization and the Stop TB Partnership set targets to achieve a 90% reduction in incidence and a 95% reduction in mortality from the 2015 values by 2035.

We investigate the global dynamics of a compartmental system of ordinary differential equations that models tuberculosis. A more detailed model is then calibrated to the HIV negative TB endemic in South Africa and used to evaluate the 2035 reduction targets. Optimal care and control strategies for fixed budgets are also identified.

Model projections for South Africa show the mortality targets can be met through combined interventions, however due to relapse and latent progression the incidence targets are unrealistic. To minimize incidence in 20 years, funding should be prioritized into linking drug susceptible cases (over multi-drug resistant cases) onto care.

Keywords: Tuberculosis, South Africa, Compartmental Disease Models

Acknowledgements

First and foremost I would like to thank both my supervisors, Dr. JF Williams and Dr. Alexander Rutherford, for all of the guidance and opportunities that they have provided for me. I would also like to thank Brian Williams from SACEMA for the continued talks we've had throughout my research.

I am especially grateful for all of my family and friends who have supported me at SFU.

Table of Contents

List of Tables

List of Figures

Nomenclature

$\mathbb{I}$	The identity matrix.
$\rho(A)$	The spectral radius of the matrix A.
A^T	The transpose of the matrix A.
$\mathbb{R}^n$	The collection of vectors $(v_1, \ldots, v_n)$ where each v_i is taken from the real number line.
$\mathbb{R}^n_{>0}$	The collection of vectors $(v_1, \ldots, v_n)$ where each v_i is taken from the positive real number line ($v_i > 0$ for all i).
$\mathbb{R}^n_{\geq 0}$	The collection of vectors $(v_1, \ldots, v_n)$ where each v_i is taken from the non-negative real number line ($v_i \geq 0$ for all i).
$\vec{e}_j$	The j^{th} standard basis vector of $\mathbb{R}^n$ $\left((\vec{e}_j)_i = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}\right)$.
$\partial\Omega$	The boundary of the set Ω.
$dom(f)$	The domain of the function f.
$range(f)$	The range of the function f.
$\dot{f}$	The time derivative of the function f.
$B(\epsilon, \vec{x})$	The set $\left\{ \vec{y} \in \mathbb{R}^n \;\middle|\; \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2} < \epsilon \right\}$.
$A \subset B$	The set A is a proper subset of the set B.

Chapter 1

Introduction

1.1 The Biology and Development of Tuberculosis

Tuberculosis (TB) is an airborne infectious disease which is caused by *Mycobacterium tuberculosis*. If *M. tuberculosis* is inhaled into the lungs some of the infection will be eliminated by the innate immune system but a proportion will be able to survive and replicate. In response, the immune system will form granulomas around the surviving mycobacterium to contain the infection. As the granulomas break down the infection will spread and more granulomas will form to contain the outbreak.

At the early stages of infection the host is noninfectious and asymptomatic. With the continued effort to contain the infection the lung tissue is gradually destroyed. This destruction eventually causes the host to develop a persistent cough which will allow them to expel droplets containing *Mycobacterium tuberculosis* into the surrounding environment.

Progression of TB from the latent (noninfectious) phase into the active (infectious) phase is highly dependent on the state of the host's immune system. If the host can mount an effective immune response, TB may remain in the noninfectious stage for significant periods of time. In a healthy population 5-25% of infections will develop active TB within 5 years, while the remaining infections have a 10% chance of developing active TB over the rest of their lives. Individuals who develop a long term latent infection do not acquire a complete immunity against reinfection and as a result, the long term latent population may develop active TB through re-exposure to the disease.

The long latency period separates tuberculosis from other infectious diseases as it means a TB endemic will be able to persist within a population with low incidence rates. This long latency period also makes TB endemics very stable within populations. Through the use of phylogenetic tree analysis the *Mycobacterium tuberculosis* complex has been mapped

back to the expansion of humans from Africa. These phylogenetic trees have also linked the spread of Tuberculosis around the world with the development of trade and migration.

1.2 Treatment and the Current State of the Tuberculosis Endemic

Historically two of the most common methods used to diagnosis active TB have been sputum smear microscopy and sputum cultures. Although smear microscopy is faster, it is also a much less sensitive diagnostic test. This means if the patient produces a sputum sample which has a low prevalence of Mycobacterium, the smear microscopy may produce a negative result. As low prevalence infections are still considered to be infectious, individuals with active TB are classified as being sputum smear positive or sputum smear negative. The sputum smear negative infections are typically diagnosed through bacterial cultures. Recently molecular tests for TB have been developed. These diagnostic tests are much faster and yield much more sensitive results than the bacterial cultures.

The primary first line drugs (isoniazid and rifampicin) for the treatment of tuberculosis were introduced in the 1950s . Treatment of sputum smear positive tuberculosis with combined drug therapy lasts for six months and with proper drug adherence, success rates[1] of over 90% can be achieved. Without proper drug adherence patients may develop drug resistant tuberculosis. Treatment of multi-drug-resistant (MDR) TB[2] requires a combined regime of second line drugs[3] which have a higher toxicity. This, combined with the fact that MDR treatment can take 2 years, means the treatment success rates are often much lower.

Although tuberculosis has largely been eradicated in developed countries, TB still remains as a significant health crisis for a large part of the world. Tuberculosis ranks alongside of HIV/AIDs as one of the deadliest infectious diseases in the world, killing 1.4 million people in 2015. Globally in 2015 there was an estimated 10.4 million incidence cases of TB, with 87% of the cases occurring in the 30 highest TB burden counties as designated by the World Health Organization [15, Table 2.2. Page 13]. The WHO estimated that 580,000 of these incidence cases were either multi-drug resistant (83%) or rifampicin-resistant (17%).

[1] A patient is said to have a successful treatment if no mycobacterium is detectable by sputum smear microscopy at the end of the treatment program.

[2] MDR TB is defined to be resistance to both isoniazid and rifampicin.

[3] Some of the second line drugs include: aminoglycosides, polypeptides, fluoroquinolones, and thioamides.

1.3 Thesis Outline

Chapter 2 is focused on compartmental systems of ordinary differential equations. In Section 2.1, the existing theory for compartmental disease models is reviewed and an explanation is provided on how this theory can be applied to tuberculosis models that include secondary reinfections. In Section 2.2, a toy model for tuberculosis is constructed and global stability results are derived in the physically relevant region for different scenarios. In Section 2.3, we introduce the Tuberculosis Incidence and Mortality Estimates (TIME) model [8] which is used by the Spectrum modelling suite to help make disease burden predictions. In the final section of this chapter the toy model and the TIME model are compared.

In Chapter 3 the TIME model is used to numerically investigate the HIV negative TB endemic in South Africa. In Section 3.1 the relation between HIV/AIDs and TB is discussed. The death rates for TB are calculated by numerically solving a pair of algebraic equations. The remaining 13 unknown parameters in the TIME model are then calibrated against 17 data points from the 2012 HIV negative TB endemic in South Africa. In Section 3.2 the calibrated model is used to investigate incidence and mortality reduction strategies for South Africa.

Chapter 4 investigates the optimal control strategies for reducing the HIV negative TB incidence in South Africa over a 20 year time period at a fixed budget. This investigation shows that treatment resources should be directed into linked non-MDR cases onto care. Once 100% of non-MDR cases are being linked onto care, then funding should be directed into increasing the diagnosis rate.

Chapter 2

Compartmental Models of Tuberculosis

2.1 Homogenous Compartmental Disease Models

Homogeneous compartmental disease models use systems of differential equations to model the spread and control of infectious diseases. To model the dynamics of a particular disease, the total population is first partitioned into compartments, which are assumed to be homogenous. Once the total population has been partitioned, the growth of the subpopulations is determined by the average flows between compartments.

One of the most important concepts in epidemiology is the basic reproduction number (R_0). This is a threshold parameter which determines whether or not a susceptible population is at risk of a disease outbreak. If a single infection is introduced into a susceptible population when $R_0 < 1$, then the disease will not be able to sustain itself and the initial infection will die off. On the other hand, if the single infection is introduced when $R_0 > 1$, then the disease will be able to sustain itself and the infection will spread throughout the population.

The basic reproduction number is defined to be the expected number of secondary infections a single infected individual will cause in a completely susceptible population over the entire infectious period. *Diekmann et al.* [3] showed that for infectious disease models in heterogenous populations, the basic reproduction number corresponds with the spectral radius of a positive linear operator.

Driessche and Watmough [16] developed a general structure for a large class of homogenous compartmental disease models and derived a closed form expression for the positive linear operator introduced in [3]. *Driessche and Watmough* proved the basic reproduction number defines a threshold parameter for the local stability of the disease free equilibrium. In this

Section we will review the class of homogenous compartmental disease models introduced in [16] and explain how relaxing some of the structural assumptions allow for the theory to be applied to a wider class of disease models.

In the disease models considered by *Driessche and Watmough*, each compartment is classified as being a susceptible or an infected state. By definition individuals in the susceptible states may only move into an infected state through transmission of the disease. Given an n state disease model with m infected states, then the compartmental model takes the form

$$\begin{aligned} \frac{dx_i}{dt} &= F_i(\vec{x}) - V_i(\vec{x}) \quad \text{for } \ i = 1, 2, \ldots, m \\ \frac{dx_i}{dt} &= g_i(\vec{x}) \qquad\qquad\quad \text{for } \ i = m+1, m+2, \ldots, n \end{aligned} \tag{2.1}$$

where $F_i(\vec{x})$ is the rate of disease transmissions entering/exiting the i^{th} compartment, $V_i(\vec{x})$ is the rate of disease transitions entering/exiting the i^{th} compartment and $g_i(\vec{x})$ describes the dynamics of the i^{th} compartment. We are specifically interested in studying compartmental disease models in which $\vec{F}$, $\vec{V}$, and $\vec{g}$ satisfy the following list of conditions:

1) The region $\{0\}^m \times \mathbb{R}^{n-m}$ is invariant with respect to (2.1).

2) The system of equations (2.1) has a positive equilibrium point which is globally stable within $\{0\}^m \times \mathbb{R}^{n-m}$.

3) $\vec{V}(\vec{x}) = \begin{bmatrix} \mathcal{V} & 0 \end{bmatrix} \vec{x}$ where:

   (a) $\mathcal{V}$ is an $m \times m$ matrix which can be rewritten as $\mathcal{V} = s\mathbb{I} - A$ for some nonnegative matrix A and some $s > 0$.

   (b) For every $j \in \{1, 2, \ldots, m\}$, $\sum_{i=1}^{n} \mathcal{V}_{ij} > 0$.

4) If $x_i = 0$, then there exists some ϵ-neighbourhood of $\vec{x}$ such that $F_i(\vec{y}) \geq 0$ for all $y \in B(\epsilon, \vec{x}) \cap \mathbb{R}^n_{\geq 0}$.

5) If $x_i = 0$ for every $1 \leq i \leq m$, then $F_i(\vec{x}) = 0$.

6) For each $i \in \{m+1, \ldots, n\}$, if $x_i = 0$ then $g_i(\vec{x}) \geq 0$.

The first two conditions guarantee any disease free population will remain disease free for all time $t > 0$ and eventually converge to a steady state solution. This steady state is referred to as the disease free equilibrium.

The third condition describes how individuals transition through the various infectious states. Biologically condition 3a) implies that the transitions out of a given state only depend on the current state of the individual. Condition 3b) guarantees that in the absence of transmission ($\vec{F} = 0$) the net outflow from the infected compartments is always positive

and as a result transmission is required to sustain an infected population.

In the general disease model presented by *Driessche and Watmough* [16] they made the assumption that for every $1 \leq i \leq m$, $F_i(\vec{x}) \geq 0$ for all $\vec{x} \in \mathbb{R}^n_{\geq 0}$. Their justification was that the function $\vec{F}(\vec{x})$ represents an inflow of infections coming from the susceptible compartments. This assumption holds true for a large class of disease models, however it does not hold for models of tuberculosis which include bilinear infection terms between infected states. As *Driessche and Watmough* only used the positivity condition in a neighbourhood of the disease free equilibrium, the established results still hold for models which only satisfy condition 4).

Combining conditions 3), 4), and 6) guarantees that for each $i \in \{1, 2, \ldots, n\}$, if $x_i = 0$ then $\dot{x}_i \geq 0$. As a result the cone $\mathbb{R}^n_{\geq 0}$ is a forward invariant region of (2.1). Physically this means that any nonnegative solution will remain nonnegative for all time $t \geq 0$.

2.1.1 The Next Generation Matrix

Diekmann et al. [3] introduced the next generation operator to calculate the expected number of secondary infections produced by introducing an initial distribution of individuals throughout the infectious state space.

As homogenous compartmental disease models have a finite state space the next generation operator corresponds with a finite dimensional matrix Θ. The $(i,j)^{th}$ entry of Θ is defined to be the expected number of secondary infections produced in compartment i by an index case introduced in compartment j over the entire period of infectiousness[1]. Given an initial population of people $\vec{y}_{gen}$ distributed over the infected compartments then the expected number of secondary infections introduced into the next generation is

$$\vec{y}_{gen+1} = \Theta \vec{y}_{gen}.$$

In this framework the local stability of the disease free equilibrium is determined by the spectral radius of Θ.

To find the next generation matrix for the homogenous compartmental disease models discussed above, (2.1) is linearized around the disease free equilibrium. As individuals in the susceptible compartments may only move into the infected compartments through bilinear infection terms, the linearization gives

$$\frac{d\vec{x}}{dt} = \begin{pmatrix} \mathcal{F} - \mathcal{V} & 0 \\ J_3 & J_4 \end{pmatrix} \vec{x}, \tag{2.2}$$

[1]This measures how many people an index case infects before they transition out of the infected states through either treatment or death.

where: $\mathcal{F}$ is a nonnegative matrix which describes the inflow of disease transmissions from the susceptible states, $\mathcal{V}$ is the matrix defined in conditions $3a$ and $3b$ which describes the transitionary flows between the infectious states, and the matrices J_3 and J_4 describe the growth of the susceptible compartments.

From the block diagonal form of (2.2) it follows that the growth of the infected compartments can be approximated near the disease free equilibrium by the $m \times m$ linear system:

$$\frac{d\vec{y}}{dt} = (\mathcal{F} - \mathcal{V})\vec{y}. \tag{2.3}$$

As the $(i,j)^{th}$ entry of $\mathcal{F}$ is the rate of secondary infections produced in compartment i by an index case in compartment j, the $(i,j)^{th}$ entry of the next generation matrix is

$$\Theta_{ij} = \sum_{k=1}^{m} \mathcal{F}_{ik} \cdot \vartheta_{kj} \tag{2.4}$$

where ϑ_{kj} is defined to be the mean time an index case introduced into compartment j spends in compartment k over the course of the infection.

From the assumptions $3a$) and $3b$) made on the transition rates, the only positive elements of the matrix $\mathcal{V}$ are the diagonal elements. This combined with the fact that $\mathcal{V}$ has strictly positive column sums, means that the matrix

$$Q = \begin{pmatrix} -\mathcal{V}^T & \mathcal{V}^T \cdot \vec{1} \\ \vec{0}^T & 0 \end{pmatrix}$$

defines the transition rate matrix[2] for a continuous time Markov chain $\{X(t), t \geq 0\}$ which has the finite state space $\{1, 2, 3, \ldots, m+1\}$. This Markov chain describes how a single individual transitions through the various infected states. The $(m+1)^{th}$ state is an absorbing state that tracks individuals who have either died or recovered and are no longer in the infected state space. We can now use this Markov chain to calculate ϑ_{kj}. By defining the transition probability functions

$$P_{jk}(t) = P\{X(s+t) = k | X(s) = j\}$$

the forward Kolmogorov equations give us that for all (j,k) pairs

$$P'_{jk}(t) = \sum_{h=1}^{m+1} q_{jh} P_{hk}.$$

[2]For each $i \in \{1, 2, \ldots, m+1\}$, q_{ii} defines the rate that an individual transitions out of state i; and for each pair (i,j) with $i \neq j$, q_{ij} defines the rate at which an individual in state i transitions into state j.

Consequently the row vector $\vec{P}_j^{\,T}(t) = (P_{j1}(t), \ldots, P_{j\,m+1}(t))$, will have the solution

$$\vec{P}_j^{\,T}(t) = \vec{e}_i^{\,T} e^{tQ}$$

where $\vec{e}_j$ is the j^{th} standard basis vector. Now the mean total time an index case introduced into the j^{th} compartment spends in the k^{th} compartment $(1 \leq k \leq m)$ is

$$\begin{aligned}\vartheta_{kj} &= \int_0^{+\infty} P\{X(s) = k | X(0) = j\} ds \\ &= \left(\int_0^{+\infty} (e^{Qs})^T \vec{e}_j ds \right)_k = \left(\int_0^{+\infty} e^{-\mathcal{V}s} ds \right)_{kj} = (\mathcal{V}^{-1})_{kj}.\end{aligned}$$

Assumptions 3a) and 3b) made on the transition matrix guarantee that $\mathcal{V}$ is a nonsingular M-matrix (these matrices are discussed in the next subsection). As a result all eigenvalues of $\mathcal{V}$ have a positive real part and $\mathcal{V}$ has a nonnegative inverse. This guarantees that the last integral is convergent and that the next generation matrix $\Theta = \mathcal{F}\mathcal{V}^{-1}$ is a positive linear operator.

2.1.2 Nonsingular M-matrices

If a matrix B can be rewritten as $B = s\mathbb{I} - A$ for some nonnegative matrix A, then B is said to be a Z-sign pattern matrix. If this can be done for some s and A where s is strictly larger than the spectral radius of A, then B is said to be a nonsingular $M-$matrix. Nonsingular M-matrices have a wide area of applications and there are 50 equivalent definitions for a nonsingular M-matrix [1, Chapter 6].

Lemma 2.1.1. *Given B is a Z-sign pattern matrix,* $\det(B)$ *will be positive if B has positive row sums; moreover if B is a Z-sign pattern matrix then the conditions i), ii), and iii) all all equivalent to the statement "B is a nonsingular M-matrix".*

i) All principle minors of B are positive.

ii) Every eigenvalue of B has a positive real part.

iii) B has a nonnegative inverse.

From conditions (3a) and (3b), $\mathcal{V}^T$ is a Z-sign pattern matrices which has positive row sums. Consequently every principle submatrix of $\mathcal{V}^T$ also has this property. This means all principle minors of $\mathcal{V}^T$ are positive and as a result $\mathcal{V}$ is a nonsingular M-matrix.

Nonsingular M-matrices are central to *Driessche and Watmough*'s stability proof of the disease free equilibrium in [16]. They proved that if $\mathcal{F}$ is a nonnegative matrix and $\mathcal{V}$ is a nonsingular M-matrix then the linear system $\dot{\vec{y}} = (\mathcal{F} - \mathcal{V})\vec{y}$ will be locally stable if and

only if $\rho(\mathcal{F}\mathcal{V}^{-1}) < 1$. Since this stability result is local to the disease free equilibrium some of the assumptions made on the structure of the disease transmissions could be relaxed to included models with bilinear infection terms between infected states.

2.2 The Sputum Smear Model

The sputum smear TB model is a compartmental disease model that we constructed to understand the dynamics of a Tuberculosis endemic. This model divides the total population into a susceptible compartment S, a latent compartment L, and two active TB compartments: the sputum smear positive (SSpos) compartment I_p, and the sputum smear negative (SSneg) compartment I_n.

Historically two of the most common methods used to diagnosis pulmonary TB have been sputum smear microscopy and sputum cultures. Although the smear microscopy is a much faster test, testing the sputum cultures yields much more sensitive results and consequently individuals with a negative sputum smear may still be culture positive. This splits the active TB compartment into a sputum smear positive and a sputum smear negative population. As the survival rates, infection rates, and diagnosis rates all differ between these two populations; they each play a different role in the disease dynamics.

This model assumes that $\sigma \cdot 100$ percent of all flows entering the active TB compartments will initially develop SSpos TB while the other $(1-\sigma) \cdot 100$ percent initially develops SSneg TB. As SSneg individuals may develop sputum smear positive TB later on, this model assumes individuals transition from the SSneg compartment to the SSpos compartment at the rate θ. As diagnosis rates depend on smear status, we also assume the SSpos and SSneg individuals are treated at the respective rates of γ_p and γ_n. In this model treatment is assumed not to provide a full cure to the disease. Consequently all treated individuals move back into the latent compartment once they are treated.

In the sputum smear model we assume that TB in transmitted by SSpos and SSneg individuals with a standard incidence force of infection rate. This means we assume that on average an SSpos individual will make β sufficient contacts[3] per unit time. We also assume the SSneg individuals will on average make $c\beta$ sufficient contacts per unit time. Since β sufficient contacts are made by each SSpos individual, on average the SSpos population will generate $\beta I_p \cdot \frac{S}{N}$ new infection from the susceptible population per unit time where N is defined to be the total size of the population.

To capture the high variability in the latent progression rate $\alpha \cdot 100$ percent of all infections

[3] to transmit the disease

proceed directly into the Active TB compartments, while the other $(1-\alpha)\cdot 100$ percent of infection proceed into the latent compartment. Individuals with latent tuberculosis are assumed to have only acquired a partial immunity against reinfection and as a result they may undergo a secondary infection event (exogenous reinfection). The degree of protection offered by a latent infection is defined to be $1-x$ where $x \in [0,1]$. Latent individuals may also transition into the active TB compartments through latent progression at the rate v.

This model assumes no vertical transmission TB so all births $\Lambda(t)$ are born into the susceptible compartment. The natural death rate for all individuals in the model is δ. Individuals with SSpos and SSneg TB die at the respect rates $\delta+\delta_p$ and $\delta+\delta_n$, where δ_p and δ_n are the death rates due to SSpos and SSneg TB.

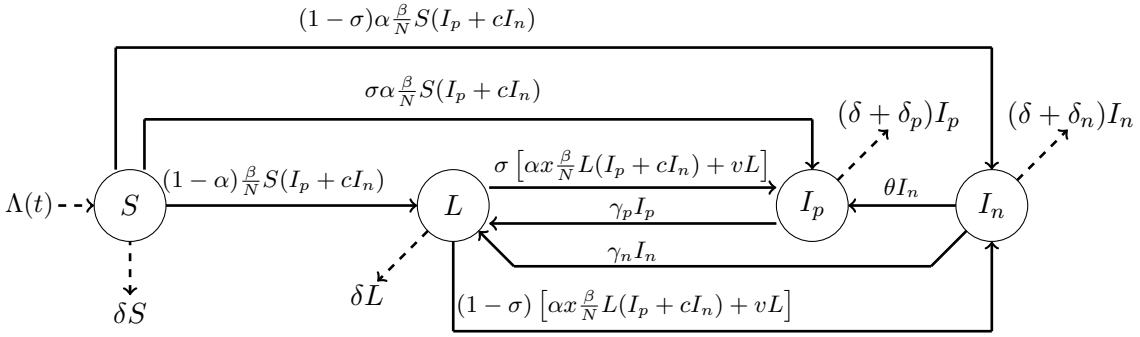


Figure 2.1: The flow diagram for the SLI_pI_n model

Parameter	Definition
β	Infectivity of SSpos individuals (infections per person per year)
c	Relative infectiousness of SSneg individuals
$1-x$	Degree of protection provided by a priori infection
v	Latent progression rate (progressions per person per year)
α	Proportion of infections that immediately develop active TB
σ	Proportion of people entering the active TB states who initially develop SSpos TB
θ	Sputum smear conversion rate (conversions per person per year)
γ_p	SSpos treatment rate (per person per year)
γ_n	SSneg treatment rate (per person per year)
δ	Natural death rate (per person per year)
δ_p	SSpos death rate (per person per year)
δ_n	SSneg death rate (per person per year)
$\Lambda(t)$	Birth Rate

Table 2.1: The SLI_pI_n model parameters

From Figure 2.1, the SLI_pI_n equations are:

$$\begin{cases} \dot{S} = \Lambda(t) - \frac{\beta}{N} S(I_p + cI_n) - \delta S \\ \\ \dot{L} = (1-\alpha)\frac{\beta}{N} S(I_p + cI_n) - \alpha x \frac{\beta}{N}(I_p + cI_n)L + (\gamma_p I_p + \gamma_n I_n) - (v+\delta)L \\ \\ \dot{I}_p = \sigma\left[\alpha \frac{\beta}{N} S(I_p + cI_n) + \alpha x \frac{\beta}{N}(I_p + cI_n)L + vL\right] + \theta I_n - (\gamma_p + \delta + \delta_p) I_p \\ \\ \dot{I}_n = (1-\sigma)\left[\alpha \frac{\beta}{N} S(I_p + cI_n) + \alpha x \frac{\beta}{N}(I_p + cI_n)L + vL\right] - (\gamma_n + \theta + \delta + \delta_n) I_n \end{cases} \tag{2.5}$$

By making the assumption that at time $t \geq 0$

$$\Lambda(t) = \delta S + \delta L + (\delta + \delta_p) I_p + (\delta + \delta_n) I_n,$$

then the total population ($S + L + I_p + I_n = N$) is constant. Under this assumption the susceptible compartment can be eliminated and the SLI_pI_n equations reduce to the three-dimensional system:

$$\begin{cases} \dot{L} = (1-\alpha)\frac{\beta}{N}[N - L - I_p - I_n](I_p + cI_n) - \alpha x \frac{\beta}{N}(I_p + cI_n)L + (\gamma_p I_p + \gamma_n I_n) - (v+\delta)L \\ \\ \dot{I}_p = \sigma\left[\alpha \frac{\beta}{N}[N - (1-x)L - I_p - I_n](I_p + cI_n) + vL\right] + \theta I_n - (\gamma_p + \delta + \delta_p) I_p \\ \\ \dot{I}_n = (1-\sigma)\left[\alpha \frac{\beta}{N}[N - (1-x)L - I_p - I_n](I_p + cI_n) + vL\right] - (\gamma_n + \theta + \delta + \delta_n) I_n. \end{cases} \tag{2.6}$$

2.2.1 The Basic Reproduction Number

The basic reproduction number for the reduced SLI_pI_n equations can be calculated from following the methods outlined by *Driessche and Watmough* in [16]; however as the nonlinear flows entering the active TB compartments in the reduced SLI_pI_n model are proportionally split between the SSpos and SSneg compartments, the calculation of the next generation from (2.6) is unnecessary. As

$$\begin{cases} \dot{I}_p = 0 \\ \dot{I}_n = 0 \end{cases} \quad \Longrightarrow \quad I_n = \frac{(1-\sigma)(\gamma_p + \delta + \delta_p)}{\theta + \sigma(\gamma_n + \delta + \delta_n)} I_p$$

by defining the constant

$$g = \frac{(1-\sigma)(\gamma_p + \delta + \delta_p)}{\theta + \sigma(\gamma_n + \delta + \delta_n)},$$

(L, I_p, gI_p) will be an equilibrium point of the reduced SLI_pI_n equations if and only if (L, I_p) is an equilibrium point of the $\mathbb{R}^2$ dynamical system

$$\begin{cases} \dot{L} = (1-\alpha)\frac{\beta}{N}[N - L - (1+g)I_p](1+cg)I_p - \alpha x \frac{\beta}{N}(1+cg)I_p L + (\gamma_p + g\gamma_n)I_p - (v+\delta)L \\ \\ \dot{I}_p = \sigma\left[\alpha \frac{\beta}{N}[N - (1-x)L - (1+g)I_p](1+cg)I_p + vL\right] + \theta g I_p - (\gamma_p + \delta + \delta_p) I_p. \end{cases} \tag{2.7}$$

This means (2.6) and (2.7) undergo the transcritical bifurcation at $R_0 = 1$ together, and as a result the basic reproduction number of the reduced SLI_pI_n equations can be calculated from (2.7).

Linearizing (2.7) at the origin gives $J = \mathcal{F} - \mathcal{V}$ where

$$\mathcal{F} = \begin{pmatrix} 0 & (1-\alpha)(1+cg)\beta \\ 0 & \alpha\sigma(1+cg)\beta \end{pmatrix} \quad \text{and} \quad \mathcal{V} = \begin{pmatrix} (v+\delta) & -(\gamma_p + g\gamma_n) \\ -\sigma v & (\gamma_p + \delta + \delta_p) - \theta g \end{pmatrix}.$$

Since $\mathcal{V}$ is a Z-sign pattern matrix (reference subsection 2.1.2), if all eigenvalues of $\mathcal{V}$ have a positive real part then $\mathcal{V}$ will be a nonsingular M-matrix. As a result the origin would be locally stable with respect to (2.7) if and only if $\rho(\mathcal{F}\mathcal{V}^{-1}) < 1$.

Applying the Routh-Hurwitz stability criterion to the polynomial $p(\lambda) = \det(\mathcal{V} + \lambda\mathbb{I})$ gives us that all roots of $p(\lambda)$ must have a negative real part as

$$\begin{aligned} \mathrm{Tr}(\mathcal{V}) &= (v+\delta) + (\gamma_p + \delta + \delta_p) - \theta\frac{(1-\sigma)(\gamma_p + \delta + \delta_p)}{\theta + \sigma(\gamma_n + \delta + \delta_n)} \\ &= (v+\delta) + (\gamma_p + \delta + \delta_p)\left[1 - \frac{\theta}{\theta + \sigma(\gamma_n + \delta + \delta_n)}(1-\sigma)\right] \\ &> 0 \\ \text{and} \quad \det(\mathcal{V}) &= (v+\delta)[(\gamma_p + \delta + \delta_p) - \theta g] - \sigma v(\gamma_p + g\gamma_n) \\ &= (v+\delta)\left\{(\gamma_p + \delta + \delta_p) - \left(\theta + \sigma\frac{v}{v+\delta}\gamma_n\right)g - \sigma\frac{v}{v+\delta}\gamma_p\right\} \\ &= (v+\delta)(\gamma_p + \delta + \delta_p)\left\{1 - \left[(1-\sigma)\frac{\theta + \sigma\frac{v}{v+\delta}\gamma_n}{\sigma(\gamma_n + \delta + \delta_n) + \theta} + \sigma\frac{v}{v+\delta}\frac{\gamma_p}{\gamma_p + \delta + \delta_p}\right]\right\} \\ &> 0. \end{aligned}$$

This means all eigenvalues of $\mathcal{V}$ have a positive real part and as a result R_0 can be determined by calculating the spectral radius of $\mathcal{F}\mathcal{V}^{-1}$. Since

$$\begin{aligned} \mathcal{F}\mathcal{V}^{-1} &= \begin{pmatrix} 0 & (1-\alpha)(1+cg)\beta \\ 0 & \alpha\sigma(1+cg)\beta \end{pmatrix} \cdot \frac{1}{\det(\mathcal{V})}\begin{pmatrix} (\gamma_p + \delta + \delta_p) - \theta g & \gamma_p + g\gamma_n \\ \sigma v & (v+\delta) \end{pmatrix} \\ &= \frac{1}{(v+\delta)[(\gamma_p + \delta + \delta_p) - \theta g] - \sigma v(\gamma_p + g\gamma_n)}\begin{pmatrix} \sigma v(1-\alpha)(1+cg)\beta & (1-\alpha)(1+cg)\beta(v+\delta) \\ \sigma^2 v\alpha(1+cg)\beta & (v+\delta)\alpha\sigma(1+cg)\beta \end{pmatrix}. \end{aligned}$$

is an 2×2 matrix with $\det(\mathcal{F}\mathcal{V}^{-1}) = 0$,

$$\begin{aligned} R_0 = \rho(\mathcal{F}\mathcal{V}^{-1}) &= \mathrm{Tr}(\mathcal{F}\mathcal{V}^{-1}) \\ &= \frac{[v(1-\alpha) + (v+\delta)\alpha]\sigma(1+cg)\beta}{(v+\delta)[(\gamma_p + \delta + \delta_p) - \theta g] - \sigma v(\gamma_p + \delta + \delta_p)} \\ &= \left[\frac{v}{v+\delta}(1-\alpha) + \alpha\right]\frac{(1+cg)\sigma\beta}{(\gamma_p + \delta + \delta_p) - \theta g - \sigma\frac{v}{v+\delta}(\gamma_p + g\gamma_n)}. \end{aligned}$$

2.2.2 Dynamic Analysis

To study the dynamic behaviour associated with the reduced SLI_pI_n equations (2.6), we will assume individuals in the SSpos compartment are treated at the rate

$$\gamma_\epsilon = \gamma_p + \frac{\theta + \sigma(\gamma_n + \delta + \delta_n)}{\sigma}\epsilon, \tag{2.8}$$

where $\gamma_\epsilon \geq 0$ and γ_p is fixed such that

$$\frac{\sigma(\gamma_p + \delta + \delta_p)}{\theta + \sigma(\gamma_n + \delta + \delta_n)} = 1. \tag{2.9}$$

By introducing the variable $W = I_n - \frac{1-\sigma}{\sigma} I_p$,

$$\begin{aligned}
\dot{W} &= -(\gamma_n + \theta + \delta + \delta_n) I_n - \frac{1-\sigma}{\sigma}[\theta I_n - (\gamma_\epsilon + \delta + \delta_p) I_p] \\
&= -\frac{\theta + \sigma(\gamma_n + \delta + \delta_n)}{\sigma} I_n + \frac{1-\sigma}{\sigma}(\gamma_\epsilon + \delta + \delta_p) I_p \\
&= -\frac{\theta + \sigma(\gamma_n + \delta + \delta_n)}{\sigma} I_n + \frac{1-\sigma}{\sigma}\left[(\gamma_p + \delta + \delta_p) + \frac{\theta + \sigma(\gamma_n + \delta + \delta_n)}{\sigma}\epsilon\right] I_p \\
&= -\frac{\theta + \sigma(\gamma_n + \delta + \delta_n)}{\sigma}\left[I_n - \frac{1-\sigma}{\sigma}\left(\frac{\sigma(\gamma_p + \delta + \delta_p)}{\theta + \sigma(\gamma_n + \delta + \delta_n)} + \epsilon\right) I_p\right] \\
&= -\frac{\theta + \sigma(\gamma_n + \delta + \delta_n)}{\sigma}\left[I_n - \frac{1-\sigma}{\sigma}(1+\epsilon) I_p\right] \\
&= -\frac{\theta + \sigma(\gamma_n + \delta + \delta_n)}{\sigma}\left[W - \epsilon\frac{1-\sigma}{\sigma} I_p\right].
\end{aligned}$$

Consequently when $\epsilon = 0$, the $\{W = 0\}$ plane is a global attractor for the reduced SLI_pI_n equations. As

$$g = \frac{(1-\sigma)(\gamma_\epsilon + \delta + \delta_p)}{\theta + \sigma(\gamma_n + \delta + \delta_n)} = \frac{(1-\sigma)\cdot\frac{\theta+\sigma(\gamma_n+\delta+\delta_n)}{\sigma}\left[\frac{\sigma(\gamma_p+\delta+\delta_p)}{\theta+\sigma(\gamma_n+\delta+\delta_n)} + \epsilon\right]}{\theta + \sigma(\gamma_n + \delta + \delta_n)} = \frac{1-\sigma}{\sigma}(1+\epsilon),$$

conditioning $\epsilon = 0$ is equivalent to conditioning $g = \frac{1-\sigma}{\sigma}$. This means when $\epsilon = 0$, the restriction of the SLI_pI_n equations to the $\{W = 0\}$ plane produces the $\mathbb{R}^2$ dynamic system which was used to calculate the basic reproduction number. This breaks the dynamic analysis into two separate scenarios: the case where $\epsilon = 0$ and the case where $\epsilon \neq 0$. In this section we will study the LI_pW equations

$$\begin{cases}
\dot{L} = (1-\alpha)\frac{\beta}{N}\left[N - L - \left(1 + \frac{1-\sigma}{\sigma}\right) I_p - W\right]\cdot\left[\left(1 + c\frac{1-\sigma}{\sigma}\right) I_p + cW\right] \\
\quad - \alpha x\frac{\beta}{N} L\left[\left(1 + c\frac{1-\sigma}{\sigma}\right) I_p + cW\right] + \gamma_\epsilon I_p + \gamma_n\left(W + \frac{1-\sigma}{\sigma} I_p\right) - (v+\delta)L \\
\\
\dot{I}_p = \sigma\left\{\alpha\frac{\beta}{N}\left[N - (1-x)L - \left(1 + \frac{1-\sigma}{\sigma}\right) I_p - W\right]\cdot\left[\left(1 + c\frac{1-\sigma}{\sigma}\right) I_p + cW\right] + vL\right\} \\
\quad + \theta\left(W + \frac{1-\sigma}{\sigma} I_p\right) - (\gamma_\epsilon + \delta + \delta_p) I_p \\
\\
\dot{W} = -(\gamma_p + \delta + \delta_p)\left[\left(\frac{1-\sigma}{\sigma} - g\right) I_p + W\right].
\end{cases} \tag{2.10}$$

for both scenarios. Before these cases are studied we will introduce the physical relevant region for the LI_pW equations and discuss some preliminary results about the endemic equilibrium points.

The Physically Relevant Region and the Endemic Equilibrium Points

Since $(L, I_p, I_n, N-L-I_p-I_n) \in \mathbb{R}^4_{\geq 0}$ if and only if (L, I_p, W) is contained in the set

$$\Omega_\sigma = \left\{ (L, I_p, W) \in \mathbb{R}^2_{\geq 0} \times \mathbb{R} \,\middle|\, W + \frac{1-\sigma}{\sigma} I_p \geq 0 \text{ and } N - L - \left(1 + \frac{1-\sigma}{\sigma}\right) I_p - W \geq 0 \right\},$$

Ω_σ defines the physically relevant region for the LI_pW equations.

Theorem 2.2.1. *The physically relevant region Ω_σ defines a compact trapping region for the LI_pW equations.*

Proof. To prove Ω_σ defines a compact trapping region for (2.10), we need to show that the inward normal flow on $\partial\Omega_\sigma$ is always positive.

Given a point $(L, I_p, W) \in \partial\Omega_\sigma$,

If $L = 0$:

$$(1,0,0) \cdot (\dot{L}, \dot{I}_p, \dot{W}) = (1-\alpha)\frac{\beta}{N} \overbrace{\left[N - \left(1 + \frac{1-\sigma}{\sigma}\right) I_p - W\right]}^{\geq 0 \text{ because } (L, I_p, W) \in \Omega} \left[\left(1 + c\frac{1-\sigma}{\sigma}\right) I_p + cW\right] + \gamma_p I_p + \gamma_n \left(\frac{1-\sigma}{\sigma} I_p + W\right) \geq 0,$$

If $I_p = 0$ then we must have $W \geq 0$ and as a result:

$$(0,1,0) \cdot (\dot{L}, \dot{I}_p, \dot{W}) = \sigma \left\{ \alpha \frac{\beta}{N} [N - (1-x)L - W] cW + vL \right\} + \theta W \geq 0,$$

If $\frac{1-\sigma}{\sigma} I_p + W = 0$ then because

$$\begin{aligned} g(\gamma_p + \delta + \delta_p) &= \frac{1-\sigma}{\sigma}(1+\epsilon)(\gamma_p + \delta + \delta_p) \\ &= \frac{1-\sigma}{\sigma}\left(1 + \frac{\theta + \sigma(\gamma_n + \delta + \delta_n)}{\sigma(\gamma_p + \delta + \delta_p)}\epsilon\right)(\gamma_p + \delta + \delta_p) \\ &= \frac{1-\sigma}{\sigma}\left[(\gamma_p + \delta + \delta_p) + \frac{\theta + \sigma(\gamma_n + \delta + \delta_n)}{\sigma}\epsilon\right] \\ &= \frac{1-\sigma}{\sigma}(\gamma_\epsilon + \delta + \delta_p) \end{aligned}$$

we must have

$$\begin{aligned} \left(0, \frac{1-\sigma}{\sigma}, 1\right) \cdot (\dot{L}, \dot{I}_p, \dot{W}) &= (1-\sigma)\left\{\alpha\frac{\beta}{N}[N - (1-x)L - I_p]I_p + vL\right\} \\ &\quad + \left[g(\gamma_p + \delta + \delta_p) - \frac{1-\sigma}{\sigma}(\gamma_\epsilon + \delta + \delta_p)\right] I_p \\ &\overset{4}{=} (1-\sigma)\left\{\alpha\frac{\beta}{N}[N - (1-x)L - I_p]I_p + vL\right\} \geq 0, \end{aligned}$$

If $N - L - \left(1 + \frac{1-\sigma}{\sigma}\right) I_p - W = 0$ then:

$$\begin{aligned} \left(-1, -\left(1 + \frac{1-\sigma}{\sigma}\right), -1\right) \cdot (\dot{L}, \dot{I}_p, \dot{W}) &= -\dot{L} - \left(1 + \frac{1-\sigma}{\sigma}\right)\dot{I}_p - \dot{W} \\ &= -\dot{L} - \dot{I}_p - \dot{I}_n \\ &= \delta L + (\delta + \delta_p) I_p + (\delta + \delta_n) I_n \end{aligned}$$

□

Biologically this is an important result as it means any solutions with initial conditions in Ω_σ will remain in Ω_σ for all time $t \geq 0$.

As the LI_pW equations are linear with respect to the L variable, the equilibrium points of (2.10) correspond to the points

$$\left(\frac{(1-\alpha)\frac{\beta(1+cg)}{N}[N-(1+g)I_p] + \gamma_\epsilon + g\gamma_n}{[(1-\alpha)+\alpha x]\frac{\beta(1+cg)}{N} I_p + (v+\delta)} I_p, I_p, \left(g - \frac{1-\sigma}{\sigma} \right) I_p \right)$$

in the (L, I_p, W) space where I_p solves

$$\left\{ \sigma \left[\alpha(x-1)\frac{\beta(1+cg)}{N} I_p + v \right] \frac{L(I_p)}{I_p} + \sigma\alpha \frac{\beta(1+cg)}{N}[N-(1+g)I_p] + \theta g - (\gamma_\epsilon + \delta + \delta_p) \right\} I_p = 0. \tag{2.11}$$

Substituting $L(I_p)$ into the above equation gives us that the equilibrium points are determined by the roots of the cubic equation

$$\left[A_1 \left(\frac{\beta(1+cg)}{N} I_p \right)^2 + B_1 \left(\frac{\beta(1+cg)}{N} I_p \right) + C_1 \right] I_p = 0 \tag{2.12}$$

where

$$\begin{cases} A_1 = -(1+g)\sigma\alpha x \\ \\ B_1 = \sigma\alpha x\beta(1+cg) - \sigma(1+g)(v+\alpha\delta) + \sigma(x-1)\alpha(\gamma_\epsilon + g\gamma_n) + [(1-\alpha)+\alpha x]\cdot[\theta g - (\gamma_\epsilon + \delta + \delta_p)] \\ \\ C_1 = \sigma\beta(1+cg)(v+\alpha\delta) + \sigma v(\gamma_\epsilon + g\gamma_n) + (v+\delta)[\theta g - (\gamma_\epsilon + \delta + \delta_p)] \\ \\ \quad = \{(v+\delta)[(\gamma_\epsilon + \delta + \delta_p) - \theta g] - \sigma v(\gamma_\epsilon + g\gamma_n)\} \left(\dfrac{\sigma\beta(1+cg)(v+\alpha\delta)}{(v+\delta)[(\gamma_\epsilon + \delta + \delta_p) - \theta g] - \sigma v(\gamma_\epsilon + g\gamma_n)} - 1 \right) \\ \\ \quad = \underbrace{\{(v+\delta)[(\gamma_\epsilon + \delta + \delta_p) - \theta g] - \sigma v(\gamma_\epsilon + g\gamma_n)\}}_{\det(\mathcal{V})>0} \cdot \underbrace{\left(\dfrac{\left[\frac{v}{v+\delta}(1-\alpha)+\alpha\right](1+cg)\sigma\beta}{(\gamma_\epsilon + \delta + \delta_p) - \theta g - \sigma\frac{v}{v+\delta}(\gamma_\epsilon + g\gamma_n)} - 1 \right)}_{R_0 - 1}. \end{cases}$$

[4] $g(\gamma_p + \delta + \delta_p) = \frac{1-\sigma}{\sigma}(\epsilon+1)(\gamma_p + \delta + \delta_p) = \frac{1-\sigma}{\sigma}(\gamma_\epsilon + \delta + \delta_p)$

When $R_0 \geq 1$, $|B_1| \leq \sqrt{B_1^2 - 4A_1C_1}$ and as a result there can only be one endemic equilibrium point in $\mathbb{R}^3_{>0}$. When $R_0 \leq 1$ the number of equilibrium points in $\mathbb{R}^3_{>0}$ depends on the sign of the B_1 coefficient. By examining the case where $R_0 \leq 1$ we find that

$$
\begin{aligned}
R_0 \leq 1 \implies \sigma\alpha\beta(1+cg) \leq & -\frac{v}{v+\delta}(1-\alpha)(1+cg)\sigma\beta + (\gamma_\epsilon + \delta + \delta_p) - \theta g - \sigma\frac{v}{v+\delta}(\gamma_\epsilon + g\gamma_n) \\
\implies \qquad B_1 \leq & \; x\left[-\frac{v(1-\alpha)}{v+\delta}(1+cg)\sigma\beta + (\gamma_\epsilon + \delta + \delta_p) - \theta g - \sigma\frac{v}{v+\delta}(\gamma_\epsilon + g\gamma_n)\right] - \sigma(1+g)(v+\alpha\delta) \\
& - \sigma(1-x)\alpha(\gamma_\epsilon + g\gamma_n) + [(1-\alpha) + \alpha x] \cdot [\theta g - (\gamma_\epsilon + \delta + \delta_p)] \\
= & \; x\left[-\frac{v(1-\alpha)}{v+\delta}(1+cg)\sigma\beta - \sigma\frac{v}{v+\delta}(\gamma_p + g\gamma_n)\right] - \sigma(1+g)(v+\alpha\delta) \\
& - \sigma\,(1-x)\,\alpha(\gamma_\epsilon + g\gamma_n) + (1-\alpha)(1-x) \cdot [\theta g - (\gamma_\epsilon + \delta + \delta_p)]\,.
\end{aligned}
$$

.

Consequently as $(\alpha, x) \in [0,1] \times [0,1]$, when $R_o \leq 1$ the LI_pW equations will never have endemic equilibrium points in $\mathbb{R}^3_{>0}$. Referring to the proof of the compact trapping region, we showed that if $(L, I_p, W) \in \partial\Omega_\sigma \backslash \{0\}$ then the inwards normal flow is never equal to zero. As a result, when $R_0 > 1$ the endemic equilibrium points of the LI_pW equations must be contained within the interior of Ω_σ.

Global Stability When $\epsilon = 0$

Under the condition that $\epsilon = 0$, the $W = 0$ plane is invariant with respect the LI_pW equations and as a result the long term dynamics in this plane can be analyzed using the Poincare-Bendixson Theorem and the Bendixson-Dulac Criterion. These are standard theorems for $\mathbb{R}^2$ dynamical systems which can be referenced in *Perko* [13, Chapter 3. Sections: 7 and 9].

Given that $\phi(t, \vec{x}_0)$ is the solution curve of an autonomous $C^1(\mathbb{R}^n)$ system of ordinary differential equations with the initial condition $\vec{x} = \vec{x}_0$, then the forward trajectory of the solution curve is defined to be the set of points

$$\Gamma^+(\vec{x}_0) = \{\vec{y} \in \mathbb{R}^n | \vec{y} = \phi(t, \vec{x}_0) \text{ for some } t \geq 0\}$$

and the $\omega-$limit set of $\Gamma^+(\vec{x}_0)$ is defined to be the set of point

$$\omega(\vec{x}_0) = \left\{ \vec{p} \in \mathbb{R}^n \;\middle|\; \begin{array}{l} \text{There exists a sequence of times } \{t_n\}_{n=1}^{\infty} \text{ such that} \\ \lim\limits_{n\to\infty} t_n = +\infty \text{ and } \lim\limits_{n\to\infty} \phi(t_n, \vec{x}_0) = \vec{p}. \end{array} \right\}.$$

The Poincare-Bendixson Theorem for analytic systems states that if the $\mathbb{R}^2$ autonomous analytic system

$$\dot{\vec{x}} = f(\vec{x}) \tag{2.13}$$

has a forward trajectory $\Gamma^+(\vec{x}_0)$ which is contained in a compact subset F of $\mathbb{R}^2$, then provided F only contains a finite number of critical points, $\omega(\vec{x}_0)$ is either:

1) a critical point of (2.13),

2) a periodic orbit of (2.13),

or 3) a finite union of seperatrix cycles of (2.13)[5]

As $\Omega_\sigma \cap \{W = 0\}$ is a compact trapping region of the LI_pW equations, given any $\vec{x}_0 \in \Omega_\sigma \cap \{W = 0\}$ then $\omega(\vec{x}_0)$ must be a Poincare-Bendixson $\omega-$limit set[6] of (2.10). By using the Poincare Dulac Criterion, we can further limit the possible $\omega-$limit sets of any solution starting with an initial condition in $\Omega_\sigma \cap \{W = 0\}$.

The Poincare Bendixson Dulac Criterion states if there exists a function $g(L, I_p)$ such that $\partial_L(\dot{L} \cdot g) + \partial_{I_p}(\dot{I}_p \cdot g)$ is sign invariant over $Int(\Omega_\sigma) \cap \{W = 0\}$, then this region will not be able to contain any closed orbits. If a closed trajectory C with interior D existed within $Int(\Omega_\sigma) \cap \{W = 0\}$, then by Greens theorem

$$\int\int_D \left(\frac{d(g\dot{L})}{dL} + \frac{d(g\dot{I}_p)}{dI_p} \right) dL dI_p = \oint_C g \cdot (\dot{I}_p dL - \dot{L} dI_p).$$

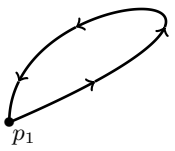
Since the line integral on the right hand side vanishes, the volume integral must evaluate to zero. This contradicts the fact that $\partial_L(\dot{L} \cdot g) + \partial_{I_p}(\dot{I}_p \cdot g)$ has a constant sign over D, and as a result no closed orbits can exits within $Int(\Omega_\sigma) \cap \{W = 0\}$.

[5] A seperatrix cycle is a continuous image of a circle which is composed of a finite number of critical points $\{p_1\}_{i=1}^n$ and a finite set of oriented orbits $\{\phi(t, \vec{x}_i)\}_{i=1}^n$ such that for each $\vec{x}_i$:

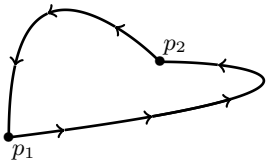
$$\lim_{t \to -\infty} \phi(t, \vec{x}_i) = p_i \quad \text{and} \quad \lim_{t \to +\infty} \phi(t, \vec{x}_i) = \begin{cases} p_{i+1} & \text{if} \quad i \neq n \\ p_1 & \text{if} \quad i = n. \end{cases}$$

Examples:

Homoclinic Cycle



Heteroclinic Cycle



[6] 1) a critical point, 2) a periodic orbit, or 3) a finite union of seperatrix cycles

Theorem 2.2.2. *When* $\epsilon = 0$ $\left(\frac{1-\sigma}{\sigma} = g\right)$, (2.10) *has a Dulac function over* $Int(\Omega_\sigma) \cap \{W = 0\}$[7].

Proof. Let F_σ be the class of differentiable functions $f : Int(\Omega_\sigma) \cap \{W = 0\} \to \mathbb{R}$ which have a constant sign. Since F_σ is closed under multiplication, if there exists a pair $(c, h) \in F_\sigma \times F_\sigma$ such that (c, h) solves

$$\dot{L}\frac{\partial h}{\partial L} + \dot{I}_p\frac{\partial h}{\partial I_p} = h \cdot \left(c - \left[\frac{\partial \dot{L}}{\partial L} + \frac{\partial \dot{I}_p}{\partial I_P}\right]\right) \tag{2.14}$$

then h is a Dulac function of (2.10). Since

$$\begin{aligned}\frac{\partial \dot{L}}{\partial L} + \frac{\partial \dot{I}_p}{\partial I_p} =& -\left[[(1-\alpha)+\alpha x]\frac{\beta(1+cg)}{N}I_p + (v+\delta)\right] \\ &+ \sigma\left\{\alpha\frac{\beta(1+cg)}{N}[N-(1-x)L-2(1+g)I_p]\right\} + \theta g - (\gamma_\epsilon + \delta + \delta_p)\end{aligned}$$

by defining

$$c(L, I) = -\left[[(1-\alpha)+\alpha x]\frac{\beta(1+cg)}{N}I_p + (v+\delta)\right] - \sigma v\frac{L}{I_p} - \sigma\alpha\frac{\beta(1+cg)}{N}(1+g)I_p,$$

$c \in F_\sigma$ and moreover

$$\begin{aligned}\frac{c - \left[\frac{\partial \dot{L}}{\partial L} + \frac{\partial \dot{I}}{\partial I}\right]}{\dot{I}_p} &= -\frac{\sigma\left\{\alpha\frac{\beta(1+cg)}{N}[N-(1-x)L-(1+g)I_p] + v\frac{L}{I_p}\right\} + \theta g - (\gamma_\epsilon + \delta + \delta_p)}{\dot{I}_p} \\ &= -\frac{1}{I_p}.\end{aligned}$$

If we now assume $h = h(I_p)$, (2.14) gives us that

$$\frac{1}{h}\frac{\partial h}{\partial I_p} = \frac{c - \left[\frac{\partial \dot{L}}{\partial L} + \frac{\partial \dot{I}_p}{\partial I_p}\right]}{\dot{I}_p} \quad \Longrightarrow \quad \frac{\partial \ln(h)}{\partial I_p} = -\frac{1}{I_p} \quad \Longrightarrow \quad h(I_p) = \frac{1}{I_p}.$$

Since $h \in F_\sigma$, $h(I_p) = \frac{1}{I_p}$ is a Dulac function for the LI_pW equation over $Int(\Omega_\sigma) \cap \{W = 0\}$. □

[7] This proof technique was introduced by *Osvaldo Osuna et al.* [12], and works for a large class of $\mathbb{R}^2$ compartmental disease models.

Since $Int(\Omega_\sigma) \cap \{W = 0\}$ cannot support a closed orbit, given any $\vec{x}_0 \in \Omega_\sigma \cap \{W = 0\}$ then $\omega(\vec{x}_0)$ is either

1) a critical point of (2.10)

or 2) a compound seperatrix cycle of (2.10) which contains the disease free equilibrium.

When $R_0 < 1$ the disease free equilibrium is locally stable and as a result it will not be able to support a compound seperatrix cycle. This means the disease free equilibrium is globally stable in $\Omega_\sigma \cap \{W = 0\}$ whenever $R_0 < 1$.

Because $g = \frac{1-\sigma}{\sigma}$, by defining J_0 to be the Jacobian matrix of the LI_pW equations at the disease free equilibrium,

$$\det(J_0) = \begin{vmatrix} \frac{\partial \dot{L}}{\partial L} & \frac{\partial \dot{L}}{\partial I_p} & \frac{\partial \dot{L}}{\partial W} \\ \frac{\partial \dot{I}_p}{\partial L} & \frac{\partial \dot{I}_p}{\partial I_p} & \frac{\partial \dot{I}_p}{\partial W} \\ 0 & 0 & -(\gamma_p + \delta + \delta_p) \end{vmatrix} = -(\gamma_p + \delta + \delta_p) \det(\mathcal{F} - \mathcal{V})$$

where $\mathcal{F} - \mathcal{V}$ is the Jacobian matrix of the (2.7) at the origin. Now by factoring the $\mathcal{V}$ matrix out of the determinant, we have that

$$\begin{aligned} \det(J_0) &= -(\gamma_p + \delta + \delta_p) \det(\mathcal{F}\mathcal{V}^{-1} - \lambda \mathbb{I})\big|_{\lambda=1} \det(\mathcal{V}) \\ &= -(\gamma_p + \delta + \delta_p)(1 - R_0) \det(\mathcal{V}). \end{aligned}$$

This shows the disease free equilibrium is a nondegenerate critical point whenever $R_0 \neq 1$.

Theorem 2.2.3. *If $R_0 > 1$, then the disease free equilibrium cannot support a compound seperatrix cycle.*

Proof. If the disease free equilibrium were to support a compound seperatrix cycle in Ω_σ when $R_0 > 1$, then disease free equilibrium would have to be a nondegenerate saddle node. This means that J_0 must have eigenpairs $(\lambda_+, \vec{v}_+)$ with $\lambda_+ > 0$ and $(\lambda_-, \vec{v}_-)$ with $\lambda_- < 0$.

Moreover since Ω_σ is a trapping region and since the disease free equilibrium is the only critical point located on $\partial\Omega_\sigma$, the seperatrix cycle cannot leave Ω_σ. This means we must have $\vec{v}_- \subseteq \Omega_\sigma$.

Applying the Hartman-Grobman Theorem[8] guarantees the existence of two open neighbourhoods V and V^* of the origin and a homeomorphism $h : V \to V^*$ such that

$$\begin{cases} \vec{x}_0 \in V \\ \phi(\vec{x}_0, t) \in V \end{cases} \implies h \circ \phi(\vec{x}_0, t) = e^{J_0 t} h(\vec{x}_0).$$

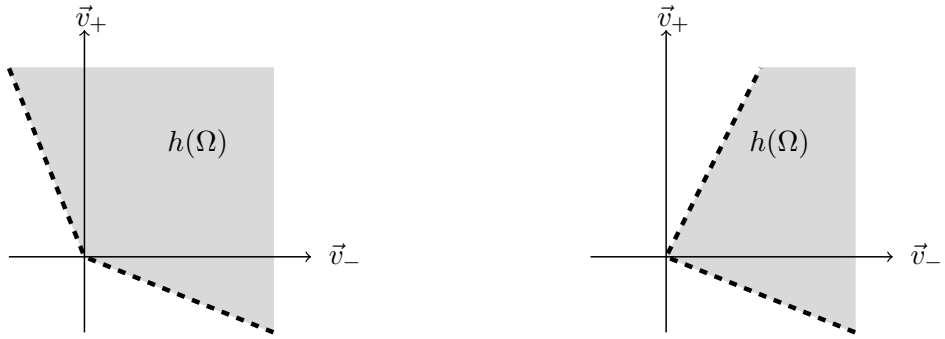
[8]This requires the equilibrium point to be nondegenerate.

As $\vec{v}_+$ and $\vec{v}_-$ are eigenvectors of J_0, $h^{-1}(\text{span}\{\vec{v}_+, \vec{v}_-\})$ is an invariant manifold of the non-linear LI_pW equations, whose dynamics are completely determined by the linear subsystem

$$\begin{pmatrix} \dot{v}_+ \\ \dot{v}_- \end{pmatrix} = \begin{pmatrix} \lambda_+ & 0 \\ 0 & \lambda_- \end{pmatrix} \begin{pmatrix} v_+ \\ v_- \end{pmatrix}$$

in V^*.

As the disease free equilibrium is located at the corner of the physical region, the disease free equilibrium must be a boundary point of both $h(\Omega_\sigma) \cap \text{span}\{\vec{v}_+\}$ and $h(\Omega_\sigma) \cap \text{span}\{\vec{v}_-\}$. From this there are only two configurations of $h(\Omega_\sigma)$ in V^* that need to be considered.



In either case the global asymptotic stability of $\text{span}\{\vec{v}_+\}$ guarantees outflow on a section of $h(\partial\Omega_\sigma)$. This contracts the fact that Ω_σ is a trapping region. □

As a result of the above theorem, the $\omega-$limit set of any physical solution starting in $\Omega_\sigma \cap \{W = 0\}$ must be a critical point of (2.10) whenever $R_0 > 1$. This proves global stability of the endemic equilibrium point in $\Omega_\sigma \cap \{W = 0\} \backslash \{0\}$ when $R_0 > 1$.

Stability Results for the general LI_pW model ($\epsilon \neq 0$)

When $\epsilon \neq 0$, the disease free equilibrium remains globally stable within Ω_σ whenever $R_0 < 1$. This result follows from *Chavez et. al.* [2] who proved the global stability of the disease free equilibrium for a general class of compartmental disease models. Since the reduced SLI_nI_p equations can be rewritten as

$$\begin{pmatrix} \dot{L} \\ \dot{I}_p \\ \dot{I}_n \end{pmatrix} = J_0 \begin{pmatrix} L \\ I_p \\ I_n \end{pmatrix} - \vec{f}(L, I_p, I_n)$$

where J_0 is the Jacobian matrix at the disease free equilibrium and $\vec{f}(L, I_p, W) > 0$ for all $(L, I_p, I_n) \in \{\mathbb{R}^3_{\geq 0} | L + I_p + I_n \leq N\}$ (this is the trapping region for the SLI_pI_n equations), the variation of parameters formula gives:

$$\begin{pmatrix} L(t) \\ I_p(t) \\ I_n(t) \end{pmatrix} = e^{J_0 t} \begin{pmatrix} L(0) \\ I_p(0) \\ I_n(0) \end{pmatrix} - \int_0^t e^{(t-s)J_0} \vec{f}(L(s), I_p, (s), W(s)) ds.$$

Since $-J_0$ is a nonsingular M-matrix when $R_0 < 1$, e^{tJ_0} is a nonnegative matrix. and as a result for all times $t \geq 0$,

$$\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \leq \begin{pmatrix} L(t) \\ I_p(t) \\ I_n(t) \end{pmatrix} \leq e^{J_0 t} \begin{pmatrix} L(0) \\ I_p(0) \\ I_n(0) \end{pmatrix}.$$

Hence the disease free equilibrium will be globally stable inside the physically relevant region whenever $R_0 \geq 1$.

To determine the stability of the endemic equilibrium point for general values of $\epsilon \neq 0$ we need to determine when all eigenvalues of the Jacobian matrix

$$J = \begin{pmatrix} \frac{\partial \dot{L}}{\partial L} & \frac{\partial \dot{L}}{\partial I_p} & \frac{\partial \dot{L}}{\partial W} \\ \frac{\partial \dot{I}_p}{\partial L} & \frac{\partial \dot{I}_p}{\partial I_p} & \frac{\partial \dot{I}_p}{\partial W} \\ \frac{\partial \dot{W}}{\partial L} & \frac{\partial \dot{W}}{\partial I_p} & \frac{\partial \dot{W}}{\partial W} \end{pmatrix} \tag{2.15}$$

have a negative real part. As J has the characteristic equation

$$\det(J - \lambda \mathbb{I}) = -\lambda^3 + c_2\lambda^2 - c_1\lambda + c_0 \quad \text{where} \quad \begin{cases} c_0 = \det(J) \\ c_1 = \begin{vmatrix} J_{11} & J_{12} \\ J_{21} & J_{22} \end{vmatrix} + \begin{vmatrix} J_{11} & J_{13} \\ J_{31} & J_{33} \end{vmatrix} + \begin{vmatrix} J_{22} & J_{23} \\ J_{32} & J_{33} \end{vmatrix} \\ c_2 = \text{Trace}(J), \end{cases} \tag{2.16}$$

all eigenvalues of J will have a negative real part if and only if: $c_0 < 0$, $c_2 < 0$, and $c_1c_2 < c_0$. These are the three Routh-Hurwitz stability conditions for the above characteristic polynomial. Because of the number of coefficients in this model, it is too cumbersome to determine the local stability of the endemic equilibrium points.

Theorem 2.2.4. *When $\epsilon \geq 0$, the endemic equilibrium point will exchange stability with the disease free equilibrium as R_0 passes through one. Moreover if $R_0 \geq 1$:*

1) the trace of the Jacobian matrix (2.15) will be negative whenever $\epsilon \geq 0$,

and 2) the determinant of the Jacobian matrix (2.15) will be negative regardless of ϵ.

Proof. By defining J to be the Jacobian matrix (2.15) and $\det(J-\lambda\mathbb{I}) = -\lambda^3+c_2\lambda^2-c_1\lambda+c_0$ it follows that,

$$c_1 = \begin{vmatrix} \frac{\partial \dot{L}}{\partial L} & \frac{\partial \dot{L}}{\partial I_p} & \left(\frac{1-\sigma}{\sigma} - g\right)\frac{\partial \dot{L}}{\partial W} \\ \frac{\partial \dot{I}_p}{\partial L} & \frac{\partial \dot{I}_p}{\partial I_p} & \left(\frac{1-\sigma}{\sigma} - g\right)\frac{\partial \dot{I}_p}{\partial W} \\ 0 & -(\gamma_p+\delta+\delta_p) & -(\gamma_p+\delta+\delta_p) \end{vmatrix} = -(\gamma_p+\delta+\delta_p) \begin{vmatrix} \frac{\partial \dot{L}}{\partial L} & \left[\frac{\partial}{\partial I_p} - \left(\frac{1-\sigma}{\sigma} - g\right)\frac{\partial}{\partial W}\right]\dot{L} \\ \frac{\partial \dot{I}_p}{\partial L} & \left[\frac{\partial}{\partial I_p} - \left(\frac{1-\sigma}{\sigma} - g\right)\frac{\partial}{\partial W}\right]\dot{I}_p \end{vmatrix}.$$

Since the LI_pW equations were created by making the substituting $I_n = \frac{1-\sigma}{\sigma}I_p - W$ into the reduced SLI_pI_n model, the W terms in the $\dot{L}$ and $\dot{I}_p$ equation always appear in the linear combination: $\frac{1-\sigma}{\sigma}I_p - W$. Now since

$$\frac{1-\sigma}{\sigma}I_p - W\Big|_{W=\left(g-\frac{1-\sigma}{\sigma}\right)I_p} = gI_p \quad \text{and} \quad \left[\frac{\partial}{\partial I_p} - \left(\frac{1-\sigma}{\sigma} - g\right)\frac{\partial}{\partial W}\right]\left(\frac{1-\sigma}{\sigma}I_p - W\right) = \frac{\partial}{\partial I_p}gI_p,$$

at the endemic equilibrium point $\det(J) = -(\gamma_p+\delta+\delta_p)\det(J^*)$ where J^* is the Jacobian matrix of the $\mathbb{R}^2$ system:

$$\begin{cases} \dot{L} = (1-\alpha)\frac{\beta(1+cg)}{N}[N-L-(1+g)I_p]I_p + (\gamma_\epsilon+\gamma_n g)I_p - (v+\delta)L \\ \dot{I}_p = \sigma\left\{\alpha\frac{\beta(1+cg)}{N}[N-L-(1+g)I_p]I_p + vL\right\} + \theta g I_p - (\gamma_\epsilon+\delta+\delta_p)I_p. \end{cases} \tag{2.17}$$

As the (2.17) is the restriction of the LI_pW equation to $W = \left(g - \frac{1-\sigma}{\sigma}\right)I_p$ plane, at the endemic equilibrium point we have

$$\begin{cases} L = \frac{(1-\alpha)\frac{\beta(1+cg)}{N}[N-(1+g)I_p]+\gamma_\epsilon+g\gamma_n}{[(1-\alpha)+\alpha x]\frac{\beta(1+cg)}{N}I_p+(v+\delta)}I_P \\ \sigma\left[\alpha(x-1)\frac{\beta(1+cg)}{N}I_p + v\right]\frac{L}{I_p} + \sigma\alpha\frac{\beta(1+cg)}{N}[N-(1+g)I_p] + \theta g - (\gamma_\epsilon+\delta+\delta_p) = 0, \end{cases} \tag{2.18}$$

and consequently

$$\det(J^*) = \begin{vmatrix} -[(1-\alpha)+\alpha x]\frac{\beta(1+cg)}{N}I_p - (v+\delta) & (1-\alpha)\frac{\beta(1+cg)}{N}[N-L-2(1+g)I_p] - \alpha x\frac{\beta(1+cg)}{N}L + \gamma_\epsilon + g\gamma_n \\ \sigma\alpha(x-1)\frac{\beta(1+cg)}{N}I_p + \sigma v & \sigma\alpha\frac{\beta(1+cg)}{N}[N-(1-x)L-2(1+g)I_p] + \theta g - (\gamma_\epsilon+\delta+\delta_p) \end{vmatrix}$$

$$= \begin{vmatrix} -[(1-\alpha)+\alpha x]\frac{\beta(1+cg)}{N}I_p - (v+\delta) & (1-\alpha)\frac{\beta(1+cg)}{N}[N-(1+g)I_p] + \gamma_\epsilon + g\gamma_n \\ \sigma\alpha(x-1)\frac{\beta(1+cg)}{N}I_p + \sigma v & \sigma\alpha\frac{\beta(1+cg)}{N}[N-(1+g)I_p] + \theta g - (\gamma_\epsilon+\delta+\delta_p) \end{vmatrix}$$
$$+\sigma \begin{vmatrix} -[(1-\alpha)+\alpha x]\frac{\beta(1+cg)}{N}I_p - (v+\delta) & -(1-\alpha)\frac{\beta(1+cg)}{N}[L+(1+g)I_p] - \alpha x\frac{\beta(1+cg)}{N}L \\ \alpha(x-1)\frac{\beta(1+cg)}{N}I_p + v & -\alpha\frac{\beta(1+cg)}{N}[L+(1+g)I_p] + \alpha x\frac{\beta(1+cg)}{N}L \end{vmatrix}$$

$$\overset{9}{=} \sigma \begin{vmatrix} -[(1-\alpha)+\alpha x]\frac{\beta(1+cg)}{N}I_p - (v+\delta) & -(1-\alpha)\frac{\beta(1+cg)}{N}[L+(1+g)I_p] - \alpha x\frac{\beta(1+cg)}{N}L \\ \alpha(x-1)\frac{\beta(1+cg)}{N}I_p + v & -\alpha\frac{\beta(1+cg)}{N}[L+(1+g)I_p] + \alpha x\frac{\beta(1+cg)}{N}L \end{vmatrix}$$

$$= \sigma\left\{\left[\alpha x\frac{\beta(1+cg)}{N}I_p + (v+\alpha\delta)\right]\frac{\beta(1+cg)}{N}[L+(1+g)I_p] - \left[\alpha x\frac{\beta(1+cg)}{N}I_p + \alpha x\delta\right]\frac{\beta(1+cg)}{N}L\right\}$$

Since $\alpha x\delta < v + \delta$, it follows that $\det(J) < 0$ whenever $R_0 > 1$. This analysis also shows that the sign of $\det(J)$ changes as R_0 passes through one.

Provided that $\epsilon \geq 0$, then if $R_0 \geq 1$ it follows that

$$
\begin{aligned}
c_2 = & -[(1-\alpha)+\alpha x]\frac{\beta}{N}\left[\left(1+c\frac{1-\sigma}{\sigma}\right)I_p + cW\right] - (v+\delta) \\
& + \sigma\alpha\frac{\beta}{N}\left[N-(1-x)L-2\left(1+\frac{1-\sigma}{\sigma}\right)I_p - W\right]\left(1+c\frac{1-\sigma}{\sigma}\right) - \sigma\alpha\frac{\beta}{N}c\left(1+\frac{1-\sigma}{\sigma}\right)W \\
& + \theta\frac{1-\sigma}{\sigma} - (\gamma_\epsilon+\delta+\delta_p) - (\gamma_p+\delta+\delta_p) \\
\\
= & -[(1-\alpha)+\alpha x]\frac{\beta}{N}\left[\left(1+c\frac{1-\sigma}{\sigma}\right)I_p + cW\right] - \alpha\frac{\beta}{N}\left[\left(1+c\frac{1-\sigma}{\sigma}\right)I_p + cW\right] \\
& + \sigma\alpha\frac{\beta}{N}\left[N-(1-x)L-\left(1+\frac{1-\sigma}{\sigma}\right)I_p - W\right]\left(1+c\frac{1-\sigma}{\sigma}\right) + \theta\frac{1-\sigma}{\sigma} - (\gamma_\epsilon+\delta+\delta_p) \\
& - (\gamma_p+\delta+\delta_p) - (v+\delta) \\
\\
\overset{10}{=} & -[(1-\alpha)+\alpha x]\frac{\beta}{N}\left[\left(1+c\frac{1-\sigma}{\sigma}\right)I_p + cW\right] - \alpha\frac{\beta}{N}\left[\left(1+c\frac{1-\sigma}{\sigma}\right)I_p + cW\right] \\
& - \sigma\left\{\alpha\frac{\beta}{N}\left[N-(1-x)L-\left(1+\frac{1-\sigma}{\sigma}\right)I_p - W\right]c\frac{W}{I_p} + v\frac{L}{I_p}\right\} - \theta\frac{W}{I_p} \\
& - (\gamma_p+\delta+\delta_p) - (v+\delta) \\
\\
= & -(1+\alpha x)\frac{\beta(1+cg)}{N}I_p - \sigma\left\{\alpha\frac{\beta c}{N}[N-(1-x)L-(1+g)I_p]\left(g-\frac{1-\sigma}{\sigma}\right) + v\frac{L}{I_p}\right\} \\
& - \theta\left(g-\frac{1-\sigma}{\sigma}\right) - (\gamma_p+\delta+\delta_p) - (v+\delta) \\
\\
< & \ 0.
\end{aligned}
$$

[9] By factoring $-\left([(1-\alpha)+\alpha x]\frac{\beta(1+cg)}{N} + (v+\delta)\right)$ out of the top row in the first determinant, the a_{12} entry becomes $-\frac{L}{I_P}$. As a result the evaluation of this determinant produces the second equation from (2.18).

[10] Using the fact that $\dot{I}_p = 0$ (2.10).

When $R_0 = 1$,

$$\left[\frac{v}{v+\delta}(1-\alpha)+\alpha\right]\sigma\beta(1+cg) = (\gamma_\epsilon+\delta+\delta_p) - \theta g - \sigma\frac{v}{v+\delta}[\gamma_\epsilon + g\gamma_n]$$

and consequently

$$c_1 = \begin{vmatrix} \frac{\partial \dot{L}}{\partial L} & \frac{\partial \dot{L}}{\partial I_p} \\ \frac{\partial \dot{I}_p}{\partial L} & \frac{\partial \dot{I}_p}{\partial I_p} \end{vmatrix} + \begin{vmatrix} \frac{\partial \dot{I}_p}{\partial I_p} & \frac{\partial \dot{I}_p}{\partial W} \\ -\left(\frac{1-\sigma}{\sigma} - m\right)(\gamma_p+\delta+\delta_p) & -(\gamma_p+\delta+\delta_p) \end{vmatrix} + \begin{vmatrix} \frac{\partial \dot{L}}{\partial L} & \frac{\partial \dot{L}}{\partial W} \\ 0 & -(\gamma_p+\delta+\delta_p) \end{vmatrix}$$

$$= \begin{vmatrix} \frac{\partial \dot{L}}{\partial L} & \frac{\partial \dot{L}}{\partial I_p} \\ \frac{\partial \dot{I}_p}{\partial L} & \frac{\partial \dot{I}_p}{\partial I_p} \end{vmatrix} - (\gamma_p+\delta+\delta_p)\left\{\frac{\partial \dot{L}}{\partial L} + \left[\frac{\partial}{\partial I_p} - \left(\frac{1-\sigma}{\sigma} - g\right)\frac{\partial}{\partial W}\right]\dot{I}_p\right\}$$

$$= \begin{vmatrix} -(v+\delta) & (1-\alpha)\beta\left(1+c\frac{1-\sigma}{\sigma}\right) + \gamma_\epsilon + \gamma_n\frac{1-\sigma}{\sigma} \\ \sigma v & \sigma\alpha\beta\left(1+c\frac{1-\sigma}{\sigma}\right) + \theta\frac{1-\sigma}{\sigma} - (\gamma_\epsilon+\delta+\delta_p) \end{vmatrix}$$
$$- (\gamma_p+\delta+\delta_p)\left\{-(v+\delta) + \sigma\alpha\beta(1+cg) + \theta g - (\gamma_\epsilon+\delta+\delta_p)\right\}$$

$$= \begin{vmatrix} -(v+\delta) & (1-\alpha)\beta\left(1+c\frac{1-\sigma}{\sigma}\right) + \gamma_\epsilon + \gamma_n\frac{1-\sigma}{\sigma} \\ \sigma v & \sigma\alpha\beta\left(1+c\frac{1-\sigma}{\sigma}\right) + \theta\frac{1-\sigma}{\sigma} - (\gamma_\epsilon+\delta+\delta_p) \end{vmatrix}$$
$$- (\gamma_p+\delta+\delta_p)\left\{-(v+\delta) - \frac{v}{v+\delta}(1-\alpha)\beta(1+cg) - \frac{v}{v+\delta}[\gamma_\epsilon + g\gamma_n]\right\}$$

$$> \begin{vmatrix} -(v+\delta) & (1-\alpha)\beta\left(1+c\frac{1-\sigma}{\sigma}\right) + \gamma_\epsilon + \gamma_n\frac{1-\sigma}{\sigma} \\ \sigma v & \sigma\alpha\beta\left(1+c\frac{1-\sigma}{\sigma}\right) + \theta\frac{1-\sigma}{\sigma} - (\gamma_\epsilon+\delta+\delta_p) \end{vmatrix}$$

$$= \begin{vmatrix} -(v+\delta) & (1-\alpha)\beta(1+cg) + \gamma_\epsilon + g\gamma_n \\ \sigma v & \sigma\alpha\beta(1+cg) + \theta g - (\gamma_\epsilon+\delta+\delta_p) \end{vmatrix} + \begin{vmatrix} -(v+\delta) & [(1-\alpha)c\beta + \gamma_n]\left(\frac{1-\sigma}{\sigma} - g\right) \\ \sigma v & [\sigma\alpha\beta c + \theta]\left(\frac{1-\sigma}{\sigma} - g\right) \end{vmatrix}$$

$$= \left(g - \frac{1-\sigma}{\sigma}\right)\begin{vmatrix} (v+\delta) & (1-\alpha)c\beta + \gamma_n \\ -\sigma v & \sigma\alpha\beta c + \theta \end{vmatrix}$$

If $\epsilon \geq 0$ and $R_0 = 1$,

$$c_0 = 0 \quad \text{and} \quad \begin{cases} c_1 < 0 \\ c_1 c_2 - c_0 < 0. \end{cases}$$

This guarantees that c_1 and $c_1c_2 - c_0$ are both less than zero in some neighbourhood of $R_0 = 1$. Since the sign of c_0 flips as R_0 passes through one, the endemic equilibrium point will exchange stability with the disease free equilibrium whenever $\epsilon \geq 0$.

□

The above analysis shows that as R_0 passes through one the LI_pW equations exhibits a transcritical bifurcation. The bifurcation diagram for the transcritical bifurcation is illustrated in Figure (2.3). The solid red line indicates the equilibrium point is local stability and the dashed blue line indicates the equilibrium point is unstable. It should be noted the local stability of the endemic equilibrium point is only known in a neighbourhood of the transcritical bifurcation.

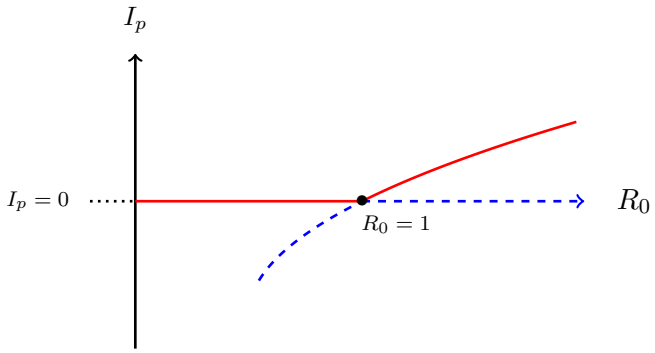


Figure 2.3: The transcritical bifurcation

2.3 The TIME Model

The Tuberculosis Incidence and Mortality Estimates (TIME) model is a compartmental disease model which was developed for the Spectrum modelling suite to help make disease burden predictions [8]. The TIME model separates the population into fifteen different compartments to capture a more realistic view of TB control measures. The latent and active TB compartments are both stratified by treatment history and multi-drug resistance (MDR) status. This model also makes the distinction between sputum smear positive (SSpos) and sputum smear negative (SSneg) infections.

Drug Resistant Status	Compartments	Symbol
	Susceptible	S
Drug Susceptible (treatable with first line drugs)	Latent Treatment Naive	L_{S_N}
	Latent Previously Treated	L_{S_P}
	Active Smear Positive Treatment Naive	I_{S_N}
	Active Smear Positive Previously Treated	I_{S_P}
	Active Smear Negative Treatment Naive	N_{S_N}
	Active Smear Negative Previously Treated	N_{S_P}
Multi-Drug Resistant (Resistant to both Rifampicin and Isoniazid)	MDR Latent Treatment Naive	L_{M_N}
	MDR Latent Previously Treated	L_{M_P}
	MDR Active Smear Positive Treatment Naive	I_{M_N}
	MDR Active Smear Positive Previously Treated	I_{M_P}
	MDR Active Smear Negative Treatment Naive	N_{M_N}
	MDR Active Smear Negative Previously Treated	N_{M_P}

Table 2.2: The TIME model compartments

Figure 2.4 illustrates the basic flow between the various compartments of the TIME model. It should be noted that this figure does not include the cross reinfections in the TIME model between the drug susceptible and multi-drug resistant strains.

As in the sputum smear model, the TIME model assumes that TB is transmitted by SSpos and SSneg individuals with a standard incidence force of infection. Drug susceptible sputum smear positive individuals have an infectivity rate of β sufficient contacts per year, whereas drug susceptible individuals with sputum smear negative TB have a reduced infectivity rate of βc sufficient contacts per year. The infectiousness of the MDR strain is reduced by the relative fitness parameter ζ. This mean that at any time t the infection rates for the drug susceptible and MDR populations are defined to be

$$\lambda_S = \frac{\beta(I_{S_N} + I_{S_P}) + \beta c(N_{S_N} + N_{S_P})}{T_{pop}} \quad \text{and} \quad \lambda_M = \frac{\beta(I_{M_N} + I_{M_P}) + \beta c(N_{M_N} + N_{M_P})}{T_{pop}}\zeta$$

respectively where T_{pop} is defined to be the size of the total population.

To capture the high variability in the latent progression rate the TIME model assumes $\alpha \cdot 100$ percent of all infections proceed directly into the active TB compartments while the other $(1-\alpha)\cdot 100$ percent of infections proceed into the respective latent compartments. Individuals with latent tuberculosis are assumed to have acquired a partial immunity against reinfection and as a result they may undergo exogenous reinfection. The degree of protection offered by a latent infection is determined by the parameter x[11].

As in the sputum smear model, $\sigma \cdot 100$ percent of all individuals entering the active TB compartments develop a sputum smear positive infection, while the other $(1-\sigma)\cdot 100$ percent develop a sputum smear negative infection. The TIME model also assumes that individuals in the sputum smear negative compartments convert to the sputum smear positive compartments at the rate θ regardless of treatment history or drug resistant status.

In the TIME model individuals may undergo cross reinfection between the drug susceptible and MDR strains from the latent compartments. The degree of protection offered by a latent infection is also assumed to protect individuals against cross reinfections. If a drug susceptible latent individual makes sufficient contact with someone with active MDR but they do not proceed into the MDR active TB compartment, it is assumed they have an $\iota \cdot 100$ percent chance of developing latent MDR TB where $\iota \equiv \frac{\zeta}{\zeta+1}$. Similarly if an MDR latent individual makes sufficient contact with an active drug susceptible TB individual but they do not proceed into the drug susceptible TB compartment, it is assumed they have a $(1-\iota)\cdot 100$ percent chance of developing latent drug susceptible TB. The cross reinfections between drug susceptible TB and MDR TB are illustrated in Figure 2.5.

Individuals who are infected with active TB may spontaneously recover back into the latent phase. This self-recovery is assumed to occur at the same rate r for all individuals with active TB. The TIME model also assumes treatment naive latent individuals develop active TB, independently from secondary reinfection events, at the rate v (the latent progression rate) and that previously treated latent individuals relapse back into the active TB phase at the rate v_0.

Regardless of treatment history or drug resistance status, sputum smear positive and sputum smear negative individuals are diagnosed at the respective rates of γ and $d\gamma$. The initial diagnostic tests are assumed to only identify people as having active TB. To determine if a diagnosed cases has MDR TB the individual must take a drug susceptibility test (DST) which has a known specificity Sp_M, and a known sensitivity Se_M[12]. The TIME model also

[11] Therefore, only $\alpha(1-x)\cdot 100$ percent of the exogenous reinfections proceed onto the active TB compartments.

[12] The specificity of a test is defined to be the percentage of negative cases which are correctly diagnosed. The sensitivity of a test is defined to be the percentage of positive cases which are correctly diagnosed.

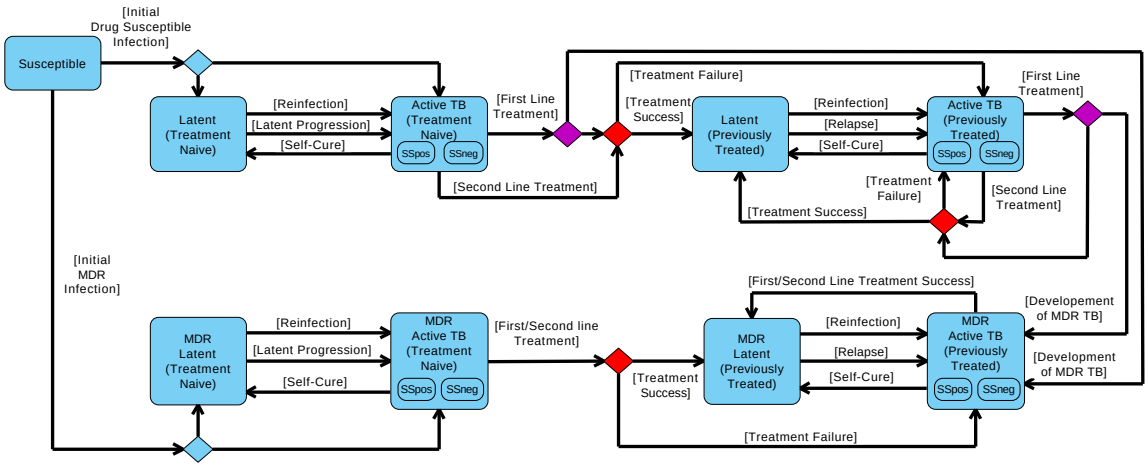


Figure 2.4: The flow diagram for the TIME model

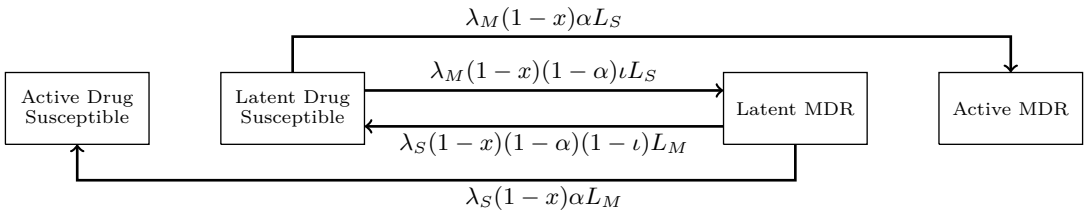


Figure 2.5: Cross reinfections in the TIME model

assumes the proportion of diagnosed cases which take a DST is dependent on the treatment history so that $\Psi_N \cdot 100$ percent of treatment naive cases and $\Psi_P \cdot 100$ percent of the previously treated cases take a DST. Individuals who test positive for MDR TB are linked onto second line care with the probability of $\eta_M \cdot 100$ and everyone else is linked onto first line care with the probability of $\eta_S \cdot 100$ (this includes individuals who have not taken a DTS).

Compartments	First Line Care	Second Line Care
Tr. Naive.	$F_{S_N} = [\Psi_N Sp_M + (1-\Psi_N)] \cdot \eta_S$	$S_{S_N} = [\Psi_N(1-Sp_M)] \cdot \eta_M$
Prev. Tr.	$F_{S_P} = [\Psi_P Sp_M + (1-\Psi_P)] \cdot \eta_S$	$S_{S_P} = [\Psi_P(1-Sp_M)] \cdot \eta_M$
MDR Tr. Naive	$F_{M_N} = [\Psi_N(1-Se_M)+(1-\Psi_N)] \cdot \eta_S$	$S_{M_N} = [\Psi_N Se_M] \cdot \eta_M$
MDR Prev. Tr.	$F_{M_P} = [\Psi_P(1-Se_M)+(1-\Psi_P)] \cdot \eta_S$	$S_{M_P} = [\Psi_P Se_M] \cdot \eta_M$

Table 2.3: Proportion of cases linked onto first and second line care

First line treatment with Rifampicin or Isoniazid for drug susceptible TB is assumed to have a treatment success rate of $\tau_n \cdot 100$ percent for treatment naive individuals and $\tau_p \cdot 100$ for previously treated individuals[13]. When first line care is used on an MDR patient, then the treatment success rates drop by $(1-R_n) \cdot 100$ percent. Second line treatment is assumed to

[13]The TIME model originally assumed that drug susceptible individuals were treated at the same rate regardless of treatment history. As this information was provided by the WHO, the TIME model was altered to include different treatment naive and previously treated treatment success rates.

have a treatment success rate of $\tau_m \cdot 100$ percent regardless the patients treatment history. The TIME model also assumes that because of the genetic instability of TB, $\xi \cdot 100$ percent of all drug susceptible cases treated with first line drugs will develop MDR TB.

This model assumes no vertical transmission TB so all births $\Lambda(t)$ are born into the susceptible compartment. Everyone is assumed to die at the natural death rate δ. Individuals with SSpos and SSneg TB die at the respective rates $\delta + \delta_p$ and $\delta + \delta_n$, where δ_p and δ_n are the death rates due to SSpos and SSneg TB.

Non-MDR Compartments

$$
\begin{cases}
\frac{dS}{dt} = \Lambda(t) - (\lambda_S + \lambda_M + u)S \\
\\
\frac{dL_{S_N}}{dt} = -\Big(v + \lambda_S(1-x)a + \lambda_M(1-x)[a + (1-a)\iota] + \delta\Big)L_{S_N} + r(I_{S_N} + N_{S_N}) \\
\qquad + \lambda_S(1-x)(1-a)(1-\iota)L_{M_N} + \lambda_S(1-a)S \\
\\
\frac{dI_{S_N}}{dt} = \sigma\Big([v + \lambda_S(1-x)a]L_{S_N} + \lambda_S a(1-x)L_{M_N} + \lambda_S aS\Big) + \theta N_{S_N} \\
\qquad - (\delta + \delta_I + r)I_{S_N} - (F_{S_N} + S_{S_N})\gamma I_{S_N} \\
\\
\frac{dN_{S_N}}{dt} = (1-\sigma)\Big([v + \lambda_S(1-x)a]L_{S_N} + \lambda_S a(1-x)L_{M_N} + \lambda_S aS\Big) \\
\qquad - (r + \theta + \delta + \delta_N)N_{S_N} - (F_{S_N} + S_{S_N})d\gamma N_{S_N} \\
\\
\frac{dL_{S_P}}{dt} = -\Big(v_0 + \lambda_S(1-x)a + \lambda_M(1-x)[a + (1-a)\iota] + \delta\Big)L_{S_P} + r(I_{S_P} + N_{S_P}) \\
\qquad + \lambda_S(1-x)(1-a)(1-\iota)L_{M_P} \\
\qquad + [F_{S_P}(1-\xi)\tau_S + S_{S_P}\tau_M](\gamma I_{S_P} + d\gamma N_{S_P}) + [F_{S_N}(1-\xi)\tau_S + S_{S_N}\tau_M](\gamma I_{S_N} + d\gamma N_{S_N}) \\
\\
\frac{dI_{S_P}}{dt} = \sigma\Big([v_0 + \lambda_S(1-x)a]L_{S_P} + \lambda_S(1-x)aL_{M_P}\Big) + \theta N_{S_P} \\
\qquad + [F_{S_N}(1-\xi)(1-\tau_S) + S_{S_N}(1-\tau_M)]\gamma I_{S_N} \\
\qquad - (r + \delta + \delta_I)I_{S_P} - \Big(F_{S_P}[\xi + (1-\xi)\tau_S] + S_{S_P}\tau_M\Big)\gamma I_{S_P} \\
\\
\frac{dN_{S_P}}{dt} = (1-\sigma)\Big([v_0 + \lambda_S(1-x)a]L_{S_P} + \lambda_S(1-x)aL_{M_P}\Big) \\
\qquad + [F_{S_N}(1-\xi)(1-\tau_S) + S_{S_N}(1-\tau_M)]d\gamma N_{S_N} \\
\qquad - (r + \theta + \delta + \delta_N)N_{S_P} - \Big(F_{S_P}[\xi + (1-\xi)\tau_S] + S_{S_P}\tau_M\Big)d\gamma N_{S_P}
\end{cases}
$$

MDR Compartments

$$
\begin{cases}
\frac{dL_{M_N}}{dt} = -\Big(v + \lambda_M(1-x)a + \lambda_S(1-x)[\alpha + (1-a)(1-\iota)] + \delta\Big)L_{M_N} + r(I_{M_N} + N_{M_N}) \\
\qquad + \lambda_M(1-x)(1-a)\iota L_{S_N} + \lambda_M(1-a)S \\
\\
\frac{dL_{M_P}}{dt} = -\Big(v_0 + \lambda_M(1-x)a + \lambda_S(1-x)[\alpha + (1-a)(1-\iota)] + \delta\Big)L_{M_P} + r(I_{M_P} + N_{M_P}) \\
\qquad + \lambda_M(1-a)(1-x)\iota L_{S_P} \\
\qquad + (F_{M_P}\tau_S R + S_{M_P}\tau_M)(\gamma I_{M_P} + d\gamma N_{M_P}) + (F_{M_N}\tau_S R + S_{M_N}\tau_M)(\gamma I_{M_N} + d\gamma N_{M_N}) \\
\\
\frac{dI_{M_N}}{dt} = \sigma\Big([v + \lambda_M(1-x)a]L_{M_N} + \lambda_M(1-x)aL_{S_N} + \lambda_M aS\Big) + \theta N_{M_N} \\
\qquad - (F_{M_N} + S_{M_N})\gamma I_{M_N} - (r + \delta + \delta_I)I_{M_N} \\
\\
\frac{dI_{M_P}}{dt} = \sigma\Big([v_0 + \lambda_M(1-x)a]L_{M_P} + \lambda_M(1-x)aL_{S_P}\Big) + \theta N_{M_P} \\
\qquad + [F_{M_N}(1-\tau_S R) + S_{M_N}(1-\tau_M)]\gamma I_{M_N} + \xi(F_{S_N}\gamma I_{S_N} + F_{S_P}\gamma I_{S_P}) \\
\qquad - (F_{M_P}\tau_S R + S_{M_P}\tau_M)\gamma I_{M_P} - (r + \delta + \delta_I)I_{M_P} \\
\\
\frac{dN_{M_N}}{dt} = (1-\sigma)\Big([v + \lambda_M(1-x)a]L_{M_N} + \lambda_M(1-x)aL_{S_N} + \lambda_M aS\Big) \\
\qquad - (F_{M_N} + S_{M_N})d\gamma N_{M_N} - (r + \theta + \delta + \delta_N)N_{M_N} \\
\\
\frac{dN_{M_P}}{dt} = (1-\sigma)\Big([v_0 + \lambda_M(1-x)a]L_{M_P} + \lambda_M(1-x)aL_{S_P}\Big) \\
\qquad + [F_{M_N}(1-\tau_S R) + S_{M_N}(1-\tau_M)]d\gamma N_{M_N} + \xi(F_{S_N}d\gamma N_{S_N} + F_{S_P}d\gamma N_{S_P}) \\
\qquad - (F_{M_P}\cdot\tau_S R + S_{M_P}\tau_M)d\gamma N_{M_P} - (r + \theta + \delta + \delta_N)N_{M_P}
\end{cases}
$$

Parameter	Definition
β	Standard incidence rate of SSpos individuals (per year)
c	Relative infectiousness of SSneg individuals
ζ	Relative fitness of the MDR Strain
x	Degree of protection provided by a latent infection
α	Proportion of infections that immediately develop active TB
v	Latent progression rate (per year)
v_0	Relapse rate (per year)
σ	Proportion of people entering the active TB states who develop SSpos TB
θ	Sputum smear conversion rate
γ	SSpos treatment rate (per year)
d	Relative treatment rate of SSneg individuals (per year)
η_S	Proportion of diagnosed drug susceptible cases linked onto care
η_M	Proportion of diagnosed MDR cases linked onto care
Ψ_N	Proportion of treatment naive cases that take a DST
Ψ_P	Proportion of previously treated cases that take a DST
Sp_M	Specificity of the drug susceptibility test
Se_M	Sensitivity of the drug susceptibility test
δ	Natural death rate (per year)
δ_I	SSpos death rate (per year)
δ_N	SSneg death rate (per year)
$\Lambda(t)$	Birth Rate

Table 2.4: The TIME model parameters

2.3.1 Screening Rates

To understand how screening rates for the active TB populations depend on the various parameters in the TIME model we need to consider a joint exponential process which models individuals leaving a compartment through different processes. Given $Y_1, Y_2, \ldots, Y_n$ are independent exponential random variables which respectively determines the time until an individual leaves the compartment through process i (i=1,2,...,n), we would like to calculate the mean time an individual spends in the compartment conditioned on the fact that they leave through treatment. By defining $Y_i \sim \exp(\mu_i)$ for each i, the random variable $X_k = \min_{i \neq k}(Y_i)$ will have the cumulative distribution function:

$$F_{X_k}(x) = P(X_k \leq x) = 1 - P(X_k > x) = 1 - \prod_{i \neq k}^{n} P(Y_i > x) = 1 - e^{-\sum_{i \neq k}^{n} \mu_i x}$$

and as a result $X_k \sim \exp\left(\sum_{i \neq k}^{n} \mu_i\right)$. It now follows that

$$\begin{aligned} P(Y_k = \min(Y_1, \ldots, Y_n)) &= P(Y_k \leq X_k) \\ &= \int_0^y \int_y^{+\infty} \mu_k e^{-\mu_k y} \cdot \left(\sum_{i \neq k}^{n} \mu_i\right) e^{-\left(\sum_{i \neq k}^{n} \mu_i\right) x} dx dy = \frac{\mu_k}{\mu_1 + \mu_2 + \cdots + \mu_n}. \end{aligned}$$

If $Y := \prod_{i=1}^{n} Y_i$, then by definition the indicator function

$$I(a,b) = \begin{cases} 1 & \text{if } a \leq b \\ 0 & \text{otherwise} \end{cases}$$

we have

$$\begin{aligned} E[Y|I(Y_k, X_k) = 1] &= \int_0^{+\infty} y_k \frac{f_{Y,I(y_k,X_k)}(y_k)}{P(Y_k \leq X_k)} dy_k \\ &= \frac{\mu_1 + \mu_2 + \cdots + \mu_n}{\mu_k} \cdot \int_0^{+\infty} y_k \mu_k e^{-\mu_k y_k} \left(\prod_{i \neq k}^{n} \int_{y_k}^{+\infty} \mu_i e^{-\mu_i y_i} dy_i \right) dy_k \\ &= \frac{\mu_1 + \mu_2 + \cdots + \mu_n}{\mu_k} \cdot \int_0^{+\infty} y_k \mu_k e^{-(\mu_1+\mu_2+\cdots+\mu_n) y_k} dy_k \\ &= \frac{1}{\mu_1 + \mu_2 + \cdots + \mu_n}. \end{aligned}$$

Therefore the mean time an individual spends in the compartment conditioned on the fact that they leave through treatment will always be one over the sum of the outflow rates. To ensure a population is being diagnosed at the desired average rate, the individuals in the population must be screened for TB at twice the conditional mean time.

Table 2.3.1 shows the required screening rates for the drug susceptible treatment naive and previously treated active TB populations in the TIME model under the assumption of perfect DSTs[14]. As an individual's smear status is unknown before they are tested, a convex combination of the SSpos and SSneg screening rates was used to obtain the screening rate for the combined population.

Compartment	Conditional mean time (years)	Screening rate (tests per person per year)
I_{S_N}	$\frac{1}{r+\gamma\eta_s+\delta+\delta_I}$	$\frac{r+\gamma\eta_s+\delta+\delta_I}{2}\sigma + \frac{r+\theta+d\gamma\eta_s+\delta+\delta_N}{2}(1-\sigma)$
N_{S_N}	$\frac{1}{r+\theta+d\gamma\eta_s+\delta+\delta_N}$	
I_{S_P}	$\frac{1}{r+\gamma\eta_s[\xi+(1-\xi)\tau_p]+\delta+\delta_I}$	$\frac{r+\gamma\eta_s[\xi+(1-\xi)\tau_p]+\delta+\delta_I}{2}\sigma$ $+\frac{r+\theta+d\gamma\eta_s[\xi+(1-\xi)\tau_p]+\delta+\delta_N}{2}(1-\sigma)$
N_{S_P}	$\frac{1}{r+\theta+d\gamma\eta_s[\xi+(1-\xi)\tau_p]+\delta+\delta_N}$	

Table 2.5: Screening rates for the drug susceptible active TB populations

[14]This assumes the drug susceptibility tests have a specificity of $Sp_M = 1$ and a sensitivity of $Se_M = 1$.

Compartment	Conditional mean time (years)	Screening rate (tests per person per year)
I_{M_N}	$\frac{1}{r+\gamma[(1-\Psi_N)\eta_s+\Psi_N\eta_m]+\delta+\delta_I}$	$\frac{r+\gamma[(1-\Psi_N)\eta_s+\Psi_N\eta_m]+\delta+\delta_I}{2}\sigma$
N_{M_N}	$\frac{1}{r+\theta+d\gamma[(1-\Psi_N)\eta_s+\Psi_N\eta_m]+\delta+\delta_N}$	$+\frac{r+\theta+d\gamma[(1-\Psi_N)\eta_s+\Psi_N\eta_m]+\delta+\delta_N}{2}(1-\sigma)$
I_{M_P}	$\frac{1}{r+\gamma[(1-\Psi_P)\eta_s R_n\tau_p+\Psi_P\eta_m\tau_m]+\delta+\delta_I}$	$\frac{r+\gamma[(1-\Psi_P)\eta_s R_n\tau_p+\Psi_P\eta_m\tau_m]+\delta+\delta_I}{2}\sigma$
N_{M_P}	$\frac{1}{r+\theta+d\gamma[(1-\Psi_P)\eta_s R_n\tau_p+\Psi_P\eta_m\tau_m]+\delta+\delta_N}$	$+\frac{r+\theta+d\gamma[(1-\Psi_P)\eta_s R_n\tau_p+\Psi_P\eta_m\tau_m]+\delta+\delta_N}{2}(1-\sigma)$

Table 2.6: Screening rates for the MDR active TB populations

2.4 Model Summary

In this chapter we introduced the sputum smear TB model and the TIME compartmental disease model which is an expansion of the sputum smear model that incorporates treatment history and drug resistance.

The size and structure of the sputum smear model allowed us to derive analytical results pertaining to the global dynamics of the system. We proved the disease free equilibrium of the SLI_pI_n model was global stable when $R_0 < 1$. We were also able to prove global stability of the unique endemic equilibrium point when $R_0 > 1$ under an exact parameter constraint. However since the sputum smear model does not included treatment history or drug resistant status, it does not have enough detail to analyze treatment strategies.

In the next chapter we will calibrate the TIME model to the HIV negative TB endemic in South Africa. This calibration will be used to examine incidence and mortality reduction targets set by the World Health Organization.

Chapter 3

South Africa

3.1 Modelling the Underlying HIV Negative TB Endemic

Co-infection with HIV drastically effects the population dynamics of tuberculosis. Because HIV is highly prevalent in South Africa, modelling TB requires an understanding of how the HIV positive and HIV negative TB endemics interact. Upon infection individuals with HIV progress rapidly progress into the active TB phase and without treatment 90%-95% would die within a few months [6]. Since individuals with HIV have a much higher chance of developing sputum smear negative TB [10], on average they will generate far few infections.

In *The Population Biology of Tuberculosis*, Christopher Dye fits a TB model stratified by HIV status to time series data of the HIV TB endemics in South Africa [4, Chapter 6]. The model fits of the time series data in Figure 3.1 indicates a large growth in HIV positive TB cases since 1990, but only a slight increase to the HIV negative TB case. Dye found the HIV positive TB epidemic had a basic reproduction number of 0.43. This indicates that the HIV positive TB incidence could be decreased by decreasing the prevalence of TB in the HIV negative population. *Corbett et. al.* [5] found similar TB incidence trends in the South African gold mines. They observed that the TB incidence in the HIV negative worker population remained constant even as the TB incidence double among HIV positive workers.

In this chapter we will use the TIME model to analyze care and control strategies for the underlying HIV negative TB endemic in South Africa. To model the endemic we assume the dynamics of the HIV negative TB endemic are independent from the HIV positive population. Moreover, we also assume the HIV negative population is in a quasi-static state, so that any changes to the HIV negative population will not effect the TB dynamics.

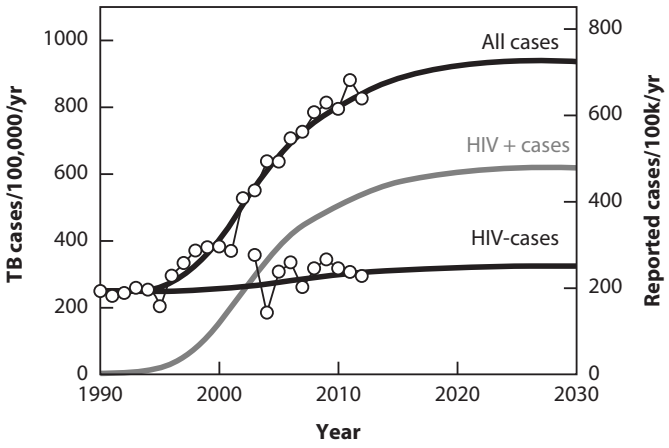


Figure 3.1: Time series model fits for the HIV positive and HIV negative TB endemics in South Africa [4, Chapter 6. Reproduced with permission of Princeton University Press]

Some of the TIME model parameters were provided in the TIME technical appendix. To determine the unknown parameters two computations were performed. The first computation solved a pair of algebraic equations which were based on the case fatality ratios to determine the TB death rates. The second computation minimized a nonlinear objective function to calibrate the remaining 13 parameters at equilibrium to 17 data points from the 2012 HIV negative TB endemic in South Africa. This optimization used Matlab's simulated annealing algorithm [11] to locate the basin of attraction for the global minimum. As the simulated annealing algorithm only converges as $t \to \infty$, this algorithm will not find the global minimum in finite time. Consequently once the simulated annealing algorithm terminated, Matlab's pattern search algorithm was used to locate the basin minimum.

3.1.1 Estimation of TB Death Rates

The least squares objective function (3.1) was used to calibrate the SSpos and SSneg death rates (δ_I and δ_N) against the case fatality ratios for the sputum smear positive and sputum smear negative populations in the absence of treatment.

The case fatality ratio measures the proportion of deaths which occur in a given compartment over the course of the disease. If μ_j is the mean time an individual spends in a compartment j over the course of the disease and δ_j is the death rate attributed to individuals in compartment j, then at equilibrium the case fatality ratio for that compartment is $\mu_j \delta_j$.

To approximate the mean time spent in the SSpos and SSneg compartments in the absence of treatment, we neglect the bilinear transmission terms and consider the finite state Markov chain in Figure 3.2. As the population of the Markov chain scales to infinity then

the average flows between the compartments will match the sputum smear model solution neglecting reinfection (assuming the degree of protection offered by a latent infection is 100%).

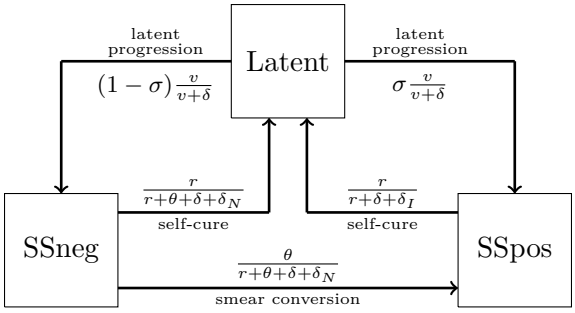


Figure 3.2: The sputum smear Markov Chain for treatment naive individuals in the absence of treatment.

Calculating the expected mean times spent in the SSpos compartment gives,

$$\begin{aligned}
\mu_I &= \frac{1}{r+\delta+\delta_I}\cdot\left[1+\sum_{n=1}^{\infty} P\{\text{Returning } n \text{ times to the SSpos compartment}\}\right] \\
&= \frac{1}{r+\delta+\delta_I}\cdot\left[1+\sum_{n=1}^{\infty}\left(P\left\{\begin{smallmatrix}\text{Transitioning from}\\ \text{SSpos to latent}\end{smallmatrix}\right\}\cdot P\left\{\begin{smallmatrix}\text{Transitioning from}\\ \text{latent to SSpos}\end{smallmatrix}\right\}\right)^n\right] \\
&\overset{1}{=} \frac{1}{r+\delta+\delta_I}\cdot\left[1+\sum_{n=1}^{\infty}\left(\frac{r}{r+\delta+\delta_I}\cdot\frac{\frac{v}{v+\delta}\left[\sigma+(1-\sigma)\frac{\theta}{r+\theta+\delta+\delta_N}\right]}{1-(1-\sigma)\frac{v}{v+\delta}\frac{r}{r+\theta+\delta+\delta_N}}\right)^n\right] \\
&= \frac{1}{r+\delta+\delta_I}\cdot\frac{1}{1-\frac{r}{r+\delta+\delta_I}\cdot\frac{\frac{v}{v+\delta}\left[\sigma+(1-\sigma)\frac{\theta}{r+\theta+\delta+\delta_N}\right]}{1-(1-\sigma)\frac{v}{v+\delta}\frac{r}{r+\theta+\delta+\delta_I}}}
\end{aligned}$$

Similarly the mean time spent in the SSneg compartment is,

$$\begin{aligned}
\mu_N &= \frac{1}{r+\theta+\delta+\delta_N}\cdot\left[1+\sum_{n=1}^{\infty}\left(\frac{r+\theta\frac{r}{r+\delta+\delta_I}}{r+\theta+\delta+\delta_N}\cdot\frac{(1-\sigma)\frac{v}{v+\delta}}{1-\sigma\frac{v}{v+\delta}\frac{r}{r+\delta+\delta_I}}\right)^n\right] \\
&= \frac{1}{r+\theta+\delta+\delta_N}\cdot\frac{1}{1-\frac{r+\theta\frac{r}{r+\delta+\delta_I}}{r+\theta+\delta+\delta_N}\cdot\frac{(1-\sigma)\frac{v}{v+\delta}}{1-\sigma\frac{v}{v+\delta}\frac{r}{r+\delta+\delta_I}}}
\end{aligned}$$

To model the TB endemic in South Africa we assume a life expectancy of 72 years for individuals in the HIV negative population[2]. This gives a natural death rate of $\delta = \frac{1}{72}$. With a value for the natural death rate, assuming only 10% of latent individuals will develop active TB over the course of their lives ($\frac{v}{v+\delta} = 0.1$) gives us a latent progression rate of $v = 0.0015$. The technical appendix of the TIME Impact paper [8] also provides values for

[1] $$P\left\{\begin{smallmatrix}\text{Transitioning from}\\ \text{latent to SSpos}\end{smallmatrix}\right\} = \frac{v}{v+\delta}\sigma + \frac{v}{v+\delta}(1-\sigma)\left[\frac{\theta}{r+\theta+\delta+\delta_N} + \frac{r}{r+\theta+\delta+\delta_N}P\left\{\begin{smallmatrix}\text{Transitioning from}\\ \text{latent to SSpos}\end{smallmatrix}\right\}\right]$$

[2] The WHO estimates a life expectancy of 60 years in South Africa, however this life expectancy includes HIV/AIDs.

the self cure rate ($r = 0.2$), the smear conversion rate ($\theta = 0.015$), and the proportion of infections which develop SSpos TB ($\sigma = 0.45$).

With the above information the mean time an individual spends in the SSpos and SSneg compartments over the course of the disease is now explicitly dependent on the TB death rates. This allowed us to calibrate the TB death rates by minimizing the objective function

$$F(\delta_I, \delta_N) = [\mu_I(\delta_I, \delta_N) \cdot \delta_I - 0.54]^2 + [\mu_N(\delta_I, \delta_N) \cdot \delta_N - 0.15]^2 \tag{3.1}$$

using Matlab's pattern search algorithm. The case fatality ratios for SSpos individuals in the absence of treatment ($\mu_I = 0.54$) and for SSneg individuals in the absence of treatment ($\mu_N = 0.15$) were obtained from The Population Biology of Tuberculosis [4, Appendix 2]. The final resulting death rate for SSpos ($\delta_I = 0.23931$) and SSneg TB ($\delta_n = 0.038346$) were similar to the TB deaths rates used in [4].

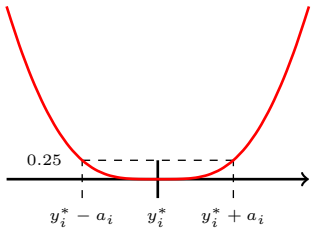
3.1.2 Calibration of the TIME model to the 2012 Endemic

In 2012, 90% of the TB patients in South Africa had a known HIV status and 36% of these cases were HIV negative. As we assumed the population could be partitioned by HIV status without effecting the dynamics of the TB endemic in the HIV negative population, we removed the HIV positive population and adjusted the observed data points appropriately. The only data points which were not adjusted were the MDR prevalence in new cases, the MDR prevalence in retreatment cases and the MDR prevalence in all cases. In these situations we assumed that the MDR prevalences occurred independently from the prevalence of HIV.

To calibrate the remaining TIME model parameters to the HIV negative TB endemic the objective function (3.2) was minimized by using Matlab's simulated annealing and pattern search algorithms. The definitions for (3.2) are documented in Table 3.1. The objective function minimized in the model calibration is a combination of squared normalized residuals and a sum of normalized special functions. The normalized residuals were used so that the order of magnitude of the observations would not bias the minimization. Special functions were used for five data points[3] instead of squared normalized residuals as these data point were estimates which came with an associated 95% confidence interval. For each of these data points we define the function $f_i(\vec{u}) = \left(\frac{y_i(\vec{u})-y_i^*}{a_i-y_i^*}\right)^4 \Big/ \left[3 + \left(\frac{y_i(\vec{u})-y_i^*}{a_i-y_i^*}\right)^2\right]$ where y_i^* is the observed data point, $y_i(\vec{u})$ is the model output for the observed data point, and $(y_i^* - a_i, y_i^* + a_i)$ defines the 95% confidence interval. An illustration of this function is in Figure 3.3.

[3] TB mortality, HIV negative TB incidence, MDR prevalence in new cases, MDR prevalence in retreatment cases, and MDR prevalence in all cases.

By design these $f_i(\vec{u})$ functions are relatively flat inside the 95% confidence interval. As a result, once the optimization algorithm finds a parameter regime in which of one the model output falls inside it's associated 95% confidence interval, a greater emphasis will be directed towards minimizing other quantities in the objective function.



$$f_i(\vec{u}) = \frac{\left(\frac{y_i(\vec{u}) - y_i^*}{a_i - y_i^*}\right)^4}{3 + \left(\frac{y_i(\vec{u}) - y_i^*}{a_i - y_i^*}\right)^2}$$

Figure 3.3: Graph of the special functions used in the model calibration.

The WHO estimated that 65% of treatment naive cases and that 71% of previously treated individuals took a drug susceptibility test in 2015. As these percentages were not documented in 2012, we made the assumption that that the 2012 values would be similar to the values in 2015. To reflect this in the model calibration, the proportion of treatment naive and previously treated cases that take a DST (Ψ_N and Ψ_P) were left as free parameters and their approximate values were included into the objective function.

In 2012, the WHO provided treatment success rates for treatment naive cases (77%), previously treated cases (66%), MDR cases (49%), and HIV positive cases (73%). As the WHO included the development of MDR TB in the definition of treatment failure for first line drugs, the treatment success rate for the treatment naive and previously treated cases are respectively defined to be $(1-\xi)\tau_n$ and $(1-\xi)\tau_p$ where ξ is the proportion of drug susceptible cases linked on to first line care which develop MDR TB. The treatment success rates as well as the ξ parameter were all left as free parameters in the model calibration, and their appropriate values were included into the objective function.

The TIME technical appendix gives a value for the relative fitness of MDR TB, however as this is a difficult parameter to measure ζ was left as a free parameter and the estimated relative fitness was included into the objective function.

With the definitions in Table 3.1 the objective function is defined to be the function

$$\text{Obj}(\vec{u}) = \sum_{i=1}^{5} f_i(\vec{u}) + \sum_{i=6}^{10} \left(\frac{y_i^* - y_i(\vec{u})}{y_i^*}\right)^2 + \underbrace{\sum_{j=1}^{2} \left(\frac{u_j^* - (1-u_{12})u_j}{u_j^*}\right)^2}_{\substack{\text{treatment success rates} \\ \text{for treatment naive and} \\ \text{previously treated cases}}} + \underbrace{\sum_{j=3}^{6} \left(\frac{u_j^* - u_j}{u_j^*}\right)^2}_{\substack{\text{MDR treatment success } (\tau_m) \\ \text{DST testing } (\Psi_P \ \& \ \Psi_N) \\ \text{Relative fitness } (\zeta)}}. \tag{3.2}$$

Measurement	Model Output	Data Value
TB Mortality	$y_1(\vec{u})$	y_1^*
HIV negative TB incidence	$y_2(\vec{u})$	y_2^*
MDR prevalence in new cases	$y_3(\vec{u})$	y_3^*
MDR prevalence in new retreatment cases	$y_4(\vec{u})$	y_4^*
MDR prevalence in all cases	$y_5(\vec{u})$	y_5^*
Confirmed MDR cases	$y_6(\vec{u})$	y_6^*
New SSpos cases	$y_7(\vec{u})$	y_7^*
New SSneg cases	$y_8(\vec{u})$	y_8^*
Relapse cases	$y_9(\vec{u})$	y_9^*
Previously treated cases (excluding relapse)	$y_{10}(\vec{u})$	y_{10}^*
Cases tested for MDR TB (Not included in fit)	$y_{11}(\vec{u})$	y_{11}^*
Treatment success rate for drug susceptible treatment naive cases (τ_n)	u_1	u_1^*
Treatment success rate for drug susceptible previously treated cases (τ_p)	u_2	u_2^*
Treatment success rate for MDR cases (τ_m)	u_3	u_3^*
Proportion of treatment naive cases which take a DST (Ψ_N)	u_4	u_4^*
Proportion of previously treated cases which take a DST (Ψ_P)	u_5	u_5^*
Relative fitness of the MDR strain (ρ)	u_6	u_6^*
Standard incidence rate of SSpos individuals (β)	u_7	
Diagnosis rate of SSpos TB (γ)	u_8	
Relative diagnosis rate of SSneg TB (d)	u_9	
Relapse rate (v_0)	u_{10}	
Proportion of drug susceptible cases linked onto care (η_S)	u_{11}	
Proportion of drug susceptible cases linked onto first line care that develop MDR TB (ξ)	u_{12}	
Specificity of the drug susceptibility test (Sp)	u_{13}	

Table 3.1: Objective function definition

Definitions of the Data Values and Model Output

The WHO estimated 31,000 (Conf. Int. [3,700-86,000]) HIV negative died from tuberculosis in 2012 [14, page 131]. In the model output TB mortality is defined to be the quantity:

$$y_1(\vec{u}) := \delta_I(I_{S_N} + I_{S_P} + I_{M_N} + I_{M_P}) + \delta_N(N_{S_N} + N_{S_P} + N_{M_N} + N_{M_P}).$$

In 2012, the WHO estimated 200,000 (Conf. Int. [160,000-240,000]) new incidence cases of TB [14, page 131]. This estimate includes all new active TB cases arising from: infection, reinfection, cross infection, latent progression, or relapse. The model output definition of incidence is:

$$\begin{aligned} y_2(\vec{u}) := {} & vL_{S_N} + v_0L_{S_P} + \lambda_S a\left[S + (1-x)(L_{S_N} + L_{S_P}) + (1-x)(1-\iota)(L_{M_N} + L_{M_P})\right] \\ & + vL_{M_N} + v_0L_{M_P} + \lambda_M a\left[S + (1-x)(L_{M_N} + L_{M_P}) + (1-x)\iota(L_{S_N} + L_{S_P})\right]. \end{aligned}$$

The National Institute for Communicable Diseases produced a tuberculosis drug resistant survey for 2012-2014. They estimated the MDR prevalence in new cases to be 2.1% (Conf. Int. [1.5%-2.7%]), the MDR prevalence in old cases to be 4.6% (Conf. Int. [3.2%-6.0%]), and the MDR prevalence in all cases to be 2.8% (Conf. Int. [2.0%-3.6%]) [9]. As prevalences were all estimated through cohort studies, these prevalences are all condition on the diagnosed individual taking a drug susceptibility test. The model output is defined such that

$$y_3(\vec{u}) = P\left\{ \begin{array}{c}\text{A new case is}\\ \text{diagnosed with}\\ \text{MDR TB}\end{array} \middle| \begin{array}{c}\text{The new case}\\ \text{took a DST}\end{array} \right\} = \frac{\gamma[(1-Sp_M)I_{S_N} + Se_M I_{M_N}] + d\gamma[(1-Sp_M)N_{S_N} + Se_M N_{M_N}]}{\gamma[I_{S_N} + I_{M_N}] + d\gamma[N_{S_N} + N_{M_N}]}$$

$$y_4(\vec{u}) = P\left\{ \begin{array}{c}\text{A retreatment}\\ \text{case is diagnosed}\\ \text{with MDR TB}\end{array} \middle| \begin{array}{c}\text{The retreatment}\\ \text{case took a DST}\end{array} \right\} = \frac{\gamma[(1-Sp_M)I_{S_P} + Se_M I_{M_P}] + d\gamma[(1-Sp_M)N_{S_P} + Se_M N_{M_P}]}{\gamma[I_{S_P} + I_{M_P}] + d\gamma[N_{S_P} + N_{M_P}]}$$

$$\begin{aligned} y_5(\vec{u}) &= P\left\{ \begin{array}{c}\text{A case is}\\ \text{diagnosed with}\\ \text{MDR TB}\end{array} \middle| \begin{array}{c}\text{The case}\\ \text{took a DST}\end{array} \right\} \\ &= \frac{\gamma[(1-Sp_M)(I_{S_N} + I_{S_P}) + Se_M(I_{M_N} + I_{M_P})] + d\gamma[(1-Sp_M)(N_{S_N} + N_{S_P}) + Se_M(N_{M_N} + N_{M_P})]}{\gamma[I_{S_N} + I_{S_P} + I_{M_N} + I_{M_P}] + d\gamma[N_{S_N} + N_{S_P} + N_{M_N} + N_{M_P}]}. \end{aligned}$$

In 2012 the WHO reported 15,419 laboratory confirmed cases of MDR TB [14, page 131]. To remove the HIV positive case this number was multiplied by reduced by 64% (the percentage of HIV positive TB case). The model outpout for laboratory confirmed cases of MDR TB is defined to be all cases who were tested and diagnosed as having MDR TB:

$$\begin{aligned} y_6(\vec{u}) = {} & (1-Sp_M)\left[\gamma(\Psi_N I_{S_N} + \Psi_P I_{S_P}) + d\gamma(\Psi_N N_{S_N} + \Psi_P N_{S_P})\right] \\ & + Se_M\left[\gamma(\Psi_N I_{M_N} + \Psi_P I_{M_P}) + d\gamma(\Psi_N N_{M_N} + \Psi_P M_{S_P})\right] \end{aligned}$$

In 2012, the WHO reported new pulmonary tuberculosis cases for three different categories: new SSpos cases (119,898), new SSneg cases (63,210), and new smear not done cases (71,421) [14, page 131]. To account for the unknown sputum smear cases in the model fit, the sputum smear not done cases were distributed between the new SSpos cases and the new SSneg cases under the assumption that unknown cases had the same sputum smear distribution as the

known cases. The new sputum smear cases were also reduced by 64% to separate the HIV negative infections. This gives:

$$y_7^* = \left(119,898 + \frac{119,898}{119,898 + 63,210} \cdot 71,421\right) \cdot 0.36 = 59,999$$
$$\text{and} \quad y_8^* = \left(63,210 + \frac{63,210}{119,898 + 63,210} \cdot 71,421\right) \cdot 0.36 = 31,631.$$

For these data points we also have the corresponding model outputs:

$$y_7(\vec{u}) = \gamma(I_{S_N} + I_{M_N}) \quad \text{and} \quad y_8(\vec{u}) = d\gamma(N_{S_N} + N_{M_N}).$$

In 2012, the WHO reported four different types of retreatment cases. The retreatment cases were classified as being a patient previously treated for TB who: (1) is started on a retreatment regimen after previous treatment has failed (treatment after failure), (2) returns to treatment after having previously defaulted[4] (treatment after default); (3) were previously declared cured or treatment complete and is diagnosed with bacteriologically-positive[5] TB (relapse), or (4) do not have a documented treatment outcome (other) [14, Box 3.1 page 30].

As the TIME model does not distinguish between treatment after failure and treatment after default, we defined the previously treated excluding relapse class to combine these two categories. It should be noted that the WHO also defines previously treated excluding relapse class, however, their definition includes individuals who do not have a documented treatment outcome [14, Box 3.2 page 131]. To distribute the unknown cases appropriately we make the assumption that the unknown retreatment cases have the same distribution as the known cases. This gives

$$y_9^* = \left(26,668 + \frac{26,668}{26,668 + 10,811} \cdot 15,007\right) \cdot 0.36 = 13,444$$
$$\text{and} \quad y_{10}^* = \left(10,811 + \frac{10,811}{26,668 + 10,811} \cdot 15,007\right) \cdot 0.36 = \ 5,450.$$

To calculate relapse and previously treated excluding relapse quantities in the model output, we define $f_{S_{pos}}$ and $f_{S_{neg}}$ to be the fraction of inflows due to relapse entering the respective drug susceptible previously treated SSpos and SSneg compartments. We will also need to define $f_{M_{pos}}$ and $f_{M_{neg}}$ to be the fraction of inflows due to relapse entering the respective MDR previously treated SSpos and SSneg compartments. With these definitions, the model output for relapse and previously treated excluding relapse are

$$y_9(\vec{u}) = \gamma(f_{S_{pos}} I_{S_P} + f_{M_{pos}} I_{M_P}) + d\gamma(f_{S_{neg}} N_{S_P} + f_{M_{neg}} N_{M_P})$$
$$\text{and} \quad y_{10}(\vec{u}) = \gamma[(1 - f_{S_{pos}}) I_{S_P} + (1 - f_{M_{pos}}) I_{M_P}] + d\gamma[(1 - f_{S_{neg}}) N_{S_P} + (1 - f_{M_{neg}}) N_{M_P}].$$

[4]The treatment outcome for a patient is classified as defaulted if the patients treatment was interrupted for two consecutive months or more.

[5]This means that the individual is diagnosed as having active TB.

Measurement	Model Output	Data Value	Relative Difference
Mortality (Conf. Int. [3,700 - 86,000])	54,559	31,000	0.76
Incidence (Conf. Int. [16,000 - 240,000])	187,323	200,000	0.0634
New SSpos cases	66,415	59,999	0.1069
New SSneg cases	33,387	31,631	0.0555
Cases tested for MDR (not included in fit)	89,658	13,431	5.6755
Confirmed MDR cases	2,595	5,550	0.5323
Relapse cases	14,132	13,444	0.0512
Previously treated cases excluding relapse	5,558	5,450	0.02
MDR prevalence in new cases (Conf. Int. [0.015 - 0.027])	0.02426	0.021	0.1537
MDR prevalence in old cases (Conf. Int. [0.032 - 0.06])	0.052384	0.046	0.138
MDR prevalence in all cases (Conf. Int. [0.02 - 0.036])	0.028895	0.028	0.0317
Relative fitness of the MDR strain (ζ)	0.730	0.73	0.0005
Proportion of treatment naive cases which take a DST (Ψ_N)	0.748	0.65	0.1507
Proportion of previously treated cases which take a DST (Ψ_P)	0.760	0.71	0.0705
Treatment success rate of second line drugs (τ_m)	0.490	0.49	0.0003
Treatment success rate of first line drugs on drug susceptible treatment naive cases (τ_n)	0.782	0.77	0.0243
Treatment success rate of first line drugs on drug susceptible previously treated cases (τ_n)	0.671	0.66	0.0183
Standard incidence rate of SSpos TB (β)	7.79		
Diagnosis rate of SSpos TB (γ)	0.472		
Relative diagnosis rate of SSneg TB (d)	0.228		
Relapse rate (v_0)	0.0276		
Proportion of drug susceptible cases linked onto care (η_S)	0.215		
Proportion of first line cases which develop MDR TB (ξ)	0.0160		
Specificity of the DST (Sp)	0.984		

Table 3.2: The best fit (0.34245) for the calibration of the TIME model to the HIV negative TB endemic in South Africa.

Source	Parameter	Definition	Point Value
TIME Technical Appendix	c	Relative infectiousness of SSneg TB	0.22
	σ	Proportion of individuals which develop SSpos TB	0.45
	α	Proportion of infections that develop active TB	0.115
	x	Protection provided by a prior infection	0.650
	θ	sputum smear conversion rate	0.015
	r	Self cure rate	0.2
Estimated	v	Latent progression rate $\left(\text{fit so } \frac{v}{v+\delta} = 0.1\right)$	0.015
	δ	Natural death rate (excluding HIV and TB)	0.0139
Fit to case fatality ratios	δ_I	Death rate of SSpos TB (for HIV - individuals)	0.239
	δ_N	Death rate of SSneg TB (for HIV - individuals)	0.0383
Calculated directly from [14, page 131]	τ_m	Proportion of MDR Cases linked onto care	0.421
No Reference	R_n	Relative treatment success of first line drugs on MDR cases	0.05

Table 3.3: Other TIME model parameters.

In the model calibration the mortality, incidence, and MDR prevalences all fit within the given 95% confidence intervals. The proportion of drug susceptible cases linked onto first line care that develop MDR TB also fit within the confidence interval provided by the TIME technical appendix ($0.014 \leq \xi \leq 0.017$).

Florian Marx et al. [7] examined the relapse and reinfection rates of successfully treated cases between 1996 and 2008 for a suburban setting in Cape Town, South Africa. They found that the relapse rate peaked at 3.93% per annum with the 95% confidence interval ranging from 2.35% to 5.96%. The relapse rate obtained from the TIME model calibration ($v_0 = 2.89\%$ per year) fits within this confidence interval.

In the model calibration the confirmed cases of MDR evaluates at approximately half of the observed data point. This discrepancy appears as the number of cases and the number of MDR cases are inconsistent with the estimated MDR prevalences.

In 2012, the WHO also provided the number of diagnosed cases tested for MDR [14, page 131]. This data point was intentionally left out of the calibration. Since re-exposure can cause secondary infections TB is often transmitted via close contacts. This combined with the low prevalence of MDR means that the spread of MDR TB can be controlled effectively through contact tracing. One of the fundamental assumptions of compartmental disease models is that any two individuals in a given compartment are indistinguishable from one another. As a result, by using a compartmental disease models we are assuming that all individuals in a given compartment undergo uniformly random testing. Comparing the model output with the number of diagnosed cases tested for MDR, indicates that the tar-

geted MDR testing in South Africa is approximately 6.7 times more efficient than uniform random testing.

3.2 The 2035 Stop TB Targets

The Stop TB Partnership set targets for 2035 to achieve a 90% reduction in TB incidence and a 95% reduction in TB mortality from the 2015 values [15, Box 2.3. Page 9]. In this section we will use the TIME model to investigate if the 20 year reduction targets can be achieved for the HIV negative population in South Africa.

As individuals who die with both HIV and TB are classified as HIV deaths [14, page 9], the percentage reduction in TB mortality can be examined by studying the HIV negative TB endemic. Moreover since the TB endemic in the HIV negative population drives the HIV positive TB incidence, the percentage reduction in total TB incidence can be roughly approximated by studying the HIV negative TB endemic.

3.2.1 Incidence Reduction

To discuss incidence projections we separate TB incidence into the two subcategories: new incidence and TB recurrence. We define new incidence to be the number of all new infection entering the treatment naive active TB compartments (drug resistant or MDR) as a result of infection, reinfection, or latent progression; and TB recurrence to be the number of previously treated individuals who relapse or undergo reinfection.

We used Matlab's ode45 function to numerically solve the TIME model equations from the 2012 calibration for different diagnosis rates (γ). The incidence was then calculated from the numerical solutions as a function of γ for different periods of time. Since we are interested in examining the percent reductions relative to 2012, these incidence γ functions were normalized with respect to the 2012 calibrated incidence (186,803) and expressed as a percentage.

Figure 3.4 illustrates the percentage of TB incidence (relative to the TB incidence in 2012) for different periods of time as a function of the diagnosis rate. To visualize the long term trends, the percentage of TB incidence at equilibrium as a function of the diagnosis rate is also included into Figure 3.4. The equilibrium incidences were approximated by using Matlab's ode15s function to run the to TIME model equations to equilibrium (10 million years). The steady state incidence plot indicates that if $\gamma \geq 20$, then the TIME model solutions will converge to the disease free equilibrium.

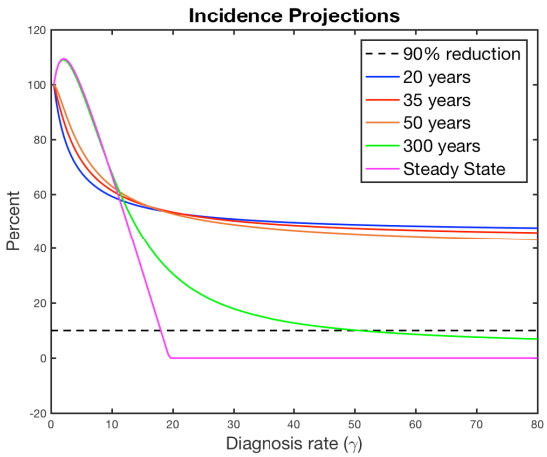
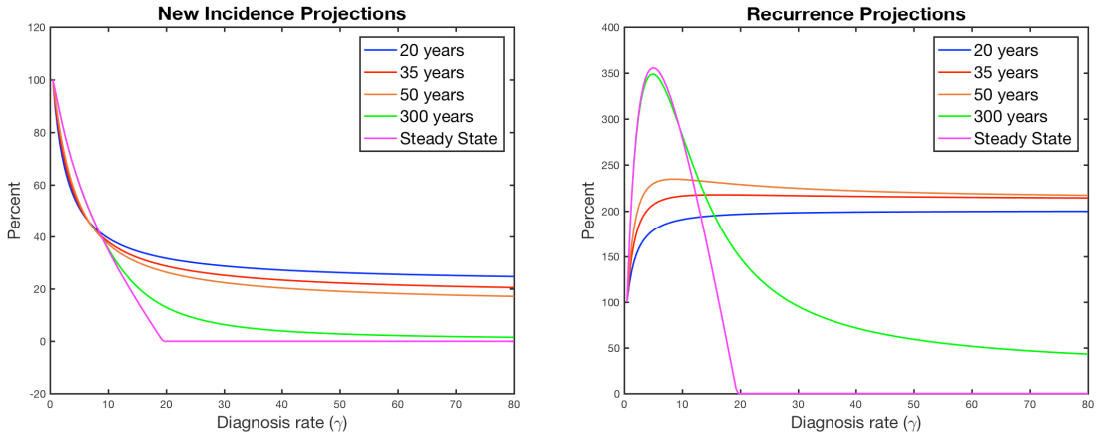


Figure 3.4: Incidence plots (expressed as a percentage relative to the 2012 incidence) for different time periods as functions of the diagnosis rate.

One of the most striking features in Figure 3.4 is that slight increases to γ will cause the incidence to increase in the long term. Increasing the diagnosis rate causes the TB cases fatality ratio to fall, and as a result fewer people are dying from TB. If γ is not sufficiently increased, the long term benefits of having a reduction in new incidence are lost to the growth of TB recurrence. The percentage of new incidence and the percentage of TB recurrence (relative to the new incidence and the TB recurrence in 2012) for different time periods have been plotted as a function of the diagnosis rate in Figure 3.5.



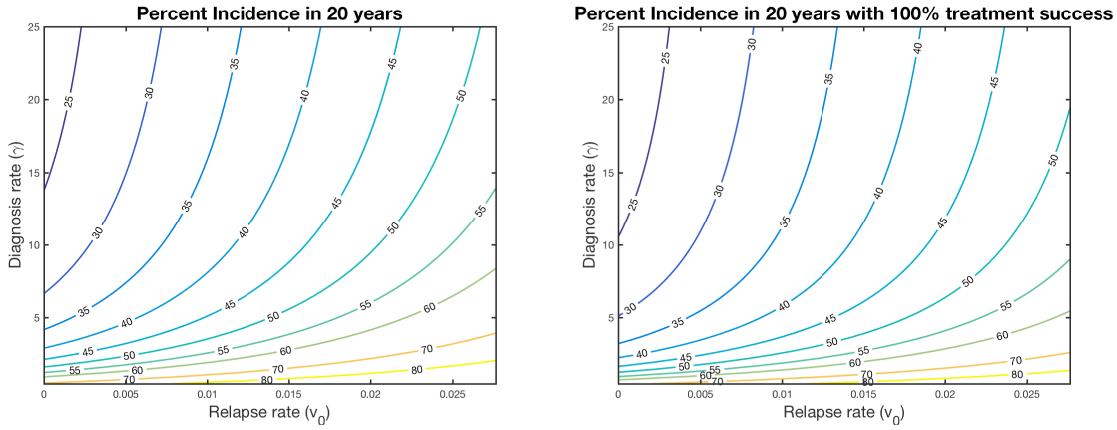
(a) (b)

Figure 3.5: New incidence and TB recurrence plots (expressed as a percentage relative to the new incidence and TB recurrence in 2012) for different time periods as a function of the diagnosis rate.

Figure 3.4 clearly shows the 20 year incidence targets cannot be achieved by only increasing the diagnosis rate. Even over the next 50 years the incidence will not drop below 40% of the calibrated incidence. By increasing the diagnosis rate a larger previously treated population

will develop; consequently this causes an increase in TB relapse. At the current state of the endemic, decreasing TB recurrence requires a high diagnosis rate and significant periods of time for the latent previously treated population to be diminished. With the current definition of incidence this makes hitting the 90% reduction targets in 20 years unrealistic.

To have a greater impact on incidence reduction, targets for the management of TB recurrence need to be established. Decreasing the relapse rate for previously treated latent individuals either through treatment advances or treatment of latent TB will significantly reduce the burden of TB recurrence. Figure 3.6a plots contours of the incidence in 20 years (expressed as a percentage relative to the 2012 incidence) with respect to the diagnosis rate (γ) and the relapse rate (v_0)[6]. In Figure 3.6b the same contours are plotted under the assumption that the treatment success rate for all individuals has been increase to 100%. Comparing these two figures shows that decreasing the relapse rate will have a much larger effect on the 20 year percent incidence reduction than improving treatment success rates.



(a) (b)

Figure 3.6: Contour plots of the incidence in 20 years (expressed as a percentage relative to the 2012 incidence) assuming the current treatment success rate (a) and assuming 100% treatment success rates (b).

Even if the relapse rate is reduced to zero, the incidence will not drop below 15% of the 2012 incidence in 20 years. As improving the diagnosis rate or the relapse rate will not effect treatment naive individuals who currently have latent TB, further declines in incidence require more time for the latent treatment naive population to diminish.

[6]The model calibrated relapse rate for South Africa is $v_0 = 0.0276$.

3.2.2 Mortality Reduction

To investigate if the 95% mortality reduction targets could be achieved in South Africa, we used Matlab's ode45 function to numerically solve the TIME model equation and calculate the percentage of mortality (relative to the mortality in 2012) in 20 years for various interventions.

In this analysis we consider combined interventions between: increasing the diagnosis rate (γ), decreasing the relapse rate (v_0), and increasing the proportion of diagnosed cases linked onto first line care (η_s). For a selection of η_s values between 0.215 (the current η_s value) and 1, a grid of points was created in the $\gamma - v_0$ plane. Once the percentage of mortality in 20 years was calculated at each grid-point, Matlab's contourc function was used to identify the 5% contours. Through combined interventions it is possible for South Africa to achieve a 95% reduction in mortality in 20 years. Figure 3.7 shows the required diagnosis rates and the required relapse rates to achieve a 95% reduction in mortality for different levels of drug susceptible care engagement ($0.215 \leq \eta_S \leq 1$).

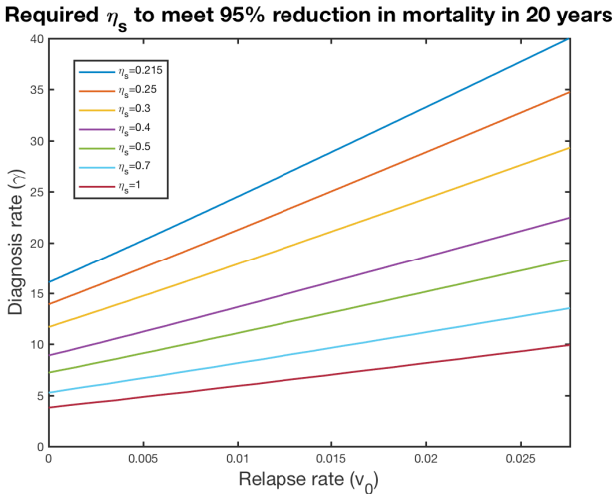


Figure 3.7: Required (γ, v_0) pairs to reach a 95% reduction in mortality for different levels of drug susceptible care engagement.

Chapter 4

Optimizing Care between Drug Susceptible and MDR Cases

4.1 Comparing Drug Susceptible and MDR Care Engagement

In this chapter we investigate optimal control strategies for reducing the HIV negative TB incidence in South Africa over a 20 year time period at a fixed treatment budget. The WHO estimates that cost per patient treated in South Africa is approximately $\$2,000$ US for drug susceptible cases and $\$40,000$ US for MDR cases [15, page 120]. To run the optimizations we assume that treatment costs are permanently set at $\$2,000$ per drug susceptible cases and $\$40,000$ per MDR cases. At the 2012 TIME model calibration this gives South Africa an annual treatment budget of 50.046 million US dollars for drug susceptible cases and 43.610 million US dollars for MDR cases. These costs were calculated by weighting the total number of drug susceptible cases and the total number of MDR cases linked onto care over a period of one year. With this annual budget, treatment for the HIV negative TB cases will cost South Africa 1.90542 billion US dollars over the next 20 years.

	Cost per Case	Current Annual Budget	Total Projected 20 Year Budget
Drug Susceptible	\$2,000 US	\$50.046 million US	\$1.90542 billion US
MDR	\$40,000 US	\$45.225 million US	

Table 4.1: Budget assumptions

As the η_s and η_m parameters determine the proportion of drug susceptible and MDR cases linked onto care, by varying these parameters the treatment budget can be reallocated between funding drug susceptible cases and funding MDR cases.

The constrained optimization for this problem was done through a direct search. Matlab's ode45 function was used to numerically solve the TIME model over a period of 20

years for a grid of (η_s, η_m) pairs in the unit square. At each point the TIME model solutions were integrated using the trapezoid rule and weighted appropriately to produce the total 20 year treatment budgets. The budgets were normalized by 1.90542 billion US as this corresponds to the total projected 20 year budget for South Africa.

$$\begin{aligned}\text{Treatment Budget} := 2000 \int_0^{20} \left(\begin{array}{c} [\Psi_N Sp_M + (1-\Psi_N)](\gamma I_{S_N} + d\gamma N_{S_N}) + [\Psi_P Sp_M + (1-\Psi_P)](\gamma I_{S_P} + d\gamma N_{S_P}) \\ + [\Psi_N (1-Se_M) + (1-\Psi_N)](\gamma I_{M_N} + d\gamma N_{M_N}) + [\Psi_P (1-Se_M) + (1-\Psi_P)](\gamma I_{M_P} + d\gamma N_{M_P}) \end{array} \right) \eta_s \, dt \\ + 40000 \int_0^{20} \left(\begin{array}{c} \Psi_N (1-Sp_M)(\gamma I_{S_N} + d\gamma N_{S_N}) + \Psi_P (1-Sp_M)(\gamma I_{S_P} + d\gamma N_{S_P}) \\ + \Psi_N Se_M (\gamma I_{M_N} + d\gamma N_{M_N}) + \Psi_P Se_M (\gamma I_{M_P} + d\gamma N_{M_P}) \end{array} \right) \eta_m \, dt\end{aligned}$$

Matlab's contourc function was then used to produce budget level sets in the (η_s, η_m) plane. Some of these budget level sets for drug susceptible and MDR care engagements are illustrated in Figure 4.1.

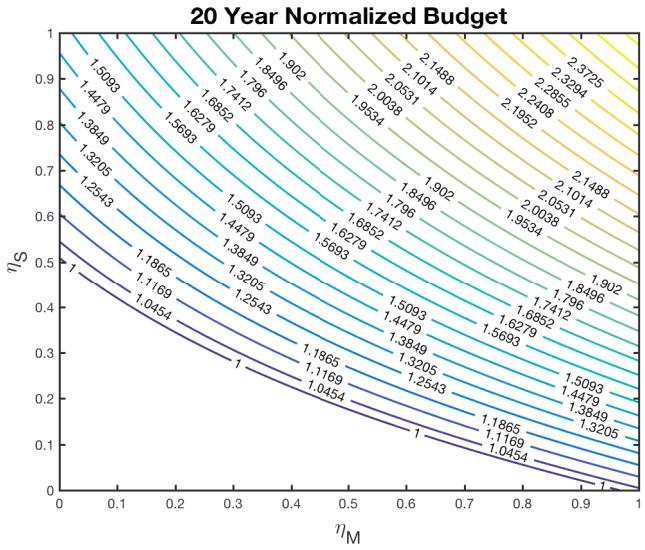
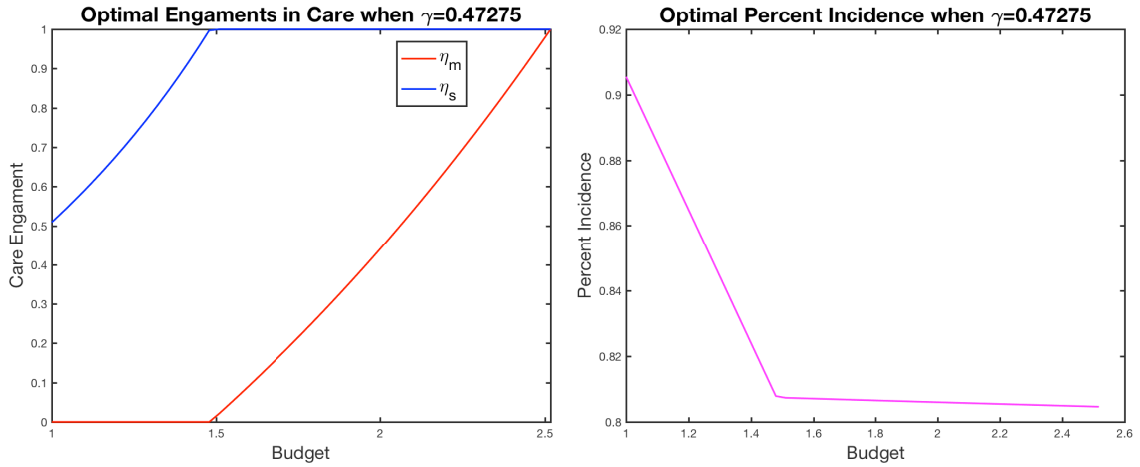


Figure 4.1: Budget level sets

Once the budget level curves were identified, the TB incidence was minimized over each set. The optimal care engagement parameters to minimize incidence in 20 years at a fixed budget are plotted in Figure 4.2a as a function of the normalized budget. The corresponding minimum incidence for each budget (expressed as a percentage relative to the 2012 incidence) is plotted in Figure 4.2b.

Figure 4.2a indicates all available funding should be prioritized into linking the drug susceptible population onto care. Comparing Figure 4.2a and Figure 4.2b shows that directing funding into MDR care engagement has a negligible effect on incidence reduction in 20 years. This is due to the fact that MDR TB has a much lower prevalence than drug susceptible TB in South Africa. Because of the lower MDR prevalence, this bang-bang behaviour between linking drug susceptible and MDR cases onto care should be unaffected by changes in cost

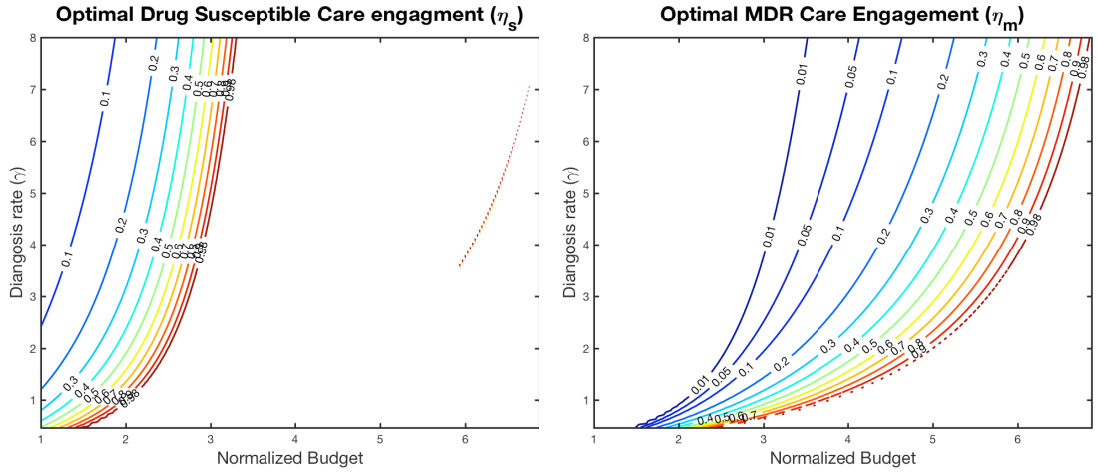
or changes to the treatment parameters provided that the drug susceptible cases remain less expensive to treat.



(a) (b)

Figure 4.2: Optimal care engagement parameters to minimize incidence in 20 years at a fixed budget as a function of normalized budget.

The contour plots in Figure 4.3a and Figure 4.3a display the optimal drug susceptible and MDR care engagement parameters to minimize incidence in 20 years at a fixed budget as a function of the diagnosis rate and the normalized budget. Contour plots of the corresponding minimum incidence (expressed as a percentage relative to the 2012 incidence) are plotted in Figure 4.4.



(a) (b)

Figure 4.3: Optimal care engagements to minimize incidence in 20 years at a fixed budget as functions of the normalized budget and the diagnosis rate (γ).

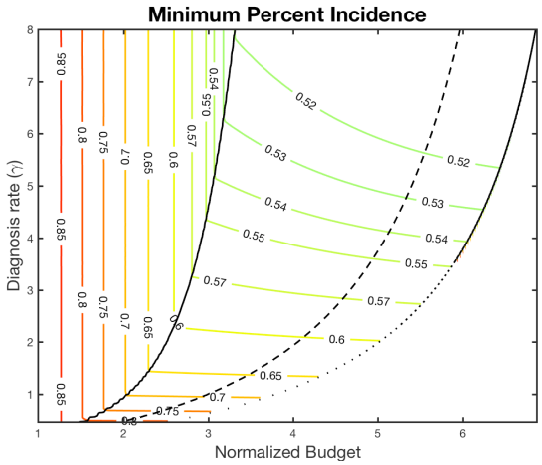


Figure 4.4: Minimum incidence for the corresponding optimal engagement parameters in Figure 4.3.

Figure 4.3a and Figure 4.3b display the same bang-bang behaviour as Figure 4.2a. This bang-bang behaviour breaks γ-budget plane into two distinct regions. The first region is defined to be the set of point where $\eta_m = 0$ and $0 \leq \eta_s < 1$, and the second region is defined to be the set of points where $\eta_s = 1$ and $0 < \eta_m \leq 1$.

The boundaries of these two regions have been included into Figure 4.4 as solid black lines. As $\eta_m = 0$ in the first region, the only way to increase the diagnosis rate at a fixed budget is to decrease η_s so that $\gamma\eta_s$ is held constant[1]. This causes the appearance of the vertical contours in Figure 4.4. Because $\eta_m > 0$ in the second region, decreasing η_m will allow γ to be increased at a fixed budget without decreasing η_s. Comparing Figure 4.4 and Figure 4.3b shows this comprise will improve the drop in incidence over 20 years.

The contour plot of current level of MDR care engagement ($\eta_M = 0.4212$) from figure 4.3b has been plotted in the second region to indicate the minimal cost required to achieve a maximal incidence reduction over 20 years without compromising MDR care engagement. This analysis shows that at least a fivefold increase to the projected 20 year treatment budget would be required to reduce the incidence by 45% in 20 years without compromising MDR care engagement.

[1]When $\eta_m = 0$, then the η_s and γ parameters always appear together as the product $\eta_s\gamma$ in the TIME model.

Chapter 5

Conclusion

In the first part of the thesis we studied a four dimensional compartmental model for tuberculosis. Under the assumption of a constant population we proved that the disease free equilibrium is globally stable in the physically relevant region when the basic reproduction number is less than unity. If the model parameters satisfy an exact condition when the basic reproduction number is greater than one, then the unique endemic equilibrium point contained within the physically relevant region is globally stable in the physically relevant region.

In the second part of the thesis we used the Tuberculosis Incidence and Mortality Estimates model to analyze the TB endemic in South Africa. The TIME model is a compartmental disease model which captures many of the dynamic processes important to TB (variation in the latent progression rate, secondary reinfections, and sputum smear status). As the TIME model is also stratified treatment history and drug resistant status, it can be used to realistically simulate control measure for tuberculosis.

Two computations were performed to determine the TIME model parameters. The first computation solved a pair algebraic equations to determine the deaths rates for the sputum smear positive and the sputum smear negative populations. The second computation minimized a nonlinear objective function to calibrate the TIME model at equilibrium to the 2012 HIV negative TB endemic in South Africa. This calibration was constructed to be overdetermined.

The calibrated TIME model was used to analyze the Stop TB incidence and mortality reduction targets. Due to post-treatment relapse and latent progression, South Africa cannot achieve a 90% incidence reduction in 20 years. If the relapse rate is not improved then even over the next 50 years, South Africa will not be able to achieve a 60% reduction in incidence through increasing the diagnosis rate. If the relapse rate is improved, combined

interventions could reduce incidence by up to 80%.

Although the Stop TB incidence reduction targets cannot be met for the HIV negative population in South Africa, we found that through combined interventions TB mortality could be reduced by 95% over the next 20 years. With the current relapse rate this would require increasing the diagnosis rate and the proportion of drug susceptible cases linked onto care so that $\gamma\eta_s \approx 9$. This means South Africa would need to screen individuals for TB approximately every 2.6 months.

To optimize the reduction in incidence over 20 years at a fixed budget, funding needs to be directed into linking drug susceptible cases (over MDR cases) onto care. Reducing incidence by 45% in the next 20 years without compromising MDR care engagement will cost South Africa five times the current projected 20 year treatment budget. This does does not included the cost of screening programs.

5.1 Model Limitations and Future Work

As the HIV negative endemic is independent from the HIV positive population we were able to accurately analyze TB control strategies for the HIV negative population in South Africa. Moreover since the TB endemic in the HIV negative population drives the HIV positive TB endemic, the incidence reductions achieved in the HIV negative population should reflect incidence reduction in the HIV positive population. To quantify the projected incidence reductions in the HIV positive population, the TIME model would have to be expanded to include the HIV dynamics. By expanding the TIME model to include the HIV dynamics, the steady state assumption may no longer be valid. Thus modelling the co-endemic could require fitting the TIME model to time series data.

A drawback of the TIME model is that the treatment process is assumed to be instantaneous. This could be improved incorporating new compartments for individuals on treatment. By making this expansion to the TIME model individuals would be less likely to relapse multiple time in a single year as treatment takes approximately six months for drug susceptible individuals. Incorporating treatment stages into the model, would allow us to simulate targeted MDR testing for individuals failing first line treatment.

Bibliography

[1] Abraham Berman and Robert J. Plemmons. *Nonnegative Matrices in the mathematical Sciences*. Academic Press, Inc., New York, 1979.

[2] Castillo Chavez, Feng Zhilan, and Huang Wenzhang. *On the computation of R_0 and its role on global stability*, volume 125 of *The IMA Volumes in Mathematics and its Applications*, pages 229–22. Springer-Verlag New York, 233 Spring Street, New York, NY 10013, USA, 1 edition, 2002.

[3] O. Diekmann, J.A.P. Heesterbeek, and J.A.J. Metz. On the definition and the computation of the basic reproduction ratio r_0 in models for infectious diseases in heterogeneous populations. *Journal of Mathematical Biology*, 28:365–382, 1990.

[4] Christopher Dye. *The Population Biology of Tuberculosis*, volume 54 of *Monographs In Population Biology*. Princeton University Press, 41 William Street, Princeton New Jersey, 08540, 2015.

[5] Corbett et. al. Morbidity and mortality in south african gold miners: Impact of untreated disease due to human immunodeficiency virus. *Clinical Infectious Disease*, 34:1251–1258, 2002.

[6] Corbett et. al. The growing burden of tuberculosis - global trends and interactions with the HIV epidemic. *Archives of Internal Medicine*, 163:1009–1021, 2003.

[7] Florian M. Marx et. al. The temporal dynamics of relapse and reinfection tuberculosis after successful treatment: A retrospective cohort study. *Clinical Infectious Diseases*, 58:1676–1683, 2014.

[8] R. M. G. J. Houben et. al. Time impact- a new user-friendly tuberculosis model to inform TB policy decisions. *BioMed Central Medicine*, 22:431–440, 2016.

[9] The National Institute for Communicable Diseases. South african tuberculosis drug resistance survey 2012-14. Technical report, National Health Laboratory Service, Johannesburg, 2014.

[10] R. K. Gupta, S. D. Lawn, L-G. Bekker, J. Caldwell, R. Kaplan, and R. Wood. Impact of human immunodeficiency virus and cd4 count on tuberculosis diagnosis: analysis of city-wide data from cape town, south africa. *The International Journal of Tuberculosis and Lung Disease*, 17:1014–1022, 2013.

[11] Alexander G. Nikolaev and Sheldon H. Jacobson. *Simulated Annealing*, volume 146 of *International Series in Operations Research and Management Science*, chapter 1,

pages 1–40. Springer-Verlag New York, 233 Spring Street, New York, NY 10013, USA, 2010.

[12] Osvaldo Osuna, Joel Rodriguez-Ceballos, Cruz Vargas-De-Leon, and Gabriel Villasenor-Aguilar. A note on the existence and construction of dulac functions. *Nonlinear Analysis: Modelling and Control*, 22:431–440, 2017.

[13] Lawrence Perko. *Differential Equations and Dynamic Systems*, volume 7 of *Texts in Applied Mathematics*. Springer-Verlag New York, Inc., 233 Spring Street, New York, NY 10013, USA, 3 edition, 2001.

[14] Global Tuberculosis Programme. Global tuberculosis report 2013. Technical report, The World Health Organization, Geneva, 2013.

[15] Global Tuberculosis Programme. Global tuberculosis report 2016. Technical report, The World Health Organization, Geneva, 2016.

[16] P. van den Driessche and James Watmough. Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission. *Mathematical Biosciences*, 180:29–48, 2002.