UNIVERSITY OF SOUTHAMPTON

FACULTY OF ENGINEERING, SCIENCE & MATHEMATICS

School of Mathematics

Investigating Household Interventions for Controlling Tuberculosis using Discrete Event Simulation

by

Georgina R. Mellor

Thesis for the degree of Doctor of Philosophy

October 2007

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF ENGINEERING, SCIENCE AND MATHEMATICS

SCHOOL OF MATHEMATICS

Doctor of Philosophy

INVESTIGATING HOUSEHOLD INTERVENTIONS FOR CONTROLLING TUBERCULOSIS USING DISCRETE EVENT SIMULATION

Georgina R. Mellor

Current control methods for Tuberculosis (TB) disease have failed to keep pace with the TB epidemics which have been particularly affected by the Human Immunodeficiency Virus (HIV) epidemic. There is still relatively little known about the interactions between HIV and TB and therefore TB control strategies that are effective in high HIV prevalent settings still need to be identified. The current policy is that active case-finding for adults living in endemic TB settings is ineffective, because transmission events between casual contacts greatly outnumber household transmission events. This policy was developed in an era of low HIV prevalence and the impact of the HIV epidemic on the relative importance of household versus community transmission has not been fully assessed.

The majority of mathematical models used to describe the epidemiology of TB and investigate methods of control have been deterministic compartmental models that have considered only homogeneous mixing. This thesis describes a discrete event simulation model that includes the effect of household structure on the transmission dynamics of TB. It is used to evaluate the effectiveness of targeted case-finding interventions in controlling TB in HIV prevalent populations.

Contents

List of Figures

List of Tables

Acknowledgements

I would firstly like to thank my supervisor Dr Christine Currie for the assistance, advice and support she has given freely and generously throughout the course of my PhD study.

I would also like to thank Professor Russell Cheng for his knowledge and guidance throughout my PhD and for all the kindhearted support.

I am grateful to EPSRC who have funded me throughout this research and to Dr Liz Corbett who through funding from WHO, SACEMA, CREATE and the Wellcome Trust has funded much of the training, feedback and expert advice gained from my travels to various meetings and conferences, which has enabled this research to be so rewarding.

Many people at the London School of Hygiene and Tropical Medicine have helped me through the course of my PhD, especially Dr Liz Corbett. Thanks also to Dr Azra Ghani, Dr Richard White and Dr Emilia Vynnycky for their invaluable advice and assistance with infectious disease modelling.

I would like to thank Dr Brian Williams of the WHO for his help with the Chapter on the parametric model of TB and HIV in Kenya.

I would also like to thank Andrew Drake, Honora Smith, Lanting Lu, Katy Hoad and Simon Doherty with whom I have had such a great time sharing an (abnormally extreme temperatured) office with... I will miss our many interesting conversations and all the hugs!

I'd like to give my recognition and thanks to all my family and friends for their continued love and support and to Andrew Drake, Anna Taylor, Tom Powell, Jo Clarke and especially Sally Mellor for ploughing through and proof-reading my thesis.

Last but not least, I would very much like to thank my mum Anna Taylor and best friend Beth Burton, whose continued encouragement, patience and good humour throughout my time of study has been absolutely invaluable to me.

Abbreviations

100k	100,000
AIDS	Acquired Immune Deficiency Syndrome
ANC	Antenatal Clinic
BCG	Bacille Calmette-Gurin
CDF	Cumulative Distribution Function
CDR	Case Detection Rate
CI	Confidence Interval
CREATE	The Consortium to Respond Effectively to the AIDS-TB Epidemic
DCM	Determinisitic, Compartmental Model
DES	Discrete Event Simulation
DOTS	Directly Observed Treatment, Short-course
EP	Extrapulmonary
GPS	Global Positioning System
HIV	Human Immunodeficiency Virus
IPT	Isoniazid Preventive Therapy
LSHTM	London School of Hygiene and Tropical Medicine
MDR	Multi Drug Resistant
MLE	Maximum Likelihood Estimation
NASCOP	Kenyan National AIDS and STD Control Programme
NLTCP	Kenyan National Leprosy and TB Control Programme
OOP	Object-Orientated Programming
PDF	Probability Density Function
Popn	Population
RNTCP	Revised National Tuberculosis Control Programme
SARS	Severe Acute Respiratory Syndrome
SS-	Sputum Smear Negative
SS+	Sputum Smear Positive
STD	Sexually Transmitted Disease
STI	Sexually Transmitted Infection
TB	Tuberculosis
TST	Tuberculin Skin Testing
UK	United Kingdom of Great Britain and Northern Ireland
UNAIDS	Joint United Nations Programme on HIV/AIDS
USA	United States of America
WHO	World Health Organisation

Chapter 1

Introduction

Tuberculosis (TB) is the leading cause of death among people with a Human Immunodeficiency Virus (HIV) infection, accounting for up to a third of Acquired Immune Deficiency Syndrome (AIDS) deaths worldwide. The epidemic of HIV has led to a dramatic resurgence of TB in sub-Saharan Africa, especially in east and southern Africa where TB notification rates have increased by over three times in the last 20 years [133]. It is increasingly clear that controlling TB in much of Africa depends on the extent to which HIV is brought under control and that to ensure improvements in the care of HIV-positive people, it is essential that they are examined for TB and treated appropriately.

The risk of developing TB disease increases as people progress from HIV infection to AIDS and although there have been studies carried out to determine the extent to which HIV-infection affects transmission of TB [46], there is still relatively little known about the interactions between these two diseases. Therefore TB control strategies that are effective in high HIV prevalent settings still need to be identified. The current policy of active case-finding for adults living in endemic TB settings is ineffective, because transmission events between casual contacts greatly outnumber household transmission events [143]. This policy was developed in an era of low HIV prevalence and the impact of the HIV epidemic on the relative importance of household versus community transmission has not been fully assessed.

The ultimate aim of this research is to develop a discrete event simulation (DES) model of TB transmission in Harare, Zimbabwe, which will allow a full assessment of the effectiveness of contact-tracing and case-finding strategies in high HIV prevalent populations. The model will make use of information regarding household size and structure and therefore the relative importance of household

versus community transmission will be fully assessed. The research is being done in collaboration with DETECTB, an organisation currently carrying out a large population-based trial in Harare, Zimbabwe. The data from this trial will be used to aid the development of the model.

1.1 Research Objectives

This thesis concerns using mathematical modelling to study the role of household versus community transmission of TB to understand the relative effectiveness of household interventions in controlling TB in HIV prevalent populations. The objectives are:

1. To develop a mathematical model of TB transmission and disease in Harare, Zimbabwe, in order to enable accurate simulation of possible active case-finding strategies for TB control in HIV prevalent populations;
2. To enable comparative projections of the likely impact of possible strategies relative to one another to allow a full assessment of the effectiveness of different contact-tracing and case-finding strategies in HIV prevalent populations;
3. To predict how variable population attributes are likely to affect the relative success of different interventions;
4. To determine the relative effectiveness of household interventions in controlling TB in HIV prevalent populations.

1.2 Modelling Approach and Methodology

The method chosen to solve the model is discrete event simulation. This will take individuals through time and incorporate the heterogeneity which leads to the variability observed in each individual's TB disease evolution. An individual's risk of infection, risk of disease progression, infectious period, and infectivity will all be affected by attributes such as the individual's age, sex, HIV status and the number of people in the individual's household. The model will provide recommendations for reducing TB incidence in HIV prevalent areas.

1.3 Summary of Findings

This research answers many questions regarding the interactions that exist between TB and HIV, the way in which household interventions for infectious diseases can be modelled and the relative effectiveness of household interventions in controlling TB in HIV prevalent populations. In summary, we conclude:

- The interactions between TB and HIV can be quantified such that it is possible to accurately estimate HIV prevalence from TB notification rates, as demonstrated with district level data in Kenya.
- Stochastic individual-based modelling is a natural methodology to use to precisely specify household or community transmission of disease, and to model the impact of local interventions.
- Targeting TB-diseased households has a relatively small effect on the TB epidemic due to the small proportion of households that are visited as part of this intervention.
- The clustering of TB disease and HIV infection in households means that targeting HIV-infected households is the most effective active-case finding strategy.
- A much larger proportion of the population needs to be reached in a untargeted or community-wide intervention in order for it to have the same benefit as targeting HIV-infected households.
- Interventions in which TB-diseased or HIV-infected households are targeted are more effective than community-wide interventions, suggesting that household transmission of TB (especially within HIV-infected households) is important.

1.4 Overview of Thesis Structure

The main body of this thesis is divided into 10 chapters which are organised to follow the logical development of how household interventions for TB control in Zimbabwe were investigated through developing a discrete event simulation model. This thesis can be divided into three parts: Chapters 2 to 5 set the background context for this work; Chapters 6 and 7 describe the research development

itself; and Chapters 8, 9 and 10 consist of the discussion of the findings, conclusions and recommendations for further work.

This first chapter gives an introduction to the problem, the problem solving approaches used, a summary of the findings and sets the scene for what follows. Chapter 2 gives the epidemiological background of TB and HIV, discusses current TB control strategies and explains the data being collected by DETECTB in Harare. A comprehensive literature review has been carried out to better understand previous models of infectious diseases (specifically TB), and to appreciate the modelling and epidemiological issues which still need to be addressed. This is given in Chapter 3. Chapter 4 describes a simple parametric model which has been developed to help understand and quantify the relationships and interactions between HIV and TB. This was followed by the investigation of different model configurations in order to establish a possible representation of the natural history of TB to be used by the discrete event simulation (DES) model. This is described in Chapter 5.

The DES model that differentiates between within-household transmission and random transmission of TB is discussed in Chapters 6 and 7.

Finally we discuss its results, undertake a sensitivity analysis and various scenario analyses in Chapters 8 and 9, and then draw some conclusions and make some suggestions for future work in Chapter 10. A glossary has been included to explain the epidemiological terms used throughout the report.

Chapter 2

Background

This Chapter aims to provide an epidemiological background of both TB and HIV and explains the dual epidemic. This Chapter will also discuss current TB control strategies and explain the data being collected by our collaborators, DETECTB, in Harare.

2.1 HIV

HIV (Human Immunodeficiency Virus) is the virus that leads to AIDS (Acquired Immune Deficiency Syndrome). AIDS was first reported in the USA in 1981 and since then more than 20 million people with HIV/AIDS have died [73]. It has become a major worldwide epidemic. A total of 3.1 million people died of HIV/AIDS related causes in 2005 alone [80] and 40.3 million people worldwide were estimated to be living with HIV/AIDS at the end of 2005 [80] with approximately two-thirds of these people living in Sub-Saharan Africa. Over time, the HIV virus weakens the body's ability to fight infections and cancers by progressively destroying cells of the body's immune system. It is only after the immune system is significantly weakened (which can take many years) that people with HIV will get "opportunistic" infections which are caused by microbes such as viruses or bacteria. These infections would not usually cause a healthy person to become sick. However, for someone with advanced HIV the diseases are life threatening. Once somebody begins to get these infections, they are said to have AIDS [3].

HIV is transmitted by blood, semen, preseminal fluid, vaginal fluid or breast milk of an HIV-infected person. The American Social Health Association [165] explains: a person can get HIV when one of these fluids enters the body by way of

the soft skin found in any opening of the body (mucous membranes) or the blood stream. The most common way of spreading HIV is by having unprotected sex with an HIV-infected partner where the virus enters the body through the lining of the vagina, vulva, penis, rectum or mouth. Other ways include contact with infected blood; through blood transfusion for example (in developing countries blood is not routinely screened) or from injecting drugs using shared needles or syringes. HIV infected women can also transmit HIV to their babies during pregnancy, birth or breast feeding.

2.2 Tuberculosis

Tuberculosis (TB) is the most common major infectious disease today [201] and causes more adult deaths worldwide that any other infectious disease [31]. It produces nine million new cases of active disease annually [201] and infects one third of the world's population [202]. "Tuberculosis (TB) is an infectious disease caused by either mycobacterium tuberculosis or mycobacterium bovis. Because these organisms are so similar, the infections they cause are given the one name - tuberculosis." [115]

A person can be infected with mycobacterium tuberculosis but not have active TB disease. This means that the TB mycobacterium are present in the body but that they are not actively causing damage to body tissues because the immune system has "walled them off". The infection can lie dormant for years and often only develops into "active" TB when the immune system is weakened.

When a person develops active disease, the TB organisms are growing and causing damage within the body. TB disease most commonly affects the lungs where it is called pulmonary TB. Seventy-five percent or more of infected people have pulmonary TB. Symptoms of the disease include a prolonged cough of more than three weeks duration, chest pain, fever, chills, appetite loss, weight loss and fatigue. Extra-pulmonary sites include the central nervous system, bones, joints and the lymphatic system. TB skin sores can develop when an infected lymph gland bursts, but this is very uncommon. TB is only infectious if the disease is in the lungs or if a TB skin sore is left uncovered.

TB is an airborne infection and so it is transmitted by the infected person expelling TB mycobacterium into the air by coughing, sneezing, talking or spitting and another person can then inhale these organisms and become infected. However, it usually takes many hours (or even days) of exposure for a previously non-infected person to become infected and so the transmission probability is low.

Investigations into the transmission of TB on aeroplanes in the United States has concluded that less than 1% of all those exposed to tuberculosis would become infected [96]. Furthermore, only about 10% of infected persons with normal immune systems will develop "active" TB in their lifetime [202].

2.3 The Natural History of Tuberculosis

Individuals who are not infected with tuberculosis are said to be susceptible. When they become infected, they are said to become latent. Individuals will then follow one of two routes: develop active disease quickly (usually defined as within five years) or retain a latent TB infection that may possibly reactivate years later. Which route a person will take depends on the effectiveness of their immune system and therefore factors such as age and poverty play a large role, and immunocompromised individuals such as those who are HIV-positive are likely to develop active disease more quickly than immunocompetent individuals.

When a person is latently infected, they can progress to active disease in two ways: reactivation or reinfection. Reactivation occurs when the immune system is weakened and the defence against the TB infection becomes inadequate and so the TB mycobacterium are able to cause damage (active disease). Reinfection occurs when someone with latent TB is infected again but they do not invoke a successful immune response and therefore progress to active disease quickly. Previous infection with TB does confer some immunity to developing active disease and therefore those that have already been infected with TB have a greater immunity to reinfection. The only exception to this is if the latent individual is also HIV positive, in which case reinfection is more likely.

Active TB disease can be infectious or non infectious, with 46% of individuals (27% of late-stage HIV-positives) developing infectious disease [197]. Therefore, HIV-positive individuals are more likely to have non infectious TB. Once a person has active disease then they may self cure, die or be diagnosed and treated. For those with infectious disease, the time until their cure, death or treatment determines an individual's duration of infectiousness, which in turn determines how many people the individual is likely to infect. It is thought that an infectious person will infect from 10 to 15 people every year [202] [160].

A person receiving treatment will either fail or successfully complete the treatment course. Those that successfully complete treatment still retain a TB infection but the TB mycobacterium are no longer actively causing damage. They therefore

have a latent infection, which as previously discussed, means that they are susceptible to reinfection from an infectious person, although they have an increased immunity compared to the uninfected population.

When individuals are said to fail treatment this indicates that they failed to correctly complete the course of drugs and therefore the TB lesions were not sterilized. Failed treatment means that the individuals have active disease. Inconsistent or partial treatment of TB can be very dangerous because it can cause the strain of TB to become drug resistant. The emergence of drug resistant TB throughout the world has caused many complications and is impeding the control of tuberculosis worldwide. A particularly dangerous form of drug resistant TB is multi drug resistant (MDR) TB, which is the disease caused by TB bacilli that are resistant to the two most powerful anti-TB drugs. Although drug resistant TB is treatable, it requires extensive chemotherapy (up to two years of treatment) and is often prohibitively expensive [202].

The natural history of tuberculosis is further complicated by its propensity to be age dependent. As previously mentioned, age is a particularly important variable in determining the risk of developing disease after initial infection with TB and in determining the type of disease developed. Older people are more likely to develop pulmonary TB [122]. Children under the age of 15 show a markedly different reaction to adults. There is a very high risk of developing the disease if infected whilst in the period of infancy which then tails off into a low risk phase from 5 to 10 years old, increasing again in adolescence and adulthood. This implies that although children between 5 and 15 can be infected with TB, they will rarely develop active TB disease but will just maintain a latent infection. A separate point is that children tend to have non infectious TB and even if they have infectious TB they are too small to generate enough of a cough to effectively aerosolise the bacteria. For these reasons, children are very poor onward transmitters and therefore play a small role in the transmission of TB.

2.4 HIV and TB: A Dual Epidemic

As HIV progressively destroys the immune system, there is a greater chance of a person infected with HIV developing TB [126]. This overlap between the epidemics is particularly important and it is increasingly being recognised that in order to successfully fight AIDS, it means fighting TB.

TB is one of the leading causes of illness and death amongst AIDS sufferers in developing countries. Up to 50% of people with HIV develop TB in Sub-Saharan

Africa and one in three die from it. Estimates for 2003 put the number of incident TB cases at 8.8 million, up from an estimated 8.3 million in 2000, with HIV being the main driving force [134]. Adult HIV prevalence rates are now 20% or higher in six southern African countries. In these countries TB case notification rates have increased 2 to 5 fold since 1990 and now between 460 and 720 people develop active TB disease per 100,000 members of the population per year [131]. It has been estimated that approximately 31% of new TB cases in adults in the Africa region can be directly attributed to HIV [47]. It is clear that the spread of the HIV epidemic has significantly impacted the TB epidemic and it is thought that one-third of the increase in TB cases over the last five years can be attributed to the HIV epidemic [39].

HIV and TB fuel each other very effectively: TB accelerates the progression of HIV to full blown AIDS; and HIV increases the risk of progression from latent TB to active TB disease. In fact, in an HIV infected person with a TB infection, the risk of progression to TB disease increases from 10% over a life time to 10% each year [4]. As a result, an HIV positive individual who develops active TB disease can expect to survive an average of just 5 to 6 weeks [194].

2.5 TB Control Strategies

The international standard for TB control is the World Health Organization's Directly Observed Treatment Short-course (DOTS) strategy [131], which aims to reduce the transmission of tuberculosis infection through prompt diagnosis and effective treatment of symptomatic TB patients who present at health care facilities, termed passive case-finding. Considerable progress was made during the last decade using this strategy in countries with small HIV epidemics but the effect of HIV on the African TB epidemic outweighs the gains being made in other regions.

In this study we describe a discrete event simulation model that has been developed to evaluate the effects of more intensive case-finding strategies (so-called active case-finding) for TB control in a high HIV prevalence setting. In essence, active case-finding involves targeted testing of the population for active disease with one commonly used strategy being to target household members of TB patients. Those found to have active disease can be treated promptly, reducing the time spent with infectious TB and so cutting transmission rates. The current policy of active case-finding for adults living in endemic TB settings is ineffective, because transmission events between casual contacts greatly outnumber household transmission events [143]. This policy was developed in an era of low HIV preva-

lence and the impact of the HIV epidemic on the relative importance of household versus community transmission has not been fully assessed.

2.6 Data

This study uses data from a large population-based trial in Harare, Zimbabwe which provides data on the size and location of every household in the study area, as well as the number of inhabitants, their ages and their TB and HIV status (Section 2.6.1). The model is fit to country-wide TB incidence and HIV prevalence statistics for Zimbabwe, available from the 2007 WHO Report [134] and the UNAIDS Epidemiological Factsheets [176] respectively.

2.6.1 The Harare Data and DETECTB

CREATE (The Consortium to Respond Effectively to the AIDS-TB Epidemic) [167] are funded by the Bill and Melinda Gates Foundation and are, along with The Bloomsbury Wellcome Trust Centre [166], currently supporting a group led by Dr Liz Corbett of the London School of Hygiene and Tropical Medicine (LSHTM), called DETECTB. The group, who are based in Harare, Zimbabwe, have been generating data that is informative for making baseline assumptions about the distribution of HIV and TB infection in populations.

Since 2005, DETECTB have been administrating periodic (6-monthly) interventions to 46 neighbourhoods (41,263 households) with a total of 107,430 adults. They have been using either door-to-door enquiry for chronic coughers, or a mobile TB clinic, and diagnosis is based on sputum microscopy. Sputum microscopy is when a sputum specimen is taken from a patient and cultured for Mycobacterium tuberculosis organisms. Communities are randomised to one of these two interventions and the main outcome measures are the cumulative yield over 6 rounds of intervention and a comparison of the point prevalence of TB disease before and after the 6 rounds of intervention.

All the households in the study neighbourhoods are demarcated and identified with GPS and then an interview with the household head is done to collect information on the household structure to a) identify previous TB disease events and b) to allocate a poverty score to the household.

A baseline survey of over 40,000 households has been completed and its results have been used throughout this study to inform the simulation of the appropriate

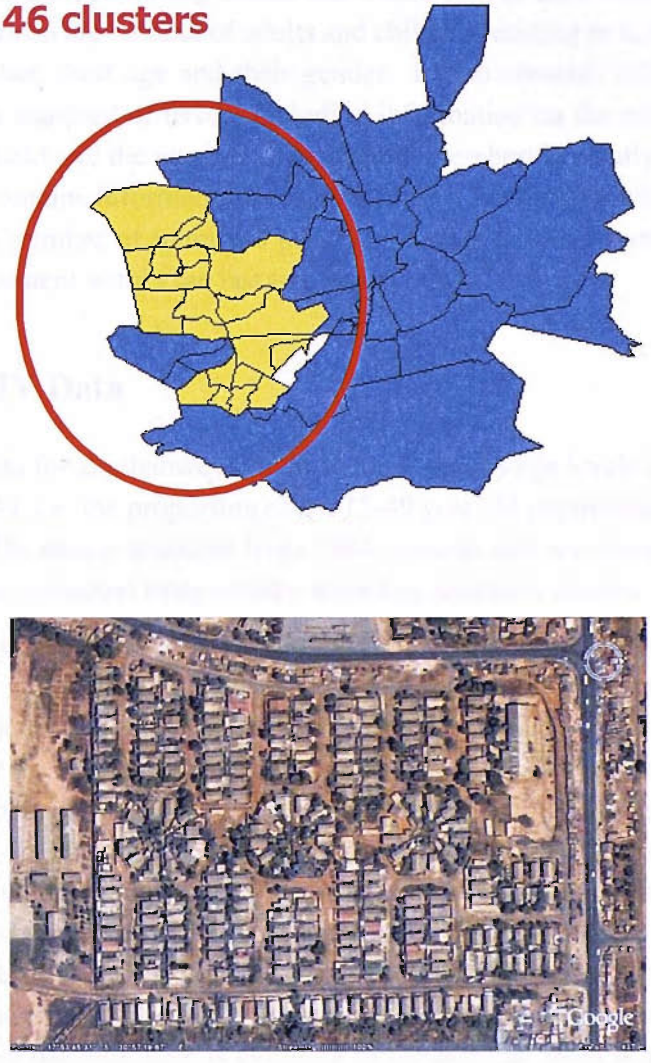


Figure 2.1: A map of Harare, Zimbabwe showing the location of the 46 high density suburbs in which DETECTB are conducting randomised trials of two active case-finding interventions and from which the baseline survey has been carried out. The photograph shows a satellite map of one of the neighbourhoods (Warren Park)

household structure and population dynamics. A copy of the questionnaire used to conduct the surveys can be seen in Appendix A.

For the study area, the survey provides data on the number of dwellings and the number of households living within one dwelling. For each household, the survey provides data on the number of adults and children residing in it, their relationships to one another, their age and their gender. It also contains information on each household's standard of living, including information on the number of rooms in each household and the number of household members typically sleeping in each. Finally, it contains information on a household's TB history, with information regarding the number of members presently on TB treatment and those that have been on treatment within the last two years.

2.6.2 HIV Data

The HIV data for Zimbabwe is in the form of prevalence levels amongst all adults aged 15 to 49, i.e. the proportion of the 15-49 year old population who are infected with HIV. The data is available from 1984 onwards and was provided by UNAIDS [176]. It was collected from women attending antenatal clinics.

The antenatal clinic (ANC) data provides an estimate of the prevalence of HIV in the general adult population, however there are concerns as to whether the ANC data is a reliable representation of the population. The concerns include whether the prevalence rates in ANCs represent the prevalence rates of women, and whether the prevalence rates among women accurately represent the prevalence rates amongst men. The data may over-estimate the prevalence of HIV amongst young women because it is biased toward those that are sexually active, but it may under-estimate the prevalence of HIV amongst older women who may have become infertile due to being sexually active for longer and contracting other sexually transmitted diseases. It may also over-estimate the prevalence of HIV amongst men because there is evidence to suggest that there is higher HIV prevalence amongst women. Despite this, the ANC data is the only data available and the Antenatal Clinic Surveillance Report for Kenya [170] concluds that the over and under estimation made within each gender and age group, actually cancel each other out and as a result, ANC prevalence is a reasonable estimate of total prevalence amongst males and females aged 15-49.

2.6.3 TB Data

The TB data for Zimbabwe is in the form of incidence of active TB disease per 100,000 members of the population. The data is available from 1984 onwards and was provided by the 2007 WHO Report [134].

The incidence of TB is calculated by dividing the notification rates of TB by the rate at which cases are detected. The standard reporting system in place for countries such as Zimbabwe include recording the number of sputum smear positive, sputum smear negative and extrapulmonary patients that are treated for TB in the public sector each year.

Chapter 3

A Review of Previous Infectious Disease and Tuberculosis Modelling Literature

This chapter gives an introduction to infectious disease modelling and a discussion of other models that have looked at the role of household transmission in intervention design. It then goes on to discuss in detail the previous mathematical models that have been built to give a better understanding of tuberculosis epidemiology and effective control measures. The chapter ends with a discussion of the issues raised from previous modelling methodology and concludes that, in line with other infectious disease models, the development of a discrete event simulation model is needed to allow the more intricate details of transmission to be understood and to enable interventions at the household level to be investigated.

3.1 An Introduction to Infectious Disease Modelling

Mathematical models have been used to help understand epidemics since the eighteenth century [26]. The simple mathematical assumptions introduced by Hamer (1906) [86] and developed by Ross (1911) [145] to model malaria provide the basis for modern disease models. These epidemic models have been used to describe for example, the progression of an infection through an individual, to investigate a person's role in transmitting infection, and to describe the spread of the disease through a community.

Kermack and McKendrick (1927) [97] made a significant contribution to the understanding of epidemics and developed some of the earlier modern disease

models. They were the first to put forward the threshold theory, which states that if an infectious case is introduced into a population, the density of susceptibles in that population must be above a threshold value for the epidemic to occur. As the density of the susceptibles increases above the threshold, the size of the epidemic increases rapidly. In later papers, Kermack and McKendrick removed some of their more restrictive early assumptions, first including the effect of a continuous introduction of new susceptible individuals [98] and then adding in a constant death rate, in addition to the disease-induced death rate [99]. Another major contribution to the field came from Soper (1929) [156] who investigated the periodicity of measles outbreaks. These early models were all deterministic, as fitted the scarcity of data at this time. As data became more detailed, the desire to model smaller numbers of people increased, resulting in a move towards stochastic models. Bailey [18] reports that McKendrick was the first to publish a stochastic model of diseases in 1926. This assumes "continuous infection", i.e. the probability of a new case is proportional to the number of susceptibles and the number of infectives in a population, with all infectives being considered to be equally infectious. Work on the continuous infection model was continued in the 1940s and 1950s by authors such as Bartlett [21] [22], who used a stochastic model to attempt to mimic the oscillations in the measles epidemic observed in practice. Bailey [16] [17] and Whittle [193] obtained more general results for stochastic epidemics.

In later decades, ideas from control theory became prevalent in mathematical epidemiology and the focus shifted from stochastic to deterministic models [7]. Deterministic models provide a good description of large-scale epidemics, where the number of people in each of the disease states is large enough to make random effects unimportant. Much progress was made in describing the dynamics of a wide range of diseases and modelling control measures.

In recent years, there has been a move towards more individual-based models as they allow for additional complexity to be incorporated and therefore a more realistic representation of disease epidemiology. In models of sexually transmitted diseases for instance, the effect of the social or contact structure of infectives is crucial. Therefore stochastic simulation models are a natural methodology to adopt and have been used to investigate the transmission dynamics of sexually transmitted diseases such as HIV [137], Gonorrhea and Chlamydia [101]. The structure of a community can also play a role in determining the transmission of other diseases and therefore microsimulation has been used to further explore vaccination strategies for diseases such as measles [60] and influenza [67] [85] [191].

To increase the efficacy of intervention design, some modellers have explored the structure of a community further by modelling individuals within specific

households. The model developed by Elveback *et al.* [67], and further developed by Halloran *et al.* [85] and Weycker *et al.* [191], is a good example of how microsimulation has been successfully used to model individuals within defined households. Weycker *et al.* [191] divides the population into communities, followed by neighbourhoods and then households, with each individual belonging to one of these households. Children and adults may then be assigned to various social groups (playgroup, school, workplace etc.) depending on their age and the composition of their households, and the mix within these various groups. Using a stochastic individual-based model allowed it to be an accurate reflection of the real world. It meant a detailed and complex depiction of social mixing could be incorporated along with variability in susceptibility, infectiousness and the length of latency and infectivity. This made it ideal for investigating child vaccination strategies against influenza in the USA as approximates of the age distribution, household structure and population size could be included and the transmission dynamics correctly captured. Other studies which realised the importance of modelling both within household and community transmission to accurately understand and identify effective interventions for an infectious disease include those investigating influenza [38] [110], SARS [14] [200]), Hepatitis A [151] and the common cold [110] [14]. As households are generally small in size, these studies have focused on stochastic models of transmission, but there have been a few exceptions, where clever model designs have allowed a deterministic treatment of household models [11] [19].

An appreciation of how infectious diseases have previously been modelled has revealed that using a stochastic individual-based approach has become the preferred method. This is due to its ability to capture the reality of transmission and mixing patterns in populations, something that many studies have highlighted as vital to obtaining the correct infection dynamics [2]. Because of this, as Levin [105] suggests, the trend towards more individual-based models is likely to continue, with more account being taken of genetic variation in infectivity and transmissibility of diseases. Previous models have also highlighted the benefit of differentiating between household and community transmission when estimating the main characteristics of transmission.

The following Section looks at the historical development of TB disease models and provides a concise description of the assumptions and methods used for modelling different aspects of TB disease. Along with its accompanying summary table (Appendix B), it contains details of previous mathematical models that have been built to give a better understanding of tuberculosis epidemiology and effective control measures.

3.2 The Historical Development of TB Disease Models

The majority of the TB models previously developed can be divided into two model types: deterministic, compartmental models or (static) mathematical models.

Deterministic Compartmental Model:

In a deterministic compartmental model, the population is divided into different epidemiological groups according to their TB (and HIV) disease status. Differential equations are then used to move proportions of the population through the various groups at specified time steps.

Mathematical/Epidemiological Model:

A mathematical/epidemiological/statistical model is when the model is static and fitted to particular data to establish various parameter values and epidemiological relationships.

The previous models of TB are summarised and explained in the corresponding table (Appendix B) and discussed in more detail in the following Section. Initially, it will discuss the historical development of TB models and then talk about the more recent models which have incorporated the effect of HIV.

3.2.1 TB Models

Simple mathematical models have been used to understand tuberculosis (TB) epidemiology since the late 1950's when Alling (1958) [6] built a Markov chain model to predict the likely course of TB disease for individual patients in America. The model comprised of six states each representing different clinical conditions of the disease, and had transition probabilities and a small number of other parameters. Alling showed that by finding just six of the parameters using maximum likelihood estimation to fit the model to the empirical data, the disease course that patients would take could be predicted. The model included those with advanced TB by incorporating an additional state and the importance of age on disease progression by dividing the patients into two age-dependent groups.

The success of this study is largely unnoticed by other modellers and is significantly overshadowed by Waaler *et al.* [186] who published a deterministic compartmental model in 1962 and went on to develop sophisticated models of TB transmission dynamics [182] [183] [187] [188] [184].

3.2.1.1 Early Models of TB

Waaler [186] was seen to be the first to bring epidemiologists and statisticians together and explicitly apply mathematical statistical methods in modelling the disease. Despite previous studies prior to this paper implying the use of a modelling approach for TB epidemiology [75] [162] [74], and studies such as Alling [6], there can be no doubt that the ingenuity of the approach used by Waaler was an inspiration to many other models and perhaps why pre-Waaler models are largely ignored.

Waaler divided the population into three epidemiological classes and advanced the model in time steps of one year by the use of difference equations. The model was fitted using data obtained by Frimodt-Moller's survey in South India [74], which collected detailed longitudinal data of a population of 60,000 people and noted the effects of introducing BCG vaccination and treatment for infectious cases. Waaler did not intend the results to be taken literally, but rather wanted to show how the epidemiological trend of tuberculosis in a country could be predicted using epidemiological models such as this, and more significantly, that they could be used to evaluate the effect of specific control programs. Waaler's enthusiasm for using a mathematical approach to epidemiology stimulated other epidemiologists to refine his approach. For example, Brogger [36] was amongst the first to improve Waaler's three category model by using systems analysis to form a relatively complex model for TB control based on data from Thailand. The model introduced heterogeneity by classifying subjects with respect to their age as well as their TB state. Persons were therefore transferred through time from one age category to the next and through six classes of TB.

ReVelle [140] used Waaler and Brogger's models as a template but was the first to introduce nonlinear differential equations to model TB dynamics. His main objective was to improve the economic allocation of TB control measures in developing countries. He ignored heterogeneity and just moved homogeneous individuals amongst nine classes of TB using differential equations, to project the course of TB with and without different control policies. This enabled him to evaluate the efficacy and cost of different forms of control.

Other epidemiologists developed new approaches to modelling, such as Ferebee [71], who built a simplistic epidemiological model to project the course of TB in the United States for fifteen years. Ferebee used exact estimates for the number of people acquiring TB, getting infected by each active case, and progressing to clinical disease. Using approximations of the efficacy of various control measures (although sources for these numbers are not provided) she was able to compare the

current control program's effect on TB cases with two other possible scenarios. The results clearly showed that the control program in place at that time in the United States was not going to decrease the number of TB cases to an acceptable level and was less effective than other possible measures.

3.2.1.2 Development of the Earlier TB Models

After these initial studies, it was Waaler who really expanded and improved the modelling methodology. Waaler embraced some of the characteristics of the models developed by Brogger [36] and ReVelle [140], such as age dependency and the incorporation of a vaccinated class, to formulate an extension to his previous model [186]. Waaler [182] developed a model that could be used to generate projections of epidemiological trends in TB given various anti tuberculosis programs. The deterministic compartmental model used difference equations to move people between eight epidemiological classes, which were also specified by five-year age groups. The model was the first to differentiate (by additional classes) between those who have been infected with TB for less or more than five years, and was also the first to separate active cases into those with infectious and non-infectious TB. Waaler then used this model to look at the effect of different BCG vaccination scenarios mainly in low prevalence countries such as Northern Europe and North America [183].

Waaler and Piot [187] continued the investigation of TB control by using the model previously developed [182] to measure the epidemiological effectiveness of BCG vaccinations in terms of "problem" reduction. The "problem" is defined as being the total sum of the individuals suffering and the related social costs caused by the disease. The simulation model used many parameters to represent, for example, measures of a) demographic, b) epidemiological, c) eligibility, d) economic and e) coverage factors. This enabled them to ascertain how to most efficiently implement say, BCG vaccinations (by finding the optimal level of coverage and age of eligibility). By investigating the sensitivity of control methods to the different parameters, the study also revealed the importance of epidemiological issues such as effective contact rates and spontaneous healing rate. This helped develop a clearer understanding of TB dynamics and the factors most likely to affect the efficacy of control measures. This study was extended by Waaler and Piot [188] to include a utility function, which measures the value of a case occurring now as compared with that for one occurring later. The paper discusses the importance of this social-time-preference parameter to the efficacy of control measures. It concludes that the parameter is strongly influential to a control measure's effectiveness

and therefore needs to be appreciated by policy makers. Another paper by Waaler [184], demonstrated the way a decision maker could use their original model [182] to answer various questions necessary for designing a control policy.

3.2.1.3 A Period of Quiet

After these initial investigations in the 1960s, relatively few theoretical studies on TB dynamics were carried out over the next twenty years. ReVelle and Male [141], Chorba and Sanders [43], Horwitz [91] and Azuma [15], were some of the few that did continue the work into the 1970s.

ReVelle and Male [141] used a decision tree to analyse the most cost effective way of testing a population to minimise the cost per active case treated. Chorba [43] developed a simulation of the TB process and applied it to the data from the United States in order to predict future TB prevalence and to provide a cost-benefit analysis of control programs. Horwitz [91] designed a mathematical model to show the dynamic interplay between the disease parameters using data from Denmark, and Azuma [15] developed a simple simulation model to calculate annual trends in TB prevalence and incidence in Japan.

In the 1980s, Goh and Fam [83] simulated the TB problem in Singapore between 1975 and 2025 to establish which control measure to introduce using the epidemiological model developed by Azuma [15]. Trefny and Hejdova [172] also used Azuma's model to do the same analysis of control measures but for the Czech Republic, and Schulzer *et al.* [152] developed a model using a Markov process which assessed which epidemiological parameters were most important to the ongoing spread of the disease in Taiwan and Korea. This was one of the first studies which included an analysis of the importance of drug resistant TB. Finally, Joesoef *et al.* [94] built a deterministic compartmental model to assess the cost-effectiveness of three control methods in Indonesia.

There is speculation that the lack of activity in this area was due to the contining decline of TB in the developed nations which implied active TB was under control [113]. Interest resumed in the early 1990s, however, when countries such as the USA and the UK started to witness outbreaks of multi drug resistant TB.

3.2.1.4 A Resurgence in TB Modelling

From 1985, the USA observed a progressive rise of TB with a total increase of about 9% per year. Likewise since 1991, an increase of 5% was being reported

in the UK with other Western European countries experiencing similar behaviour [173]. In 1993, the World Health Organisation (WHO) declared TB a global emergency and since this time, a large number of TB studies have been started with many TB models being developed.

The models vary from simple parametric models [147] to simulation [122], cluster [11] [155] and Bayesian models [78], but the majority of them [30] [31] [37] [138] [28] [107] [206] [181] [69] [68] [12] [77] [84] extend Waaler's deterministic compartmental model by incorporating various improvements and intricacies to consider more advanced epidemiological issues such as exogenous reinfection [181] [69] [84] age dependency [181], multi drug resistance [31] [37] [68] and self cure [30].

Amongst those that developed deterministic compartmental models is Blower *et al.* [30]. Their intention was to gain a better understanding of the intrinsic transmission dynamics of untreated TB epidemics and the historical epidemiology of tuberculosis. The model was meant for immunocompetent populations and attempts to reflect more biological complexities than previous models by including a spontaneous cure rate (those cured without treatment), by having only a fraction of TB cases as infectious, and by including a recovered class where individuals are able to either die of other causes or develop TB again. This model successfully gave quantitative answers which identified the mechanisms that drive TB epidemics.

Understanding these dynamics meant that Blower's results could be used to design and understand how to control the disease. In their next paper, Blower *et al.* [31] extended their model to include the population level effect of treatment and thereby "developed a theoretical framework for designing effective tuberculosis control strategies" It was felt that control strategies could not be considered efficiently without also considering treatment failure and the subsequent evolution of drug resistance (a significant challenge to control programs). Blower therefore extended the model to include two strains of TB, and thus developed a transmission model that included the dynamics of both drug sensitive and drug resistant TB. The theoretical framework allowed the counterproductive control programs to be identified and successful control programs to be improved upon.

Most other compartmental models developed after this time were intended to evaluate control strategies. These are discussed in the next Section.

DCMs for Evaluating Control Strategies

Castillo-Chavez and Feng's [37] four stage model was modified to include re-

sistant strains of TB. They incorporated two additional classes specifically to represent the development of resistant strains, with the purpose of determining the role that the lack of treatment compliance plays on the maintenance of resistant TB strains.

Porco and Blower [138] produced a time dependent uncertainty and sensitivity analysis of their previous models [30] [31] to identify which input parameters significantly affected the severity of a TB epidemic, and which contributed to the variability in the epidemiological outcomes. This type of understanding of an epidemic is invaluable, especially when considering the role of control strategies. For example, Blower and Gerberding [28] incorporated this information into the compartmental model they developed, which could be used as a health policy tool, to predict the epidemiological outcome of specified approaches to control.

In 2000, Lietman and Blower [107] developed this tool to predict the epidemiological effect of both preexposure and postexposure vaccines. The model was based on the compartmental models they had already built [30] [31] [138] [28] [148] [108] [27] [29]. The model divides the susceptible and latent classes into vaccinated and unvaccinated subclasses and the authors applied the model to both developing and developed countries. The results showed that using both vaccines in developing countries will help to eliminate TB by preventing reactivation, reinfection and new infection at the same time. In developed countries, the results showed it would only be necessary to prevent new infections from occurring (preexposure vaccine), as a small percentage of the population are latently infected. These conclusions are developed further by Ziv *et al.* [206] who evaluated the effect of targeting therapy specifically to persons with recently acquired latent TB compared to those who have been latently infected over a long period. The model was similar to Blower *et al.* [31] and concluded that fewer early cases need to be treated to eliminate TB than if you were to treat those of a large scale population with evidence of a latent infection. These were amongst the first models to stress the importance of treating individuals with latent TB.

DCMs not for Evaluating Control Strategies

Other models, that were not intended to explore the effects of different interventions, were also developed. Sutherland *et al.* [164] and Vynnycky and Fine [181] developed models solely to help understand the role of exogenous reinfection on TB dynamics. Whilst Sutherland had already used a mathematical model to quantify exogenous reinfection and its relative risk compared to that of primary infection and endogenous reactivation, Vynnycky and Fine were the first to model reinfection, by means of a deterministic compartmental model. The model was de-

veloped because few of the previous models had included age structure and none of them (apart from Sutherland's) until this time had considered the importance of reinfection. For example, Blower *et al.* [30] explicitly ignored exogenous reinfection in their model and acknowledged that the model is therefore unsuitable for immunocompromised populations. Vynnycky and Fine's analysis showed that the epidemiology of TB has changed considerably since the end of the 19th century due to HIV; and that age and reinfection are important factors of TB transmission dynamics.

Recognising that exogenous reinfection cannot be ignored in populations having high HIV prevalence, Feng *et al.* [69] also developed a model to understand the effect of reinfection on TB dynamics in developing countries and the inner cities of developed countries. The basic TB model from a previous paper [37] was extended to incorporate exogenous reinfection by introducing a new term into the differential equation that effects the dynamics of the 'infectious' epidemiological class and includes a new parameter which measures the level of reinfection. The results showed that the dynamics of TB would change with the consideration of exogenous reinfection. Biologically it implied that exogenous reinfection increases the number of individuals at risk of becoming infectious.

Feng *et al.* [68] later did a study to investigate the effects of variable periods of latency on TB disease dynamics. Using previous models [37] [70], they developed another compartmental two-strain model with the intention of determining whether the conclusions change when both multiple strain TB and distributed delays of latency are considered. Unlike the previous study the results showed that the introduction of host heterogeneity in latency did *not* change the basic conclusions and therefore the dynamics of the disease remain unchanged from the earlier model [37].

Aparicio *et al.* [12] developed a model to understand the possible reasons for the reductions in active TB incidence in the United States. The roles of demographic, epidemiological and social components were explored by use of a deterministic compartmental model similar to those being published at the time, but with time dependent parameters. Particular importance was placed on the effect of urbanisation and an increase in the standard of living on the disease evolution, and these were incorporated into the model, which provided evidence to suggest the decrease in incidence of active TB was due to a reduction of progression from latent to active disease.

Other compartmental models were developed simply to estimate the possible evolution and trajectories of the disease. Garcia *et al.* [77] present a model where

individuals are vaccinated at birth and only those not protected by the vaccine can become infected with TB. The natural history of the disease is represented by seven epidemiological classes and the transfer rates between the groups are obtained from the literature of TB situations in developed countries. Gomes *et al.* [84] developed the same model but with adjustments to incorporate the effect of exogenous reinfection.

Other Types of Models

Although the majority of models seem to be deterministic compartment models, other types of models have been developed. Salpeter and Salpeter [147] developed a purely mathematical model and applied epidemiological data to it to ascertain estimates of the parameters of TB epidemiology in the United States. The model's results were validated using other published results, and estimates were established for case rates in different age groups and the time delay between the initial infection and active disease.

Aparicio *et al.* [11] proposed a new dynamic model which incorporated the effects of clusters on TB transmission. The model enabled the authors to focus on the effect of long and systematic exposure of infectious individuals on susceptible individuals, which previous models had failed to address. The cluster model is essentially still a deterministic compartmental model but the population is split into two - those individuals belonging to an epidemiologically active cluster (N1) and those that do not (N2). An epidemiological cluster (of size n) is a generalised household with at least one actively infected individual. Therefore if an individual is newly infected, they activate a new cluster and increase the risk of TB for all those susceptibles in their cluster. The cluster model therefore differs from the usual compartment model because instead of moving individuals through the various stages, it moves clusters of individuals. For example, when an individual becomes infectious, this creates an epidemiologically active cluster which means n individuals are moved from population N2 to N1. This model was extended by Song *et al.* [155] who further explored the role of close and casual infections on TB dynamics. The cluster model allows two levels of mixing to occur, with transmission processes occurring at both the population level and the individual level. This allows individuals to acquire TB through membership of an epidemiologically active cluster (close contacts) as well as from random (casual) contacts in the population.

Murray [122] also looked at the clustering of TB cases and defined a cluster by using molecular techniques to identify identical TB isolates in communities. She used a discrete event simulation model to track the chain of disease trans-

mission through different clusters, enabling her to see the effect of variability of strain behaviour and the transmission dynamics of TB in determining cluster size. The study was not intended to obtain specific cluster distributions and TB incidences but to help to interpret empirical studies. Murray concludes by endorsing the microsimulation approach to be used in the study of epidemiology of infectious diseases.

Another method, used by Getoor *et al.* [78] is to create a Bayesian model. They used data collected in San Francisco to create a Bayesian network which was extended using statistical relational models. This allowed the rich and complex data to be explored and TB disease transmission to be better understood. The study revealed the potential of this type of analysis in answering fundamental questions about tuberculosis biology, addressing issues such as heterogeneity of susceptibility and the virulence of different strains.

The models discussed so far have solely been for modelling TB disease. However, there is a significant relationship between the HIV and TB epidemics, as described in Chapter 2, Section 2.4. The following Section will discuss models which have tried to incorporate the intricate relationship between the two diseases.

3.2.2 TB and HIV Models

TB and HIV fuel each other very effectively. TB accelerates the progression of HIV to full blown AIDS; and HIV increases the risk of progression from latent TB to the active TB disease. In fact, in an HIV infected person with a TB infection, the risk of progression to TB disease increases from 10% over a life time to 10% each year [4]. As a result, with both diseases active the average survival time is just 5 or 6 weeks [194]. Models that try to capture the dynamics between the two diseases, and their impact on each other are therefore important, particularly in regions with a high HIV prevalence such as Sub-Saharan Africa.

Schulzer had previously developed a mathematical model to predict TB infection in Taiwan (Schulzer *et al.* [152]) but recognised the urgent need to develop a model which addressed the interaction between TB and HIV. Schulzer was among the first to incorporate the effects of HIV in a TB model and in 1992, Schulzer *et al.* [153] and Bermejo *et al.* [25] published simple parametric models which looked at the impact that HIV was having on TB incidence in developing countries. Bermejo *et al.* [25] developed a static model which investigated the relationship between the two epidemics to show what percentage of new TB cases will be HIV-positive and what TB incidence levels can be expected under various situations of HIV

prevalence. The study formed general conclusions about the dramatic increase of TB incidence that can be expected in developing countries due to the HIV epidemic. Similarly, Schulzer *et al.* [153] developed a model which predicted the likely extra numbers of TB cases due to HIV infection in Sub-Saharan Africa. Unlike Bermejo, Schulzer looked at four different scenarios so that a range of risks of infections could be investigated. This study similarly highlighted the dramatic increase in the number of TB cases that can be expected due to HIV infection. This model was extended to allow it to be applied to other countries and regions effected by the dual epidemic. Schulzer *et al.* [154] discuss the use of the mathematical model in Sub-Saharan populations and the Canadian needle exchange sub-population and predict the expected progression of TB disease in these populations, given the acceleration from the impact of HIV. All three of these early studies emphasise the potential of their models to be used to design appropriate control programmes.

These studies discuss simple mathematical models that were only able to generate general results and conclusions about the impact of HIV on TB. Other more sophisticated models were also developed. Massad [113] for example, developed a deterministic compartmental model which allowed a more comprehensive analysis to be done and the complexities resulting from the interaction of the two infections to be investigated. The actual structure of the model is based on the TB model previously discussed by ReVelle [140] and an HIV model already developed by the author [64]. The model separated individuals into two susceptible classes - those who are susceptible to both infections, and those who are susceptible to both infections but who have previously suffered from active TB. The population is able to progress through the 16 compartments, which represent various stages of the two diseases, and the parameters control the transitions through the model depending on HIV status. The model was meant as a theoretical study and enabled specific relationships between the diseases to be explored and quantified. Firstly, the simulations showed that the pathogenicity of HIV is greatly enhanced by the presence of TB, secondly that the prevalence of AIDS almost doubles in the presence of TB, and vice versa, and thirdly, there is a stronger influence of AIDS on TB than there is of TB on AIDS. These findings increased the understanding of the epidemiological interaction between the two diseases at that time, and were also suggested to be useful for the designing of control strategies.

Heymann [89] developed a 10-stage Markov model to analyse how the interaction between HIV and TB affects both the HIV-positive and HIV-negative population in Africa. The model was implemented using a computer simulation of one million adults over ten years. As well as the effect of HIV prevalence on

TB infection rates, the impact of expanding chemoprophylaxis programs was also evaluated. Data and parameter values derived from the literature were used to determine transition rates. The complex mechanisms behind the increase in TB deaths due to the HIV epidemic were identified and the importance of chemoprophylaxis in decreasing the prevalence of TB in HIV-infected and non-infected individuals by reducing the spread of TB was clearly ascertained. It concludes that providing chemoprophylaxis to HIV-positive individuals is more cost effective than treatment but whether the same is true for HIV-negative individuals depends on various conditions, some of which are identified in the paper.

Heymann was also involved in a similar model which simulated the USA's general population for ten years using available epidemiological data. In the paper by Brewer *et al.* [33] they discuss the development of a semi-Markov model, where, unlike in a Markov model, the probabilities of moving between states can vary over time. There are 18 states defined in the model, which are dependent on TB status (both drug sensitive and drug resistant) and HIV status, and the population is split into three different age groups. Three different prevention strategies and two treatment strategies are introduced singly and in various combinations to assess their impact as TB interventions. The study concluded that on their own, the treatment and prevention strategies would not be sufficient to eliminate TB in the USA and that a combination of control strategies is needed.

Another type of model that was used to investigate and quantify the potential impact of the HIV virus on TB was discrete event simulation. Porco *et al.* [139] extended their previous deterministic models [30] [31] [138] [28] [107] [206] of tuberculosis by including stochastic effects and the effect of HIV on both the pathogenesis and transmission of TB. The model had six states to represent the disease progression of TB in HIV negative individuals and a further six states for each of the four WHO defined stages of HIV. The stochastic model simulated, under various HIV and treatment rate scenarios, the average outbreak size from introducing one infectious case of tuberculosis. The results showed that in areas with very high treatment rates for TB, HIV epidemics are unlikely to substantially increase the number of TB cases but that in areas where treatment rates are moderate or less, "HIV is likely to significantly amplify the TB epidemic". The study highlighted that WHO target levels for tuberculosis treatment are well below what they need to be and that in developing countries, decreasing the prevalence of HIV will decrease the incidence of TB.

After 1998, a series of deterministic compartmental models were developed specifically to investigate the efficacy of directly observed treatment, short course (DOTS), the World Health Organisation's recommended strategy for TB control.

Murray and Salomon [120] elaborated on a previously discussed model which was developed by Blower *et al.* [30] [31]. The finite difference model describes the progression through a possible 19 states of TB and was applied to populations from five regions of the world [1]. This was one of the most complicated compartment models developed so far and incorporated features such as superinfection, fast or slow breakdown to TB disease, fast or slow diagnosis rates and three clinical categories of TB. The profound effect of HIV on the development of TB is also captured by creating two sub models to represent the HIV-negative and HIV-positive populations. The paper concludes that using DOTS alone will be inadequate and that extensions need to be implemented. In a subsequent paper [121], they go on to advocate the use of active case finding in high HIV prevalent populations as a cost-effective extension to DOTS, by incorporating costs into the original analysis.

A very similar model was developed by Dye *et al.* [61] to quantify the worldwide effect of the DOTS strategy and hopefully justify it. The model is a deterministic compartmental model with two sub models to incorporate the effect of HIV; however it also includes age structure. The model was applied to the six WHO regions of the world and showed that improvements in case finding and cure rates are vital, as even if WHO targets were met by 2010, "three-quarters of the worldwide TB burden would still not have been averted in the next 23 years."

This model became a popular foundation for other models and studies. An adaptation of it was used by Currie *et al.* [55] to compare and investigate preventative methods of TB control with case detection and cure. The study was motivated by the observation that DOTS was failing to prevent increases in TB cases in high HIV prevalent populations. The Dye [61] model was adjusted so that age structure was removed but that options for TB control were extended to include three preventative methods (one of which was to reduce HIV transmission) as well as case detection and cure. The model was applied to data from South Africa, Kenya and Uganda; all countries where the TB epidemics are driven by HIV. The results showed that reducing the burden of HIV is an effective way of reducing TB, but that the effect is delayed and less dramatic than finding and curing active TB. The authors concluded that both HIV and TB control would be needed to give a long-term decline in TB incidence.

Currie *et al.* [56] then used the DCM approach to develop a different model to look at the effect that the duration of TB infectiousness amongst late-stage HIV-positives has on the dual epidemics of TB and HIV. A dynamic TB transmission model is split into two sub models to represent HIV-negative and late-stage HIV

[1] The world was divided into five regions based on patterns of TB epidemiology

individuals, as in the earlier model, but the TB states are more complex with the latent TB infection state split into a state to represent latently infected individuals who progress quickly to active disease, and a state for individuals with a long-term TB infection. The model was applied to data from Kenya and showed, using various scenarios for duration of infectiousness, death and diagnosis rates, that when the duration of infectiousness is short, the HIV epidemic has relatively little effect on TB prevalence. This is due to the fact that the majority of transmission events are attributable to HIV-negative individuals, because the HIV-positive individuals are infectious for such a small amount of time. This could have important implications for TB intervention design, especially when studies in South Africa showed that the duration of infectiousness could be as little as 2 months in late-stage HIV individuals, rather than the 6 months assumed by the WHO.

Dye and Williams [63] also adjusted their previous model [61], to investigate multi drug resistant (MDR) TB and how best to eliminate it. The model allows both drug resistant and drug sensitive TB to be modelled and was one of the first, along with Vynnycky *et al.* [181], to incorporate a disease state for those with fast progression to active TB, which, as already discussed was later used by Currie [56]. The results were obtained by applying the model to data from 6 countries to represent the current variations in cure rates. The study allowed general conclusions about the control of MDR TB to be made, mainly that current treatment and cure rates are inadequate if an MDR-TB epidemic is to be prevented.

In another study, Dye and Williams adjust their previous models [61] [63] and discuss a new model for HIV and TB. Williams *et al.* [197] discuss a model to investigate the capability of India's Revised National TB Control Program (RNTCP) DOTS program, to reach the United Nations Millennium Development Goals of halving TB prevalence and death rates by 2015. This time the model was replicated 5 times to represent each possible HIV stage (negative and stages 1 through to 4), as the impact of HIV on TB and the RNTCP's ability to control the TB epidemic was the primary goal. The model's other modifications, were the removal of age dependency [120] and the lack of a 'fast latent' stage [63]. The model results showed that HIV should have little impact on India's RNTCP and its ability to reverse the increase in TB incidence but that more treatment is needed to halve TB mortality by 2015.

As the review shows, the majority of previous studies have used deterministic compartmental models which have focused on modelling TB at the population level. These studies concentrated on the effect of interventions at a large scale and although many of them addressed the implications of reducing transmission, none of them were able to look at the actual mechanics behind it. These studies

have been vital in understanding and quantifying TB disease progression in populations. However it is felt that a discrete event simulation is more appropriate for investigating interventions at the household level and enables the more intricate details of transmission to be understood. There is also a need to further investigate TB control strategies in areas of high HIV prevalence; in particular with relation to active case-finding.

3.3 Issues from the Literature

When designing the discrete event simulation model it is useful to appreciate the issues uncovered by the literature regarding TB modelling, some of which remain unresolved or continue to be inadequately addressed. The rest of this Section discusses some of these issues and considers which of them will be important to the study.

3.3.1 HIV

The majority of previous models have not incorporated the effect of HIV on TB disease evolution and the effectiveness of control strategies, however it is now widely accepted, as reflected in the most recent literature, that it is essential to simultaneously control both epidemics, especially in developing countries [139]. This means that TB models must include HIV and that better data is needed to resolve the uncertainties surrounding the parameters connecting the two epidemics [197].

3.3.2 Homogeneity

The majority of the studies divided the population into homogeneous groups based on the natural history of TB disease and presumed homogeneous mixing of the population.

Age Dependency:

Age structure is particularly important in TB epidemiology as a lot of the disease parameters are dependent on it. Age affects the mortality rate and life expectancy and also determines both the risk of developing disease after infection [11] [155] and the type of disease developed (pulmonary /extrapulmonary). Also, the behaviour of children versus adults in disease evolution can be very different

as children are less likely to infect others or progress to active disease. Some studies explicitly ignored children because they are not a priority group for TB control programs in developing countries [25]; however it is unclear how important they are when considering the disease dynamics within households.

A few studies recognised the importance of age on the epidemiology of TB and tried to incorporate age dependency [6] [36] [182] [181] [61]; but all these studies failed to include the effects of either gender or non homogeneous mixing.

Gender:

Gender differences are important and have been observed in TB epidemiology; for example Sutherland *et al.* [164] found that the differences between the sexes were statistically significant and that there are genuine differences between male and female rates of mortality, incidence and disease progression. However, none of the models discussed in this review incorporated this phenomenon into their models.

Homogeneous Mixing:

The only models that address non homogeneous mixing are those by Aparicio *et al.* [11] and Song *et al.* [155]. These models are still deterministic compartmental models with no age or gender structure; however they do try to model two different levels of mixing with transmission processes occurring at both the population level and the individual level. Although this provided some insight on TB dynamics between close and casual contacts, the groups were still homogeneous as the population structure was ignored and therefore the modelling was unrealistic. Previous models have therefore only been able to look at the effect of large scale interventions and have been unable to look at the actual mechanics behind transmission. It is felt that discrete event simulation is more appropriate for investigating the intricate details of transmission and the effect of household level interventions, because different mixing patterns and heterogeneous contact rates can be incorporated into the model.

3.3.3 Endogenous Reinfection

Although reinfection has been incorporated by many studies, its importance in immunocompetent populations is still debated within the literature [68]. However, most studies agree that exogenous reinfection plays an important role in disease progression in developing countries where high incidence rates are observed.

3.3.4 Lengths of Latency and Infectiousness

One of the criticisms of all the previous models is that they do not take into account the long and variable periods of latency which is an important feature of TB. Only Feng *et al.* [70] recognised this and tried to incorporate it into one of their basic TB models however it was unable to capture the effect sufficiently. They found that the disease either died out or remained endemic regardless of the shape of the length of latency distribution.

The length of infectiousness is also variable and influenced by an individual's age, sex and disease characteristics. The varying lengths of infectiousness are particularly important when considering the effect of HIV on the TB epidemic. HIV-positive individuals are infectious for a relatively short amount of time - recent studies in South Africa suggest it could be as little as two months [49] [48]. When the duration of infectiousness is short the HIV epidemic has relatively little effect on TB prevalence because the majority of transmission events are not coming from HIV-positive individuals. This effect has only been specifically modelled by Currie *et al.* [56] and they concluded that it has important implications for the design of control strategies and is therefore an important feature to incorporate into a model of TB and HIV.

3.3.5 Multi Drug Resistance

Multi drug resistant (MDR) TB is generated by inadequate treatment and is a growing impediment to the effective design and success of control strategies. A handful of studies have incorporated multi drug resistance into their models ([152] [31] [37] [68] [33] [63]), as it was felt that control strategies could not be considered efficiently without also considering treatment failure and the subsequent evolution of drug resistance. Many of the studies were successful in identifying how current control strategies could be improved in light of MDR TB and also in determining the role that the lack of treatment compliance plays on the maintenance of resistant TB strains. The models conclude that the prevalence of MDR TB is an essential consideration when designing TB control strategies and therefore its incorporation into the design of a TB model needs to be evaluated.

3.3.6 Immigration

Immigration is an important element to understand when modelling tuberculosis in a population because of the flux of infectious individuals that it can introduce

into the population. Studies in developed countries such as America have shown that foreign-born persons composed 60% of the increase of TB cases between 1986 and 1992 and in Auckland, 63% of new cases were found among immigrants from Asia and the Pacific Islands during 1992 [69]. This implies that foreign-born individuals with TB may be responsible for much of the transmission of TB in Auckland and that TB incidence rates in developed countries are affected by immigration. Some models, such as Feng *et al.* [69], appreciate the importance of immigration and its impact on TB control, but all fail to actually model it. The importance of immigration on TB disease dynamics in developing countries is still uncertain.

3.3.7 Poverty

Tuberculosis has long been a disease of poverty for several reasons. The risk of being infected with TB is higher among poor people because there are higher contact rates in overcrowded homes and areas, the risk of developing active disease is higher amongst people with poor immune systems due to below average nutrition and working conditions, and the chance of being successfully diagnosed and treated depends on good infrastructure and the availability of health services [185].

Previous models have been unable to capture the effect of poverty because of the complications and ambiguity surrounding modelling it. Some have acknowledged the strong evidence which suggests that "a correlation exists between TB progression rates and the corresponding average standard of living" [12]. However, it has only led to hypothesising over the effects on TB control interventions and not the modelling of it. It is currently thought that the clustering of TB infection and HIV infection in the poorest of households may be exacerbating the biological interaction between these two diseases. How this may reduce the success of TB control interventions (since the poorest households may also tend to have the least access to health care) still needs to be determined.

3.4 Summary and Conclusions from the Review

The use of mathematical modelling in the study of TB has helped to illuminate the general epidemiology of the disease as well as to optimise the effectiveness of control measures. As such, it has evolved over time to answer questions such as those about exogeneous reinfection in the 1990s moving onto questions about the interaction with HIV in the late 1990s and early 2000s. Following shortly

behind advances in vaccinations, TB treatment and prophylaxis, and occasionally anticipating what these advances might be, it has been used to compare different interventions and to find the optimal strategies for their delivery in terms of effectiveness and cost.

Moving to the future, there are still a number of unanswered questions about the effects of HIV on TB and the impact this has on control. Answering some of these questions, in particular those concerning HIV, where the effect of the contact network is needed to understand transmission, may require different modelling methods from the deterministic compartmental models that have traditionally been used to model TB. For example, there is undoubtedly a place for more stochastic, individual-based microsimulation models.

The purpose of this study is to evaluate the effects of more intensive case-finding strategies for TB control in a high HIV prevalent setting. As active case-finding can involve targeting household members of TB patients, a clearer understanding of the role of household versus community transmission of TB is needed. A review of both TB and other infectious disease modelling literature points to discrete event simulation as the most appropriate method for investigating interventions at the household level, as it enables the population dynamics to be more accurately represented and therefore the more intricate details of transmission to be understood.

The importance of incorporating various epidemiological issues into a TB model have been identified. Which issues are addressed by a model are dependent on the questions needing to be answered. For our discrete event simulation model, which is looking to determine the relative importance of household interventions in controlling TB in HIV settings, all of the issues could be considered as important. In our model we only aim to address HIV, age dependence, non-homogeneous mixing, reinfection, and varying lengths of latency and infectiousness. It would be ideal to also incorporate MDR TB, gender and immigration, however more comprehensive data on their impact is required. It would also be valuable to explore the impact of poverty on the likely success of interventions, and cross-sectional data on this is being collected by DETECTB in Harare. The work involved in using this data to quantify poverty is time-consuming and has not been done for this model. It is considered feasible however, and would be a very interesting and useful extension to this research (see Chapter 10, Section 10.3).

Chapter 4

Examining the Dynamic Relationship between HIV and TB

4.1 Introduction

In an initial study of the interactions between the HIV and TB epidemics a simple parametric model was developed at the population level. This was done in collaboration with Dr Brian Williams of the World Health Organisation and much of the following discussion is taken from the currently unpublished paper entitled "The impact of HIV on TB at a district level in Kenya" [116]. Using district level data from Kenya on TB notification rates and HIV prevalence among women attending ante-natal clinics, the dynamic relationship between TB and HIV was explored. Kenya was an appropriate study area because of the availability of comprehensive TB data and because TB notification rates have risen by up to ten times in some of its districts over the last ten years, almost certainly as a result of the HIV epidemic. The presence of an HIV led TB epidemic allowed the relationship between the epidemics to be investigated at the population level. An understanding of how HIV impacts on TB epidemiology is important, not only for Kenya, but for TB modelling in all other countries with substantial HIV epidemics.

The model developed is a simple parametric model which was fitted to, where available, district level data of both infections. The model predicts the likely course of the HIV and TB epidemics. It is particularly useful for those districts where TB data, but not HIV data, are available as it is capable of estimating the HIV prevalence from just the TB notification rates. This is useful because in some countries such as Kenya, TB notification data is considerably more reliable than the HIV prevalence data both in quantity and quality. The results from the initial

modelling work described in this Chapter are not used as a direct input into the DES model.

4.2 Background

The risk of developing TB disease increases as people progress from HIV infection to AIDS and studies have been carried out to determine the odds ratio for TB in HIV-positive and HIV-negative people [132], changes in the risk of TB as people progress toward AIDS [196], the relative infectiousness for TB of HIV-positive and HIV-negative people [58] [79], and the extent to which HIV-infection affects transmission of TB [46]. Attempts to model the population level impact of the TB epidemic have relied heavily on these and other studies to estimate the necessary parameters, assuming that both the parameter estimates and the model structure adequately capture the dynamics of the two epidemics.

As already discussed in Chapter 3, a number of attempts have been made to model the impact of HIV on TB using dynamical compartmental models [89] [33] [120] [121] [61] [63] [139] [53] [123] [55] [47] [199]. However, without good population level data for both diseases it is difficult to be confident about the validity of these results and the confidence limits in population level estimates are wide [196]. The problem is that where TB notification rates are high and HIV is prevalent, some countries, such as South Africa, have good data on the prevalence of HIV but poor data on the incidence of TB, while other countries, such as Kenya, have good data on TB but less certain data on the prevalence of HIV.

TB notification rates are available from all of the 41 districts in Kenya since 1985. The number of districts changes over time as district boundaries are redrawn and districts merged or split. The district level data for HIV is less complete and only 10 districts have reliable and consistent data for more than 10 years. However, this is the most comprehensive set of sub-national data on both diseases for any developing country in the world and allows us to develop preliminary estimates of the relationship between TB and HIV in Kenya.

Because of the limited extent and coverage of the HIV data, even in Kenya, we use a parametric model to describe the interaction between the two diseases. The data cannot support detailed parameter estimation for a full compartmental model and we seek the simplest possible model that is biologically plausible and is sufficiently flexible to fit the data. This simpler model enables us to identify important relationships between the two epidemics.

The results of the model show that using the TB incidence data to predict the HIV prevalence curve in Kenya's districts is a realistically precise method. This work is our own contribution toward the evidence of a strong link between the HIV and TB epidemics, a key fact underpinning the need to develop a discrete event simulation (DES) model of TB transmission (Chapter 1).

4.3 Data

In addition to information on the changes in district boundaries over the last 18 years, we use three sets of data: TB data from the routine surveillance system of the Kenyan National Leprosy and TB Control Programme (NLTCP), HIV data from the sentinel surveillance system of the Kenyan National AIDS and STD Control Programme (NASCOP) [10], and data from the Kenyan National Census for the years 1979, 1989 and 1999 [8].

4.3.1 District Boundaries

Over the 18 years for which TB data are available for this study there have been a number of changes to the district boundaries in Kenya (Appendix C). In 1990, for example, Kericho was divided into Bomet and Kericho. These two districts were again divided in 2000 with part of Bomet and part of Kericho combining to form a new district, Buret. In this study we use the district names as given for the period 1985 to 1991, which gives a total of 41 districts, and combine data from the present districts where necessary. Figure 4.1 is a map of the Kenya districts as referred to in this study.

4.3.2 TB Data

The TB data are part of the standard reporting system in which the number of sputum smear positive (SS+), sputum smear negative (SS-), and extrapulmonary (EP), patients that are treated for TB in the public sector each year is recorded. Ideally one would use the incidence of TB rather than notification rates. Multiplying the notification rate by the case detection rate (CDR) gives the incidence but the CDR is difficult to estimate precisely. However, the World Health Organization estimates that the case detection rate has remained fairly constant, between 47% and 49%, for the last ten years [133]. Provided the case detection rate has not varied over time this will only introduce a scaling factor into the model fitting.

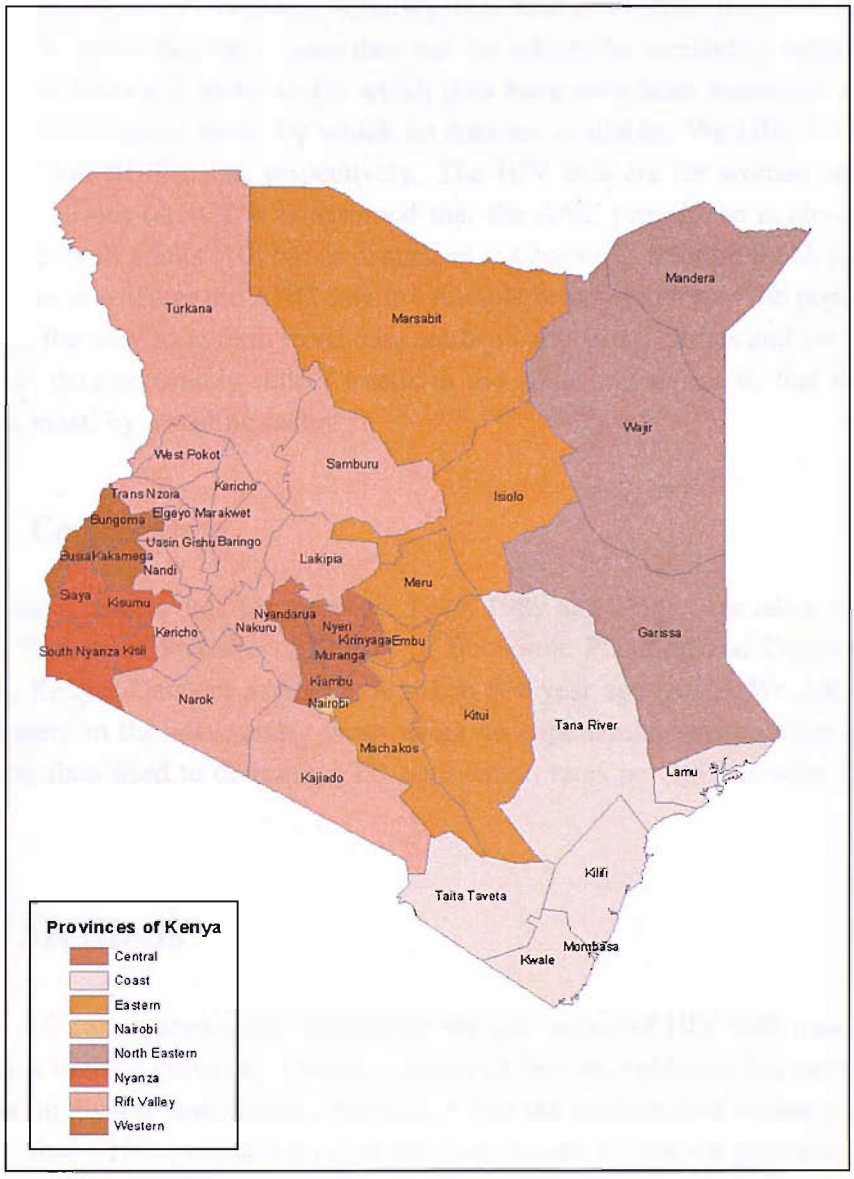


Figure 4.1: The provinces and districts of Kenya

4.3.3 HIV Data

HIV data are only available for some districts for some years. Visual inspection allows us to classify the districts broadly into three groups: a) those that have consistently smooth trends in time (within the binomial confidence limits) for at least 11 years; b) those that have some data but for which the variability substantially exceeds the binomial errors or for which data have only been measured at a few points in time; and c) those for which no data are available. We refer to these as Type I, II and III districts, respectively. The HIV data are for women attending ante-natal clinics (ANC). It is assumed that the ANC prevalence is close to the prevalence in all adults [10] but as discussed in Chapter 2, Section 2.6.2, there are concerns over whether the ANC data is a reliable representation of the population. However, the only long term trend data are from ante-natal clinics and we assume that these data accurately reflect trends in the adult prevalence so that they too differ, at most, by a scaling factor.

4.3.4 Census Data

The national census data for the years 1979, 1989 and 1999 were taken from the Central Bureau of Statistics, Ministry of Economic Planning and Development, Nairobi, Kenya. Data are available by sex in five year age bands. We determined the numbers in the intervening years using an exponential interpolation. These data were then used to determine TB notification rates per 100k people for each district.

4.4 Methods

In most African countries the variation of the prevalence of HIV with time can be fitted by a logistic curve or, if there is evidence that the epidemic has peaked and is declining, by a double logistic curve [55]. For the Kenyan data we use a logistic curve to fit the HIV-prevalence curve for each district so that the prevalence $P(t)$ is

$$P(t) = \frac{ae^{\alpha(t-\tilde{t})}}{1 + e^{\alpha(t-\tilde{t})}}, \tag{4.1}$$

where a is the asymptotic prevalence, α is the rate at which the HIV prevalence increases at the start of the epidemic, and $\tilde{t}$ determines the timing of the epidemic.

In particular we assume that the TB notification rate does not affect the prevalence of HIV.

TB incidence can also be fitted using a logistic curve but the initial value of the notification rate is not zero and so we include a non-zero asymptote as an additional parameter. Furthermore, we are interested in exploring the impact of the HIV epidemic on the rates of TB in both HIV-positive and HIV-negative people and we include these separately in the model.

Several studies have shown that even if TB notification rates increase dramatically as a result of the HIV epidemic, the annual risk of infection and the prevalence of TB disease in HIV-negative people may increase only slightly, if at all, because the increase in the individual risk of developing TB is balanced by the lower infectivity and the higher mortality and rate of disease progression of HIV-positive TB patients [46] [81] [45] [142] [117]. We therefore assume that $I^-(t)$, the TB notification rate in HIV-negative people, is constant and equal to $I^-(0)$, the TB incidence observed before the HIV epidemic began to have an impact on TB, and therefore

$$I^-(t) = I^-(0) \tag{4.2}$$

The incidence of TB in HIV-positive people depends on both the rate at which HIV-positive people acquire new infections and the rate at which latent TB infections break down. The latter depends on the historical rates of TB infection which will determine the proportion of people that have a latent TB infection, but is otherwise independent of current rates of infection; the former depends only on the current risk of TB infection. The median life expectancy after infection with HIV is about 10 years [196] and HIV-positive people present on average about 8 years after infection with HIV, and so we anticipate a delay between the rise of the HIV epidemic and the resulting rise in the TB epidemic. We therefore model the incidence of TB in HIV-positive people as

$$I^+(t) = I^-(t)\left\{1 + \beta P(t - \hat{t})\right\} + \gamma P(t - \hat{t}) \tag{4.3}$$

The first term on the right hand side, which represents reinfection, is proportional to the incidence of TB in HIV-negative people increased by a factor that is proportional to the prevalence of HIV infection $\hat{t}$ years earlier; the second term, which represents the activation of latent infection, is independent of the current incidence of TB but is proportional to the prevalence of HIV infection $\hat{t}$ years earlier. The overall incidence of TB is given by

$$I(t) = I^-(t)\left\{1 - P(t)\right\} + I^+(t)P(t) \tag{4.4}$$

The model therefore involves three parameters that depend on the course of the HIV-epidemic (the timing, $\tilde{t}$, the rate of increase, α, and the asymptote, a), one parameter that determines the TB incidence prior to the advent of HIV, $I^{-}(0)$, and two parameters that determine the contribution of reinfection and the breakdown of latent infection to the incidence in HIV-positive people, and $\hat{t}$ that determines the delay between the onset of the TB and HIV epidemics.

We first fit the model simultaneously to the districts with the best HIV data. In doing this we use the same four 'global' parameter values for all the districts. The parameters are chosen to be 'global' as it is considered that α the initial rate of increase of the prevalence of HIV; β and γ, that link the TB epidemic to the HIV epidemic; and $\hat{t}$, the parameter that determines the relative timing of the two epidemics should be the same regardless of geographical location. The remaining three 'local' parameters are allowed to vary among the districts. These parameters are considered 'local' as the timing of the HIV epidemic, $\tilde{t}$; the asymptotic HIV prevalence, a; and the initial value of the TB notification rate, $I^{-}(0)$ will vary depending on the region.

The Nelder-Mead optimization routine [55] is used to find the maximum likelihood [195] estimates for the model parameters (see Appendix D for details of the Nelder-Mead and maximum likelihood estimation methods). We estimate the covariance matrix by finding the negative inverse of the Hessian (matrix of double derivatives) at the maximum likelihood parameters values. We use bootstrap resampling methods to obtain confidence intervals for the fitted curves. This is a well established and convenient (although computer intensive) way of calculating the distributional properties of the statistics of interest [66] [42] [57] [177]. Appendix E gives a summary of the bootstrap methodology that was employed. We use the districts with less good data as a check on the fitted parameter values and then apply the model to those districts for which no HIV data are available in order to estimate the trends in HIV in those places.

4.5 Results

We first fitted the full model to the data from the Type I districts (listed in Table 4.2 and shown in Figure 4.2) including both of the parameters that determine the relationship between the two epidemics (β and γ) and then examining the effect of dropping each in turn. Including γ did not significantly improve the fit over that obtained with β only. We therefore kept the simpler optimisation by including only the parameter β in the following analysis. Statistically acceptable fits were

			Covariance matrix		
Parameter	Estimate	95% C.I.	α (yr^{-1})	$\hat{t}$ (yr)	β
α (yr^{-1})	0.458	0.034	0.00031	-0.00138	-0.09848
$\hat{t}$ (yr)	7.65	0.23	-0.00138	0.0132	0.71702
β	325	13	-0.09848	0.71702	47.22267

Table 4.1: Estimates of the global parameters obtained by fitting the model to Type I districts

District	a (%)	95% CL	$\bar{t}$ (yr)	95% CL	$\bar{I_0}$ (/100k)	95% CL
Busia	0.207	0.171- 0.244	87.56	86.73- 88.33	14.82	10.66- 21.09
Kakamega	0.141	0.129- 0.155	91.22	90.48- 91.94	20.24	17.09- 24.02
Kisumu	0.276	0.258- 0.294	89.38	88.86- 89.88	24.28	21.07- 27.58
Kitui	0.094	0.087- 0.103	88.71	87.95- 89.50	67.16	58.66- 74.98
Meru	0.147	0.138- 0.157	92.01	91.48- 92.52	37.99	33.63- 42.18
Mombassa	0.127	0.114- 0.141	88.26	87.43- 89.07	141.87	117.04- 169.15
Nairobi	0.167	0.156- 0.179	89.21	88.94- 89.49	69.86	61.71- 78.67
Nakuru	0.180	0.155- 0.208	89.47	88.82- 90.16	24.21	18.28- 31.88
Nyeri	0.168	0.148- 0.193	89.47	88.85- 90.03	24.42	18.44- 30.94
Trans-Nzoia	0.104	0.096- 0.113	89.48	88.84- 90.17	33.27	29.15- 37.51

Table 4.2: Estimates of the local parameters for the Type I districts with their confidence intervals. Years are relative to 1900

obtained for all the Type I districts, giving the global parameters (with confidence intervals and covariance) in Table 4.1, and the local parameters listed in Table 4.2.

The three parameters retained in the model are all significant ($p < 0.001$) and results show that the confidence intervals on the global parameters are small and we are therefore reasonably certain about the values obtained. The fitted curves are shown in Figure 4.2. The model fits for the TB notification rates are generally good. The data for HIV are more variable and in several districts the model estimates are outside the binomial confidence limits on some of the data points. We suggest that this maybe the result of errors in data reporting rather than the inadequacy of the model as the HIV data can be unreliable. For example it would seem infeasible that in Nakuru the number of HIV-positive individuals reduced by 60% in 1996 compared to the previous year with HIV prevalence reducing from 26.2% to 10.0% in just one year.

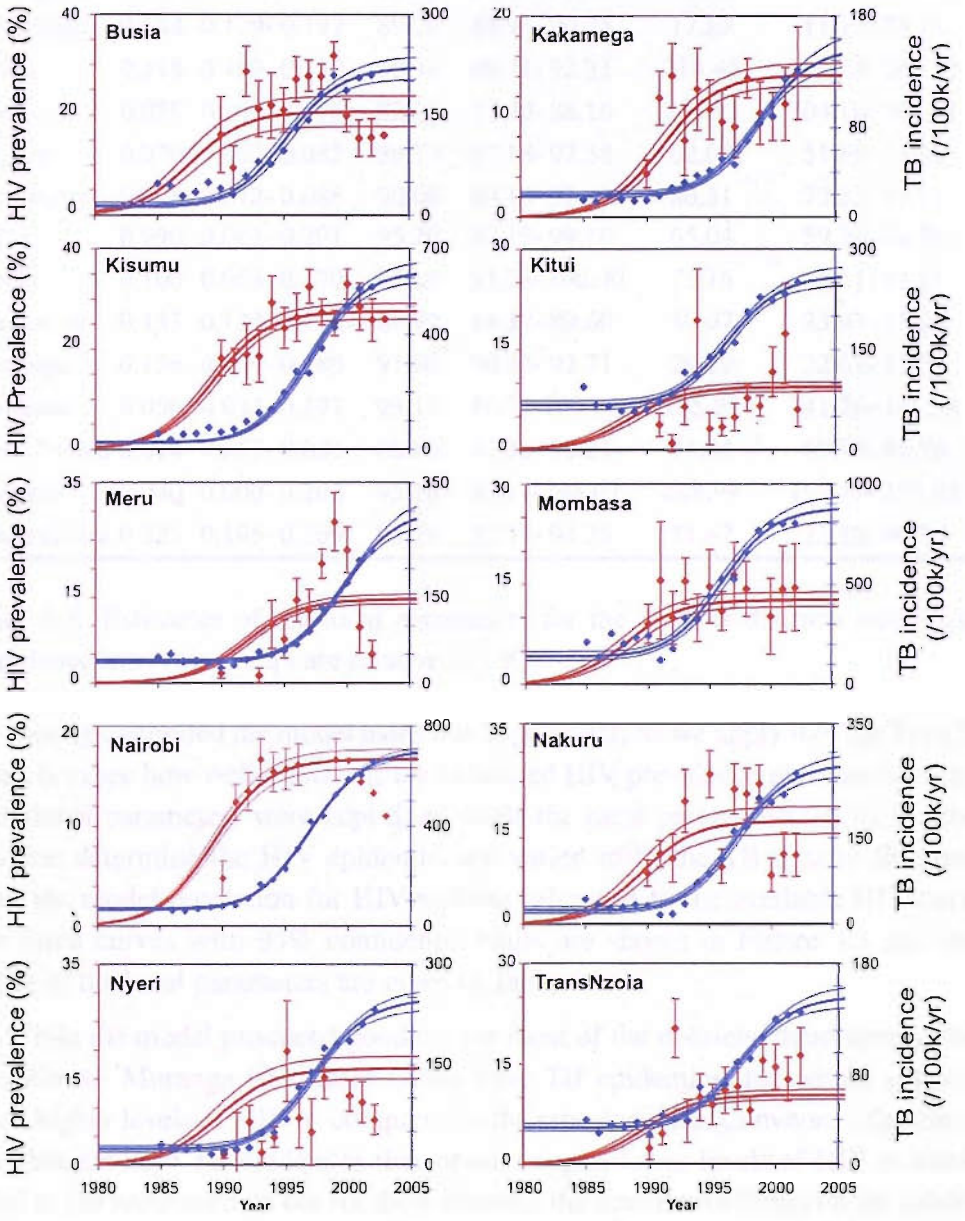


Figure 4.2: TB incidence (blue) and HIV prevalence (red) for the Type I districts. The graphs show, for each district, the TB incidence data and HIV prevalence data (with binomial confidence limits), and the TB incidence and HIV prevalence curves produced by the model with their 95% confidence intervals

District	a (%)	95% CL	$\bar{t}$ (yr)	95% CL	$\bar{I_0}$ (/100k)	95% CL
Baringo	0.106	0.089- 0.130	91.58	90.16- 93.38	38.18	29.81- 45.66
Bungoma	0.152	0.129- 0.197	89.79	88.96- 90.75	17.29	11.13- 23.21
Embu	0.118	0.103- 0.140	90.92	89.71- 92.31	214.45	161.29- 260.62
Garissa	0.058	0.045- 0.076	82.27	77.90- 88.16	146.93	104.10- 202.01
Kajiado	0.070	0.062- 0.082	89.77	87.18- 92.58	62.09	51.86- 71.62
Kericho	0.078	0.072- 0.086	90.08	89.10- 91.11	80.31	72.53- 87.61
Kilifi	0.090	0.061- 0.201	95.20	92.13- 99.10	65.04	59.70- 69.76
Kwale	0.100	0.063- 0.270	95.20	91.38- 100.00	75.16	65.11- 83.25
Machakos	0.135	0.123- 0.154	88.99	88.37- 89.60	30.07	23.97- 35.16
Muranga	0.156	0.139- 0.180	91.80	90.93- 92 71	29.19	22.61- 35.05
Samburu	0.050	0.032- 0.197	93.11	86.72-100.00	165.95	141.76- 181.38
Taita Taveta	0.074	0.065- 0.083	88.85	87.36- 90.21	81.04	69.99- 91.76
Turkana	0.040	0.000- 0.202	95.16	83.14-148.67	248.99	193.42- 279.92
Uasin Gishu	0.225	0.195- 0.269	93.29	92.39- 94.28	31.67	22.88- 40.26

Table 4.3: Estimates of the local parameters for the Type II districts with their confidence intervals. Years are relative to 1900

Having calibrated the model using the Type I districts we apply it to the Type II districts to see how well it predicts the measured HIV prevalence in those districts. The global parameters were kept fixed while the local parameters, including the two that determine the HIV epidemic, are varied to fit the TB data so that this gives the model prediction for HIV without reference to the available HIV data. The fitted curves with 95% confidence limits are shown in Figure 4.3 and the values of the local parameters are given in Table 4.3.

While the model produced good fits for most of the districts, there were some exceptions. Muranga and Uasin Gishu have TB epidemics that would suggest much higher levels of HIV as compared to the reported data. Conversely Samburu and Turkana have TB epidemics that would suggest lower levels of HIV as compared to the reported data but for these districts the confidence limits on the model fits for HIV are very wide and the HIV data are very limited.

The confidence intervals on the estimated local parameters (Table 4.3) are, on average, more than two times wider than those obtained for the Type I districts (Table 4.2). This is to be expected as the parameters were obtained by fitting the model using only the TB data and ignoring the HIV data available; one would therefore expect the estimates to be less precise. Similarly, the confidence limits on the HIV

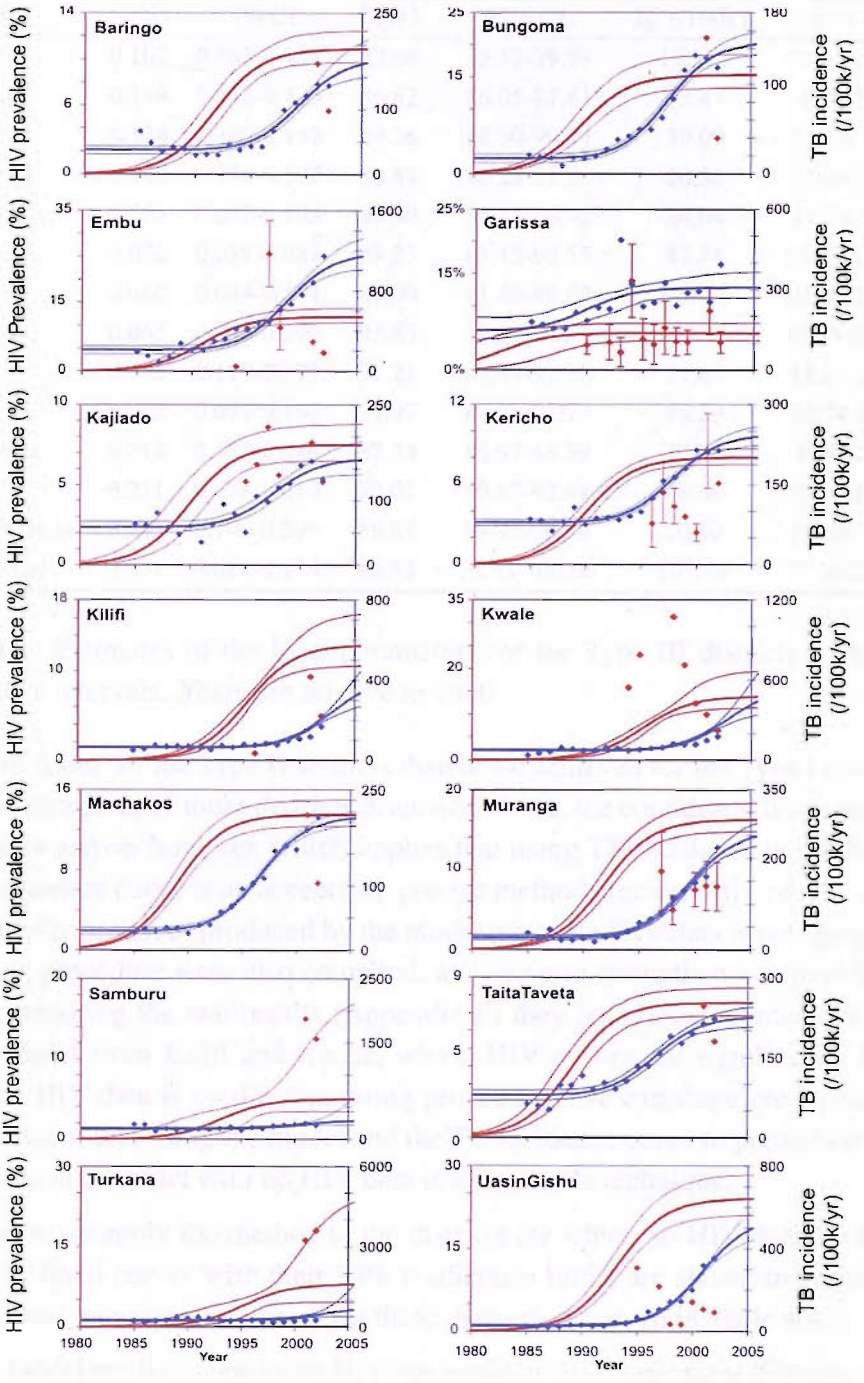


Figure 4.3: TB incidence (blue) and HIV prevalence (red) for the Type II districts. The graphs show, for each district, the TB incidence data and HIV prevalence data (with binomial confidence limits), and the TB incidence and HIV prevalence curves produced by the model with their 95% confidence intervals

District	a (%)	95% CL	$\bar{t}$ (yr)	95% CL	$\bar{I_0}$ (/100k)	95% CL
Isiolo	0.102	0.082-0.146	87.66	85.52-89.59	135.21	72.06-187.01
Kiambu	0.189	0.146-0.335	86.82	86.05-87.61	13.47	4.36-21.56
Kisii	0.124	0.107-0.152	89.26	88.30-90.24	30.09	21.16-38.70
Kirinyaga	0.149	0.134-0.167	90.89	90.28-91.50	20.56	16.49-24.41
Laikipia	0.093	0.083-0.105	89.60	88.33-90.70	26.04	21.74-30.61
Lamu	0.052	0.038-0.087	88.23	83.12-93.55	87.74	57.31-109.27
Mandera	0.060	0.044-0.104	85.00	81.45-89.07	196.27	86.29-258.26
Marsabit	0.065	0.045-0.290	85.85	79.59-91.36	159.36	10.85-223.48
Nandi	0.145	0.119-0.177	91.21	89.89-92.38	17.64	12.03-23.12
Narok	0.082	0.057-0.198	91.97	86.28-97.73	46.39	26.74-58.73
Nyandarua	0.214	0.158-0.446	87.74	86.97-88.59	7.79	1.80-13.64
Siaya	0.231	0.178-0.410	92.02	90.87-93.44	18.68	5.71-31.19
South Nyanza	0.227	0.192-0.296	89.88	89.39-90.38	20.50	12.19-28.38
West Pokot	0.031	0.014-0.176	86.93	76.33-100.00	203.04	25.06-229.54

Table 4.4: Estimates of the local parameters for the Type III districts with their confidence intervals. Years are relative to 1900

curve are wider for the Type II districts than those achieved for the Type I districts. With the exception of those districts discussed above, the confidence limits are still reasonably narrow however, which implies that using TB incidence to predict the HIV prevalence curve is an acceptably precise method. Incidentally, results of the HIV prevalence curves produced by the model when the HIV data is not ignored in the fitting procedure were also compiled, and serve to strengthen our confidence. When comparing the two results (Appendix F) they are almost identical in most districts apart from Kilifi and Kwale, whose HIV curves are significantly lower when the HIV data is used in the fitting procedure. We can therefore be reasonably confident that using the model and the TB incidence curve to predict the HIV prevalence in a district with no HIV data is a justifiable technique.

Finally, we apply the method to the districts for which no HIV data are available. The fitted curves with their 95% confidence limits are shown in Figure 4.4 and the local parameter estimates for these districts are given in Table 4.4.

The model predicts significant HIV epidemics in all except three districts: Tana River, Wajir and Elgeyo Marakwet. The TB data in these three districts is very uncertain and does not show clear trends. It is possible that the TB data are not sufficiently good to be able to draw any conclusions for these three districts.

The TB data suggest that the prevalence of HIV in South Nyanza is very high

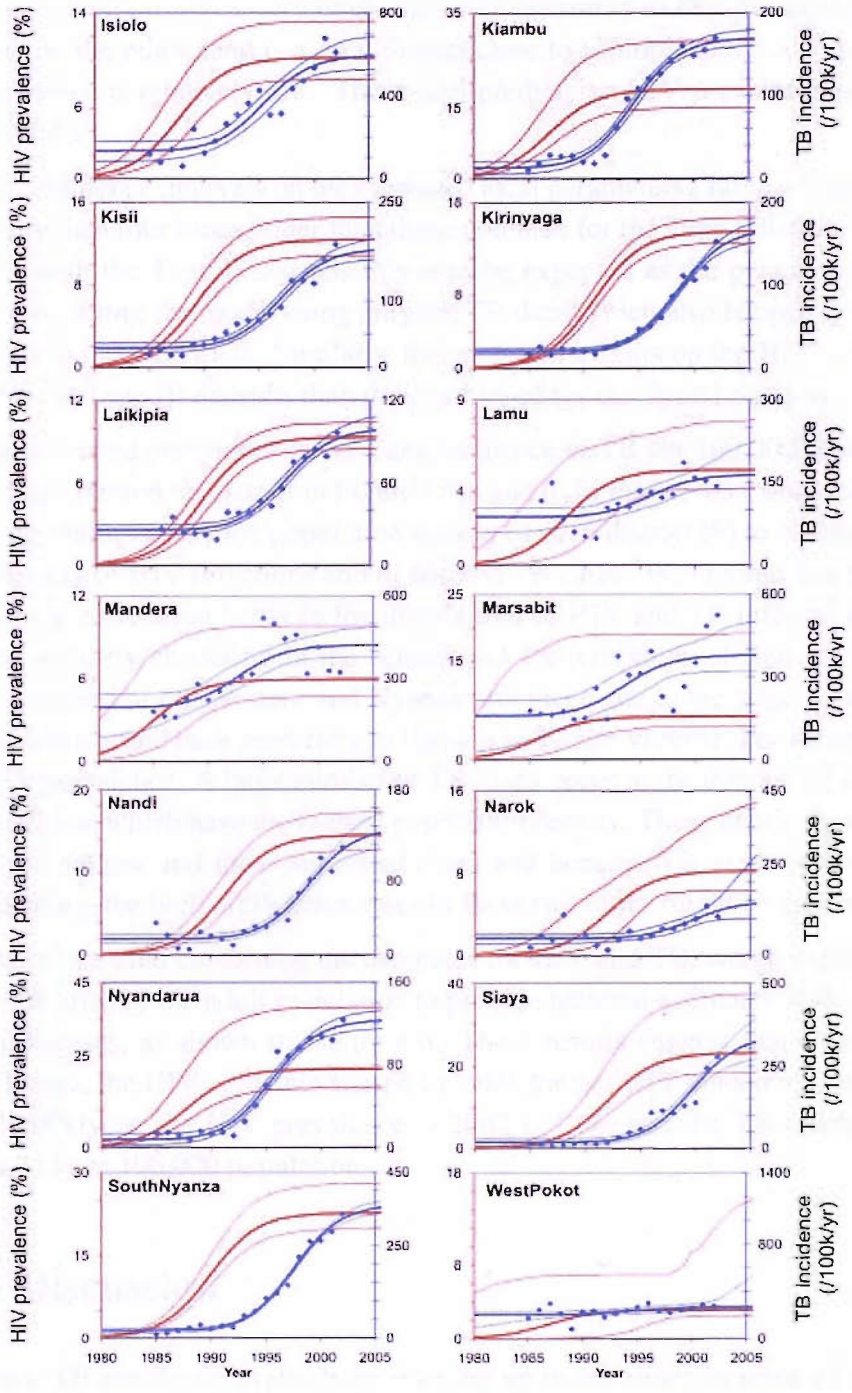


Figure 4.4: TB incidence (blue) and HIV prevalence (red) for the Type III districts. The graphs show, for each district, the TB incidence data and HIV prevalence data (with binomial confidence limits), and the TB incidence and HIV prevalence curves produced by the model with their 95% confidence intervals

(23%; 19%-28%) which is very likely due to its proximity to Kisumu and Uganda. Mandera on the other hand is a rural district close to Ethiopia and Somalia where the prevalence is relatively low. The model predicts an HIV prevalence of 6.0% (4.6%-9.1%).

The confidence intervals on the estimated local parameters (Table 4.4) are typically more than four times wider than those obtained for the Type I districts (Table 4.2). As with the Type II districts this is to be expected as the parameters were obtained by fitting the model using only the TB data, which also happen to be less consistent in these districts. Similarly, the confidence limits on the HIV curve are wider for the Type III districts than those achieved for the Type I districts.

The estimated prevalence of HIV and incidence of TB per 100,000 population for 2005 are plotted by district in Figure 4.5 a and b. In Figure 4.5 c and d the data have been multiplied by the population density of each district [8] to obtain a map of the density of HIV infections and of notified TB cases. We can see that there is a noticeable correlation between the distribution of HIV and TB infected people, with the majority clustering in the central and western areas of Kenya. HIV is highly clustered in the Western and Nyanza provinces due to the large population in these districts and their proximity to Uganda and Lake Victoria, known areas for high HIV prevalence. A large number of TB cases occur in the districts of Nairobi and Mombasa, which have the highest population density. These districts are home to Kenya's largest and most important cities and because TB is associated with overcrowding, the high notification rates in these two cities might be expected.

Finally, we tried combining the estimates for HIV and TB, weighting the data for each district by the adult population to provide national estimates of the trends in both diseases, as shown in Figure 4.6. These results suggest that on average across Kenya, the HIV epidemic started in 1989, the pre HIV epidemic rate of TB was 73/100k/year, the HIV prevalence in 2007 is 14%, and the TB incidence in 2007 is 313 per 100,000 population.

4.6 Discussion

In Kenya, TB notification rates have risen by up to ten times in some of its districts over the last ten years as a result of the HIV epidemic. Because of this phenomenon and due to the rich TB data available from the country's surveillance program, building a simple parametric model to investigate the interactions between HIV and TB gave some understanding of the link between the epidemics.

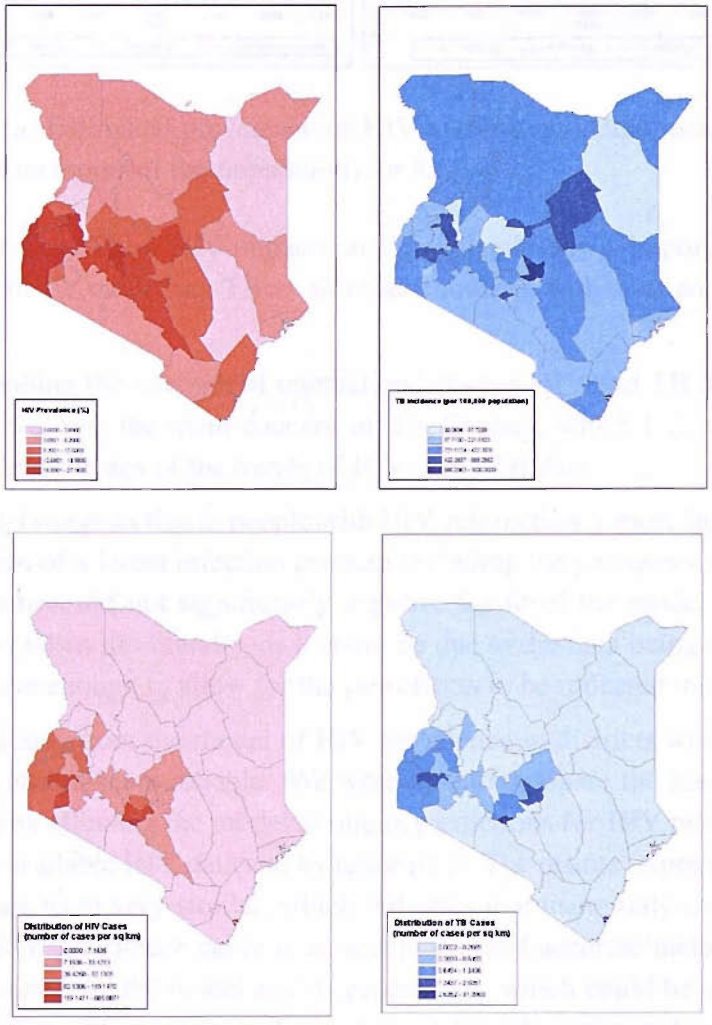


Figure 4.5: Estimated (a) adult prevalence of HIV and (b) TB incidence for each district of Kenya 2005. Estimated (c) number of prevalent cases of HIV among adults and (d) number of notified TB cases per square kilometre in 2005.

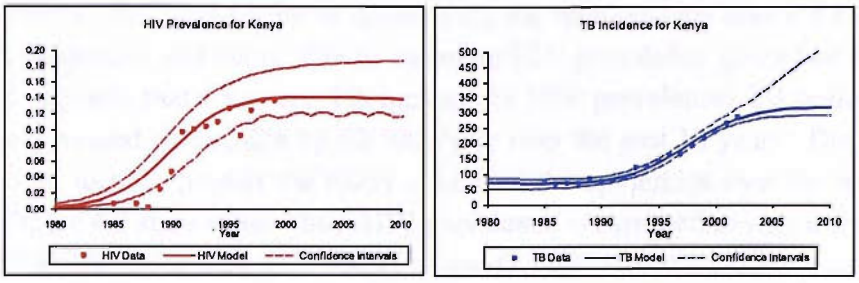


Figure 4.6: (a) Estimated prevalence of HIV and (b) estimated incidence of TB (per 100,000 members of the population) for Kenya

An understanding of how HIV impacts on TB epidemiology is important, not only for Kenya, but for modelling TB in all other countries with substantial HIV epidemics.

Understanding the strength of interaction between HIV and TB disease at the population level was the main concern of this Chapter, which has enabled us to make sensible estimates of the trends of HIV using TB data.

The model suggests that in people with HIV, reinfection is more important than the breakdown of a latent infection because including the parameter which represents reactivation did not significantly improve the fit of the model. This result should not be taken too literally, as it could be due to the data being deficient and not informative enough to allow for the two effects to be reflected independently.

The estimates from the model of HIV prevalence in districts where HIV data are weak or absent seem reliable. We were able to compare the resulting prevalence curves by allowing the model to obtain predictions for HIV prevalence both by using the available HIV data and by ignoring it. The results (Appendix F) show that the curves were very similar, which indicates that using only the TB data to predict the HIV prevalence curve is an acceptable and accurate method. This increases confidence in the model and its predictions, which could be tested further against field data. The confidence limits obtained for the estimated curves of each of the districts were very encouraging. The limits are relatively narrow which implies that using the TB incidence to predict the HIV prevalence curve in a district is a realistically precise method.

The fact that the results were generally so consistent and the predictions so reliable gives us evidence that there is justification for thinking that the force of infection of HIV, the time delay between the two epidemics (7.65 years; 7.43-7.88 years) and the effect of reinfection on HIV associated TB should not vary depending on district.

The model has been useful in quantifying the relationships between the HIV and TB epidemics and being able to calculate HIV prevalence given just the TB data. It suggests that for every 1% increase in HIV prevalence, TB notification rate has increased on average by 62/100k/year over the past 10 years. The model can also be used to predict the likely course of the epidemics over the next ten years. Figure 4.7 shows that whilst HIV prevalence is expected to start decreasing from 2004, TB incidence levels will continue to increase but at a slower rate than has previously been experienced. Currently in some countries in Africa, a very sharp decrease in HIV prevalence has been observed. By making the scenario assumption that the future course of the HIV epidemic will follow this dramatic reduction and reduce by 10% each year until it settles at half its current level, we can use the model to estimate the effect that this will have on TB. Figure 4.8 shows that as HIV prevalence is reduced, TB incidence is also expected to fall by on average 15/100k/year for every 1% decrease in HIV prevalence.

The key assumptions that we have made in this model are that:- the TB notification rate in HIV-negative people is constant, TB incidence is linearly dependent on HIV prevalence, TB incidence is proportional to TB notification, and adult HIV-prevalence is proportional to ANC prevalence. The importance of these assumptions is still unclear and until better data sets will allow a close analysis of these issues, they will remain unresolved. However, the assumptions that the model use are widely accepted throughout current HIV and TB literature.

It was initially thought that the work would have been useful in quantifying the two processes responsible for HIV-associated TB, and could supply the DES model with the value of the parameters which measure the relative contribution of reinfection and reactivation. The data did not allow these effects to be reflected independently however, and the parameter which represented reactivation ended up being removed. The results of this initial modelling work could not be used as a direct input into the DES model, however, this work does contribute toward the evidence of a strong link between the HIV and TB epidemics. This link is a key justification behind the development of the DES model of TB transmission which will allow the impact of the HIV epidemic on the relative efficacy of household interventions for TB control to be fully assessed.

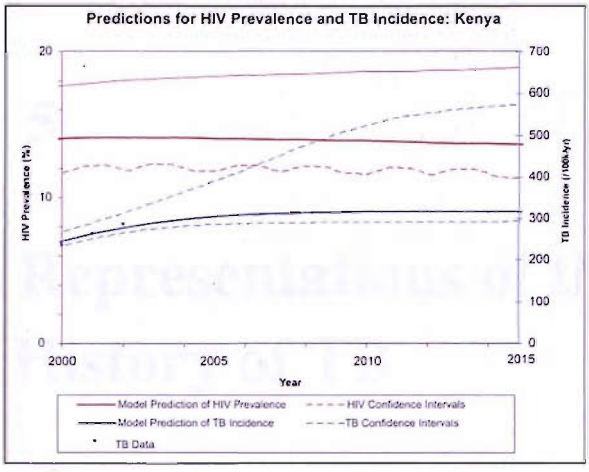


Figure 4.7: The model predictions of the likely course of the HIV and TB epidemics in Kenya over the next ten years

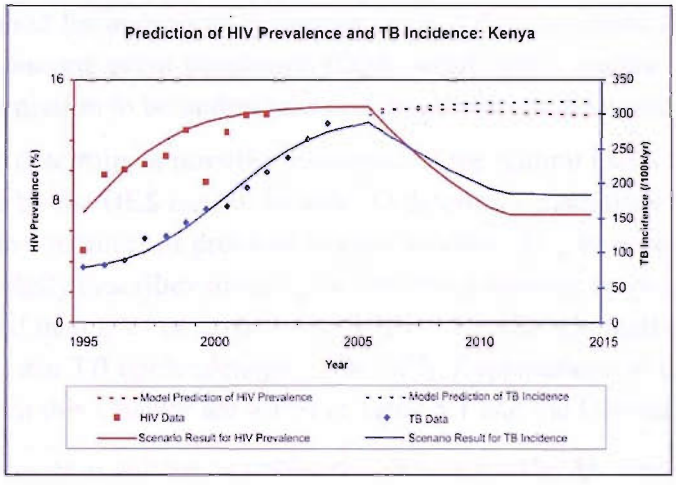


Figure 4.8: The likely course of the HIV and TB epidemics in Kenya and the likely course of the scenario assumption that HIV will reduce by 10% each year over the next ten years

Chapter 5

Possible Representations of the Natural History of TB

5.1 Introduction

The literature review (Chapter 3) showed that the majority of previous TB models had been deterministic compartmental models (DCMs) [1]. It concluded that despite all of the studies there is still a need to identify TB control strategies that are effective in high HIV prevalent settings. The review discussed why DCMs are an unsuitable method for investigating interventions at the household level and highlighted how a discrete event simulation (DES) would allow for the more intricate details of transmission to be understood and therefore would be more appropriate.

We need to determine a possible schematic of the natural history of TB which can be adopted by the DES model. In order to determine a possible representation we consider the structures of previous studies and how they have represented TB. This Chapter briefly describes some of the structures adopted by previously developed DCMs and then goes on to describe SEEINITR, the schematic that has been chosen to represent TB epidemiology in the DES. Explanations of the majority of the terms used in this Chapter are given in Table 5.1 and the Glossary.

[1] A DCM divides the population into different epidemiological groups according to their TB disease status and uses differential equations to move proportions of the population through the various groups/states at specified time steps

Term	Abbreviation	Description
Susceptible	S	Those that are uninfected and therefore susceptible to infection
Latent	E	Those that have TB infection. An ***infection*** means that the TB mycobacterium are present in the body but that they are not actively causing damage to body tissues
Fast Latent	E_f	Those that have a TB infection and will progress quickly to **active disease** (A).
Fast Latent – Infectious	E_{fI}	Those that have a TB infection and will progress quickly to **infectious** (I) active disease.
Fast Latent – Non Infectious	E_{fNI}	Those that have a TB infection and will progress quickly to **non infectious** (NI) active disease.
Active Disease	A	Those that have TB disease. The *disease* means the TB organisms are growing and causing damage within the body. Within this compartment you can be **infectious** (I) or **non infectious** (NI).
Infectious	I	Those that have active TB disease and are able to transmit the infection. $I = f A$.
Non Infectious	NI	Those that have active TB disease but are unable to transmit the infection.
Treatment	T	Those that are having treatment for **active disease** (A).
Treatment – Infectious	T_I	Those that are having treatment for **infectious** (I) active disease.
Treatment – Non Infectious	T_{NI}	Those that are having treatment for **non infectious** (NI) active disease.
Recovered	R	Those that have successfully completely **treatment** (T) for **active disease** (A).
Self Cure	N_I, N_{NI}	Those that have self cured from either infectious (N_I) or non infectious (N_{NI}) active disease.

Table 5.1: An explanation of the terms used

5.2 The Structures of Previous Models

The majority of previous TB models have been deterministic compartmental models. This Section briefly describes some of the structures adopted by these previously developed DCMS.

5.2.1 SEA Model

The SEA model has a similar structure to that used by Blower *et al.* [31] [28], Ziv *et al.* [206] and Song *et al.* [155] where the population is divided into Susceptible (S), Latent (E) and Active Disease (A) states. It is thought that this structure would perhaps be appropriate for investigating the effect of a very basic vaccination programme in a homogeneous population.

5.2.2 SEAR Model

The SEAR model has a similar structure to that used by Blower *et al.* [30], Garcia *et al.* [77] and Porco and Blower [138] where the population is divided into Susceptible (S), Latent (E), Active Disease (A) and Recovered (R) states. This structure would perhaps be appropriate for investigating the effects of a very basic treatment programme in a homogeneous population. The structure was used in this way by Garcia.

5.2.3 SEAT Model

The SEAT model has a similar structure to that used by Castillo-Chavez and Feng [37] and Gomes *et al.* [84] which divided the population into Susceptible (S), Latent (E), Active Disease (A) and Treatment (T) states. This structure would perhaps be appropriate for investigating drug resistant TB. The structure was used in this way by Castillo-Chavez and Feng.

5.2.4 SEATR Model

The SEATR model is a similar but simpler version of that used by Debanne *et al.* [59]. This model has five states: Susceptible (S), Latent (E), Active Disease (A), Treatment (T) and Recovered (R). Debanne used it to look at projections of TB

incidence in different sociodemographic groups where multi drug resistant TB was an issue.

5.2.5 SEEATR Model

The SEEATR model is the same as the SEATR model but it introduces the concept of having two 'latent' compartments. One compartment (Latent, E) represents those that have effective immune responses to TB infection and will therefore have a latent infection for many years until either they die or it is reactivated by the immune system being weakened; or they are reinfected and do not have an effective immune response. The other compartment (Fast Latent, E_f) represents those that do not have an effective immune response to the TB infection and will therefore progress quickly to 'active disease' (A). In all the previous models this has been represented by a flow directly from 'susceptibles' to 'active disease', however this implies that the progression to active disease is instantaneous which is not realistic. The introduction of a 'fast latent' compartment is to enable the time frame of developing active disease, which can be as long as five years, to be modelled. A few models, which include the division of the latent compartment to reflect fast and slow progression to active disease, have been developed [181] [120] [197] [56]. This structure is appropriate for investigating exogenous reinfection and endogenous reactivation and the role they play on TB morbidity. The structure was used in this way by Vynnycky and Fine [181].

5.2.6 SEEINITR Model

This model builds on the SEEATR model but introduces the concept of dividing the active disease state into those that are infectious and those that are non infectious. It also includes states to represent those that self cure. This model is largely inspired by the model developed by Williams *et al.* [197]. This structure is appropriate for investigating the impact of HIV and heterogeneity on TB control. The structure was used in this way by Williams.

5.3 Choosing a Structure

We have considered the structures of the main deterministic compartmental models from literature. It has been shown that the models with simple configurations

have been useful for addressing questions surrounding vaccination and treatment of homogeneous populations [206] [77]. As the questions have become more involved however, the structures of the DCMs have become more complicated. For example, questions regarding the impact of multi drug resistance and exogenous reinfection have required the structures to incorporate additional states so that specific questions regarding their role on TB morbidity, for example can be explored.

If the questions require many aspects of TB to be considered, it is not possible to keep the model simple and instead a detailed structure is required. Williams *et al.* [197] for example, needed to capture spatial and temporal variation of TB amongst the different risk groups in India to allow them to explore the impact of HIV on TB control in different areas, and used the detailed SEEINITR structure (Section 5.2.6).

In Chapter 3, Section 3.3 we discussed the epidemiological aspects of TB identified in the literature, and determined which of these aspects would be included in our study. In order to determine the relative importance of household interventions in controlling TB in HIV prevalent settings, we identified HIV, age dependence, non-homogeneous mixing, reinfection, and varying lengths of latency and infectiousness to be necessary to the model.

Due to the number of aspects we expect to incorporate into the model to capture the epidemiology of TB in HIV prevalent settings, it is clear that a simple structure will not be sufficient. As many aspects need to be addressed, it is felt that a model which enables all these issues to be incorporated and their effects to be seen, would need to be a comprehensive model which allows the epidemiology of TB to be fully represented. This would mean using a configuration similar to SEEINITR (Section 5.2.6). The schematic of this model is given in Figure 5.1.

This structure was suggested to the collaborators, DETECTB, and was presented to subject matter experts such as epidemiologists, mathematical disease modellers and clinicians. The consensus from these meetings was both that we have correctly identified the aspects important to the epidemiology of TB in HIV prevalent settings, and that the structure proposed is sufficient to model these aspects. They were also satisfied because the model allows all of the possible pathways an individual infected with TB might experience, to be represented.

All of the aspects of TB we highlighted for accurately modelling an HIV-associated epidemic were considered important by the subject matter experts and therefore none were removed. There was some discussion over the importance of adding multi drug resistant (MDR) TB, however the consensus was that although this is important due to its potential impact on the efficacy of control methods, the

questions it raises are different to those being considered by this research and it would require a slightly different model (where TB 'hotspots' such as hospitals are represented). Currently unavailable data on the prevalence and virulence of MDR TB in Harare would also be required.

We have chosen to implement the SEEINITR structure in the DES model. It is thought that it gives an appropriate schematic representation of the natural history of TB for the purposes of our research. The structure allows issues such as reinfection, conversion of non infectious TB and the self curing of TB, which were raised in the literature and by the subject matter experts, to be captured. These issues could not be addressed by simpler structures.

Basing the structure of the discrete event simulation model on the SEEINITR model implies that individuals within the DES model will belong to seven epidemiological groups according to TB status: Susceptible, Latent (with fast and slow progression), Infectious, Non Infectious, Self Cured, Treated and Recovered.

In the next Section we describe how a deterministic compartmental model was built using the SEEINITR structure, in order to ascertain the spread of the population amongst the epidemiological groups once the model has reached a steady state. This distribution will be an input into the DES model and will inform it of the likely TB status of the individuals when they are created at the start of the simulation.

5.4 Building a DCM of the SEEINITR Structure

We use Berkeley Madonna [24] to develop a DCM using the SEEINITR structure. Berkeley Madonna is a general purpose differential equation solver. It allows you to assemble a graphical flow chart of the "system" meaning we can construct a visual representation of TB disease progression whilst the program generates the differential equations. The DCM can be developed very quickly and conveniently and the use of graphics makes the structure easy to evaluate. Berkeley Madonna also allows parameter exploration by using parameter sliders, plots and sensitivity analysis which means the role of various parameters can easily be ascertained.

Because the majority of models developed for Africa in the past have been DCMs, the relevant parameter values have already been established and can therefore be found in the literature already reviewed. Descriptions and details of the parameters and the values used can be seen in Table G.1 in Appendix G.

5.4.1 General Notes

The infection of an individual:

The risk that an individual becomes infected with TB during a given time step depends on two factors:

1. The number of infectious individuals in the population (I)
2. The probability that the individual comes into effective contact with an infectious individual (β)

An 'effective' contact is a contact that will lead to an infection. If one knows the effective contact rate (ecr), which represents the number of individuals each person effectively contacts per time step, and the total population size (N), then β can be estimated as

$$\beta = \frac{ecr}{N}, \tag{5.1}$$

which is the probability that two specific people will come into effective contact per time step. The number of susceptibles infected per time step can therefore be given by βIS, where IS gives the total number of possible contacts between a susceptible and an infectious individual and β gives the probability of each of those contacts being 'effective'.

Births and Deaths:

Although not specifically mentioned in the analysis, people can leave all of the states discussed by dying at the death rate (μ) given in Table G.1. When people are in any of the states involving active disease their death rate is increased by the disease induced death rates (d) in Table G.1.

People are also recruited into the susceptibles class at the birth rate. This corresponds with babies not inheriting any special immunity from their mothers. The birth rate in these models is equal to the death rate and so the number of births per time step (Λ) is given by the total number of deaths in that time step. This enables the total population size (N) to stay constant.

We keep the population size stable as this is something seen in the majority of previous DCM studies. Stationary populations are useful as they allow the equilibrium states of the models to be characterised analytically. This can lead to a better understanding of the relationships between fundamental epidemiological parameters. A stable environment also means that when the model is simulated,

for instance to investigate the consequences of different interventions for control, we can have confidence that demographic factors are controlled, and the different outcomes can therefore be explained by purely epidemiological factors.

Fast and Slow Latent Individuals:

When a person has been infected with TB, their immune system may invoke a successful response which "walls off" the infection stopping it from causing damage. This is called a latent (E) TB infection. Alternatively, the immune system may fail to respond successfully and the infected person will then quickly progress to active TB disease, usually within months.

To represent these two possibilities, when susceptibles have been infected they can follow one of two routes. The first represents an effective immune response, where people become latently infected (E), and the second represents fast progression to active TB disease (E_f). The proportion of those infected that will develop primary active disease is given by the parameter p. Therefore, using the expression βIS, the number of susceptibles that will be infected and will develop primary TB disease is equal to βISp, and the number of susceptibles that will have a long-term latent infection is given by $\beta IS(1-p)$.

Reinfection and Reactivation:

Persons with a latent infection can progress to active disease in two ways: reactivation of their TB infection at a rate v or through reinfection. An individual with a latent infection has some immunity to a new infection, with only a proportion x being susceptible to active disease. Therefore $vE + \beta IEpx$ individuals with a latent infection move to active disease in each time step.

Failed Treatment:

When individuals are said to fail treatment this indicates that they failed to correctly complete the course of drugs (usually a 6 month course of Isoniazid) and therefore the TB lesions were not sterilized. Failed treatment therefore means that individuals have active disease.

5.4.2 The Model

The following section gives a description of the model and shows how the system of differential equations were derived. The schematic of the model can been seen in Figure 5.1 and a summary of the parameters and their values is provided in Appendix G.

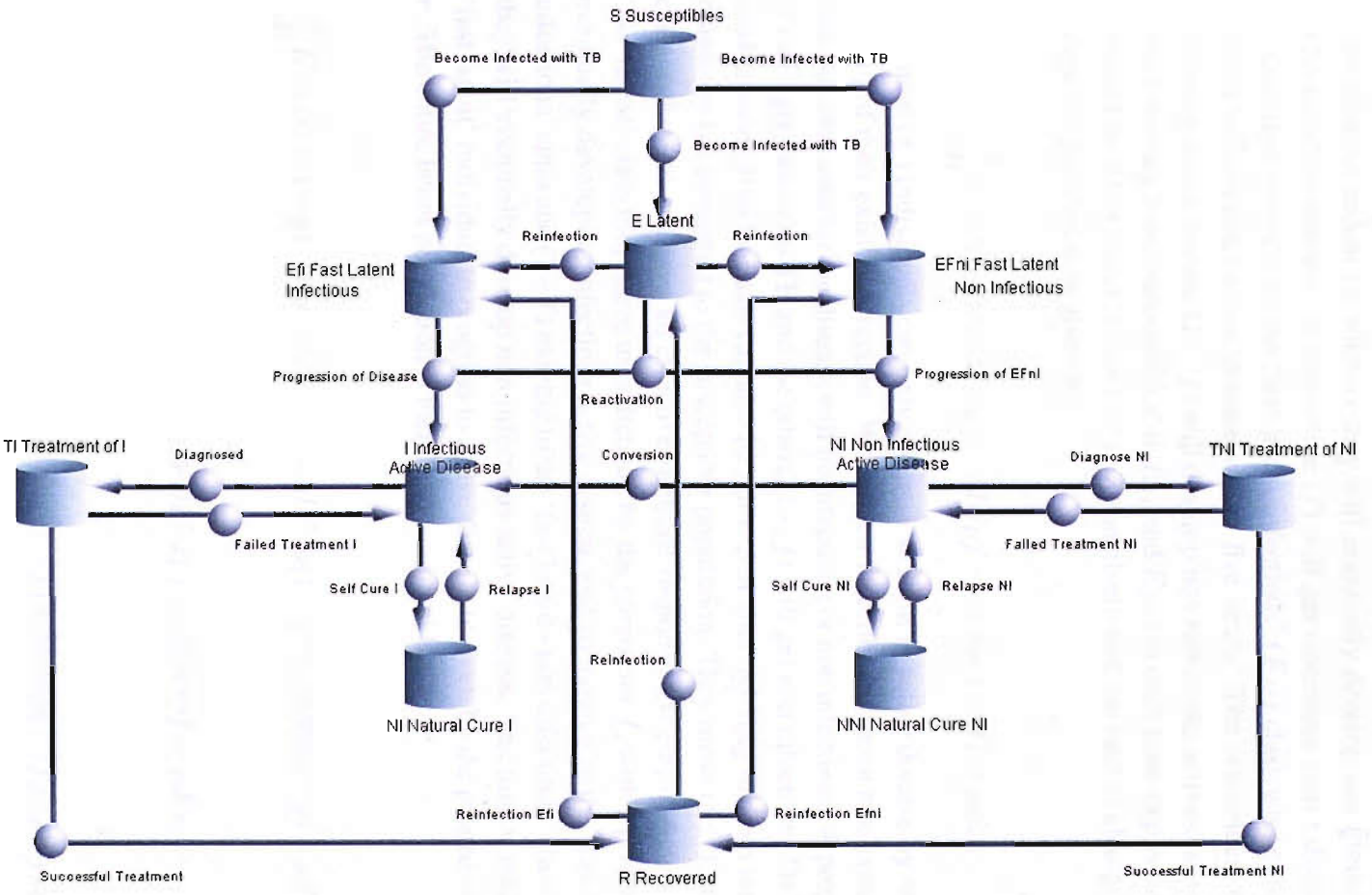


Figure 5.1: Schematic of the SEEINITR Model. This model is based on the model developed by Williams *et al.* [197]

When a susceptible is infected they will have either a 'latent' or 'fast latent' status ($\beta IS(1-p)$ or βISp respectively). Those moving to a 'fast latent' status are divided dependent on whether they will eventually develop infectious or non infectious active disease. A proportion (f) will get infectious active disease and will therefore progress to the 'fast latent infectious' (E_{fI}) class where they will progress to infectious active disease within five years. The remainder of those developing active disease ($1-f$) will develop non infectious active disease. The number moving from 'susceptibles' to E_{fI} and E_{fNI} in each time step is therefore expressed as $\beta ISpf$ and $\beta ISp(1-f)$ respectively and the rate of change of the susceptible population is given by

$$\frac{d}{dt}S = \Lambda - \beta IS(1-p) - \beta ISpf - \beta ISp(1-f) - \mu S. \tag{5.2}$$

Latent (E) individuals can only progress straight to active disease by reactivation (v) of their existing infection. When their infection has been reactivated, it is then decided whether the disease will be infectious or non infectious. A proportion (f) will get infectious TB and the others ($1-f$) will get non infectious TB. Latent (and recovered) individuals can also be reinfected although they have an increased immunity (x) compared to the susceptible population. They move to a 'fast latent' class if they do not have an effective immune response ($\beta IExp$). Again, which 'fast latent' class they move to is dictated by the parameter f, with a proportion f eventually developing infectious active disease and so moving to the 'fast latent - infectious' class and $(1-f)$ moving to the 'fast latent - non infectious' class, where they will eventually develop non infectious active disease. The time it takes for a 'fast latent' individual to progress to active disease is given by the progression rate r. The entire latent population is therefore described by

$$\frac{d}{dt}E = \beta IS(1-p) - Ev - \beta IExpf - \beta IExp(1-f) + \beta IRx(1-p) - \mu E. \tag{5.3}$$

$$\frac{d}{dt}E_{fI} = \beta ISpf - rE_{fI} + \beta IExpf + \beta IRxpf - \mu E_{fI} \tag{5.4}$$

$$\frac{d}{dt}E_{fNI} = \beta ISp(1-f) - rE_{fNI} + \beta IExp(1-f) + \beta IRxp(1-f) - \mu E_{fNI}. \tag{5.5}$$

When an individual has active disease, the model allows them to be diagnosed at rate ϕ and move out of the class to receive treatment. It also allows them to

self cure at the natural cure rate scr and to die at the disease induced death rate d. All of these rates are dependent on whether the individual is infectious or non infectious.

After diagnosis an individual will move to the treatment state. Individuals receive treatment for a duration of time (td) and a percentage of them are cured cr. Cured individuals are said to be recovered however those individuals that fail treatment ($1 - cr$) will return to the active disease state.

When in the 'self cured' state individuals are able to relapse into whichever active disease state they previously occupied at the relapse rate s.

The population with active disease can be represented by

$$\frac{d}{dt}I = vEf + rE_{fI} - \phi_I I - scr_I I + \frac{1-cr}{td}T_I + sN_I + nNI - (\mu + d_I)I \quad (5.6)$$

$$\frac{d}{dt}NI = vE(1-f) + rE_{fNI} - \phi_{NI}NI - scr_{NI}NI + \frac{1-cr}{td}T_{NI} + sN_{NI} - nNI - (\mu + d_{NI})NI. \quad (5.7)$$

The relapse rate s of the self cured individuals does not depend on whether they were previously infectious or non infectious. We can describe the rate of change of the self cured population as

$$\frac{d}{dt}N_I = scr_I I - sN_I - \mu N_I \quad (5.8)$$

$$\frac{d}{dt}N_{NI} = scr_{NI}NI - sN_{NI} - \mu N_{NI}. \quad (5.9)$$

There are two treatment states to represent those being treated for infectious and those being treated for non infectious TB. When individuals fail treatment they automatically return to their previous type of active disease. These arrangements assume that treatment outcomes occur at the end of treatment and that those on treatment are not infectious. The rate of change of those on treatment for TB is given by

$$\frac{d}{dt}T_I = \phi_I I - \frac{1}{td}T_I - \mu T_I \quad (5.10)$$

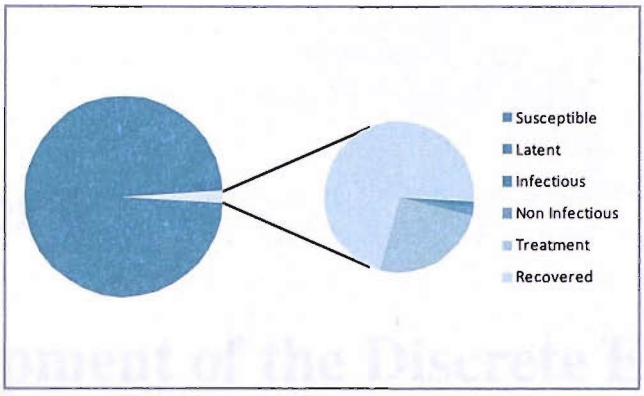


Figure 5.2: The distribution of the population amongst the various epidemiological groups once the SEEINITR model has reached a steady state. This is used to establish the initial distribution of the DES model's population amongst the epidemiological classes.

$$\frac{d}{dt}T_{NI} = \phi_{NI}NI - \frac{1}{td}T_{NI} - \mu T_{NI}. \tag{5.11}$$

Finally then, the recovered population is described by the following equation,

$$\frac{d}{dt}R = (T_I + T_{NI})\frac{cr}{td} - \beta IRx - \mu R. \tag{5.12}$$

A useful output of the model is TB incidence. The equation used to obtain TB incidence consists of those latent individuals whose infection reactivates (vE) and those fast latent individuals who progress to active disease ($r(E_{fI} + E_{fNI})$),

$$\text{TB Incidence} = vE + r(E_{fI} + E_{fNI}). \tag{5.13}$$

In order to determine the spread of the population amongst the various epidemiological groups at a steady state, we solved the deterministic compartmental model using Berkeley Madonna. Figure 5.2 shows the distribution of the population which will be used to inform the DES model of the likely TB status of the individuals when they are created at the start of the simulation.

Chapter 6

Development of the Discrete Event Simulation Model

In this Section we describe the development of a discrete event simulation model which will be used to evaluate the effects of more intensive case-finding strategies for TB control in a high HIV prevalent setting.

As previously discussed, the end goal of the modelling is a discrete event simulation (DES) model of TB transmission in Harare, Zimbabwe which will allow an assessment of the effectiveness of contact-tracing and case-finding strategies. The first stage, which we describe in this Chapter, is designing the simulation and its processes for generating and maintaining the population, warming up the model, incorporating HIV and managing its inputs and outputs. The second stage is discussed in the next Chapter which will look at how the various aspects of the natural history of TB are incorporated into the model and how it is parameterised so that it corresponds with earlier deterministic models of TB and HIV. It will also discuss how the model was validated using country-wide statistics for Zimbabwe and findings from previous studies of the distribution of TB amongst populations.

6.1 An Overview of the Simulation

The simulation follows the structure of the population, the TB disease characteristics of individuals, the HIV epidemic, the intervention scenarios and the model outcomes. The user defines the number of years that the simulation will run for, the underlying population structure, the available prevalence and incidence data on the HIV and TB epidemics and the number of replications required. The model then warms-up the initial population and runs for a predefined length of time in order

to eliminate transients and to establish a population which correctly reflects that of 1980 in both its characteristics and TB incidence. We have chosen to start the simulation from 1980, warming the population up to this point, as TB incidence data becomes available from this point onwards (discussed in Section 6.4).

Activities associated with the dynamics of the population (discussed in Section 6.3), TB infection events leading to the formation of various disease, transmission and treatment events (discussed in Chapter 7), and HIV transmission events (discussed in Section 6.5) are all evaluated throughout each year using a next event method as shown in Figure 6.1. At the end of each year various outputs regarding the model's population, such as its TB incidence and HIV prevalence are recorded (discussed in Section 6.6).

The user can define various interventions which take effect in 2008, and aim to adjust an individual's disease pathway by reducing the individual's risk of getting active disease and of reactivating, and by decreasing their period of infectiousness and thus the number of likely transmission events produced.

6.2 The Simulation Design

The model was implemented using Object-Orientated Programming (OOP) techniques with the C++ programming language and the .Net framework. The main development of the model has been done using a console interface, which outputs comma separated value files for analysis using Excel. The maximum population size is limited by computer memory and constraints on run time, otherwise populations of any size can be represented and evaluated. The model has been designed to allow easy maintenance for further development in order to incorporate improved data and understanding of parameter values, disease processes and epidemiological complexities such as contact networks. Full documentation of the simulation model is given in Appendix H.

The population is first established by creating a house, which is an object or instance of the household class. "The term object refers to an instance of a class and thus a class defines the behaviour of possibly many objects (instances)." [129] A class will define the attributes or *members* that all of its instances will possess and therefore each instance of the class will have a copy of the class members. A comprehensive and detailed documentation of the simulation's classes and their members is given in Appendix H.3. The household class (Table H.5) defines three main attributes: its unique identification number (*itsID*), a list of its occupants (*itsOccupants*) and the number of occupants (*itsHouseholdSize*). As well as each

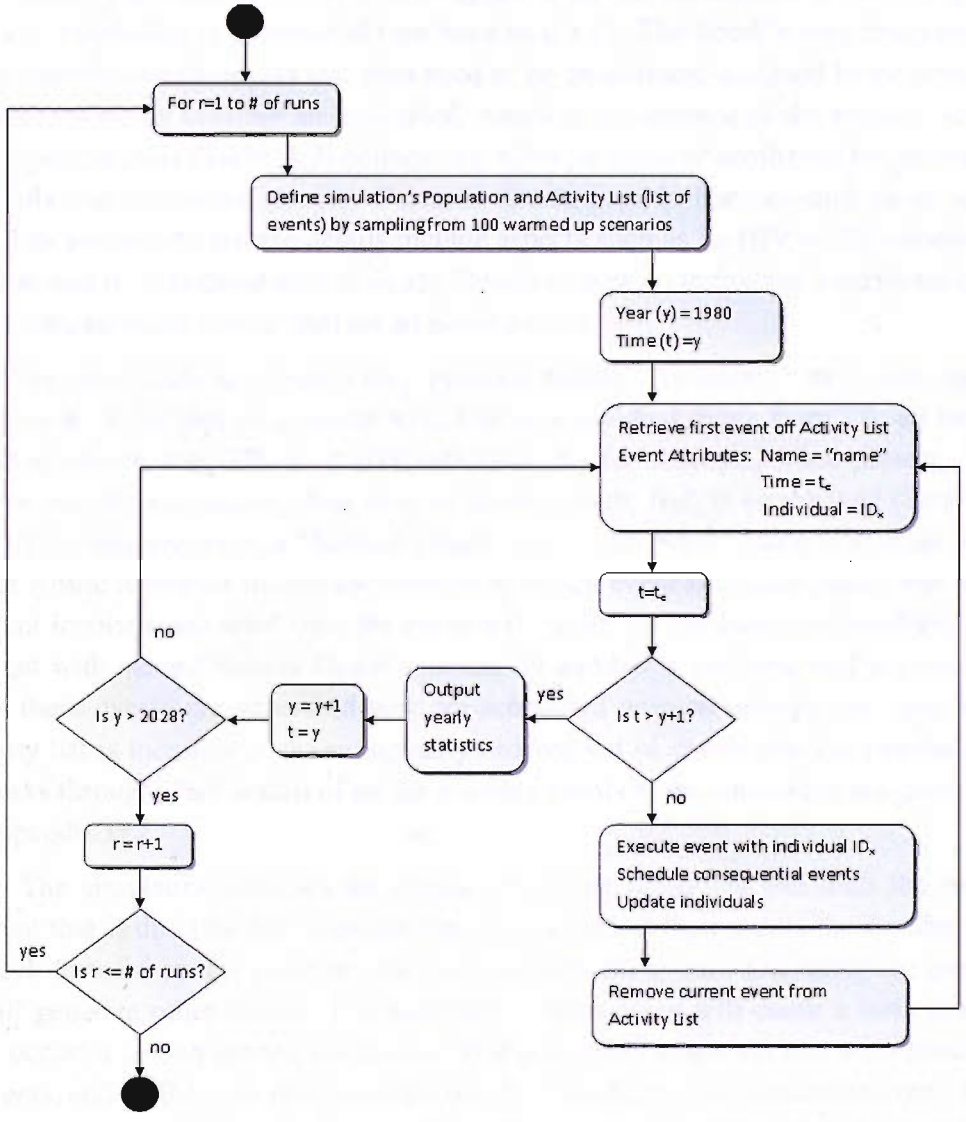


Figure 6.1: Computation activity diagram

household having this member data, many member functions are also defined. These functions enable each house to answer questions such as whether it contains an adult and who an individual lives with, which is useful when considering household transmission of disease.

When a house is created, it is assigned a unique identification number and *itsHouseholdSize* is determined (see Section 6.3.1). The house's size determines the number of individuals that then need to be created and assigned to the house. This is done by creating an individual, which is an instance of the person class. The person class (Table H.3) defines two different types of attributes; its personal details and its disease details. Personal details include features such as its age and its gender; its disease details include aspects such as its HIV or TB infection status and its scheduled time of death. Details of how an individual's attributes are determined when it is created are in Section 6.3.3.

As individuals are created they generate events. An event is an action upon a person. Examples of an event would be an individual dying from TB, an individual transmitting TB or an HIV infection. So for example, when person x is born into the simulation, their time of natural death, tnd, is established (Section 6.3.3.3); this generates a "Natural Death" event. An event (Table H.4) is an object whose attributes inform the simulation which event is to take place, who the event involves and what time the event will occur. In the example therefore, an event with name "Natural Death", person ID number x and time tnd is created. As these events are generated they are scheduled onto an activity list. The activity list is therefore a chronologically ordered list of events that the simulation works through. Full details of all the possible events in the simulation are given in Appendix H.2.

The simulation searches the events which are scheduled and finds the next event that is due (the first event on the activity list); it then moves the simulation clock forward to this point in time and executes the event. Executing the event will generate other events. For example, a death event will cause a birth event to occur; a person getting infectious TB disease may cause various transmission events; and a TB transmission event will cause an infection or reinfection event. As the simulation executes events therefore, other events are generated and scheduled onto the activity list.

Once the simulation has completed all of the events scheduled within a certain year, various details and disease indicators for that year are recorded before moving onto the following event or year. At the end of each year, for instance, TB incidence and HIV prevalence are calculated and recorded.

6.3 Modelling the Population

From the data available on the age distribution of the population, life expectancy, distribution of household size and ratio of adults to children within households, we are able to define a model population which has the characteristics and captures the appropriate dynamics of the study population in Harare, Zimbabwe. The following Section discusses in detail the data and methods used to generate the population.

The model population is achieved by first generating a household and then creating its occupants. This process is done iteratively until the user-defined population size (N) is reached. In our model we have a population of size 10,000 in which each individual belongs to a defined household, of which there are on average 2500. A discussion of why a population of size 10,000 was chosen is given in Chapter 7, Section 7.6.

6.3.1 Distribution of Household Size

The baseline data from Harare (Chapter 2, Section 2.6.1) contains information on the number of individuals within each household. Using maximum likelihood estimation (Appendix D) we fitted various statistical distributions to the data using Microsoft Excel but incorporated Akaike's information criterion in order to measure their goodness of fit whilst taking into consideration the number of estimated parameters required. This enabled us to establish a distribution which could be used to represent the data on household size in Harare, with a minimum of estimated parameters. Analysis showed that this was accomplished by using the Poisson distribution with mean 3.99, which is given in Figure 6.2.

Each time a household is created, we sample from this Poisson distribution to give us its household size which determines how many occupants to generate and assign to it.

6.3.2 Adult to Child Ratio

Once the number of occupants in a household is established, we use the baseline data from Harare, which gives us information on the number of adults and children within each of the survey households, to generate a distribution of the proportion of the household's occupants that should be children. The data showed that nearly 30% of households were childless and that on average the proportion of children

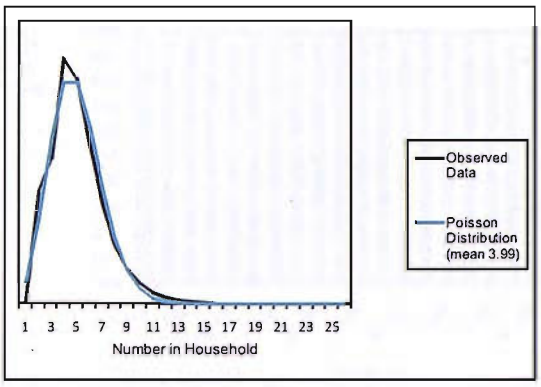


Figure 6.2: A Poisson distribution (mean 3.99) fitted to the Harare baseline data of the distribution of household size

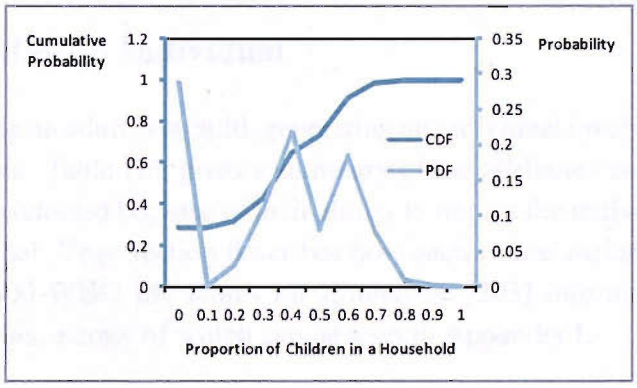


Figure 6.3: The distribution of the proportion of children in households according to the Harare baseline survey data

within a household is 0.34, which indicates that in a typical household with 4 occupants, there would be 1 child. The data also showed that a very small percentage of households (0.01%) were adultless and contained only individuals under 16 years old. The distribution generated from the data is shown in Figure 6.3 and gives us the likely proportion of children within a household.

Sampling from this distribution when generating a household meant that we knew the number of adults and children to assign to the household to make up its occupants.

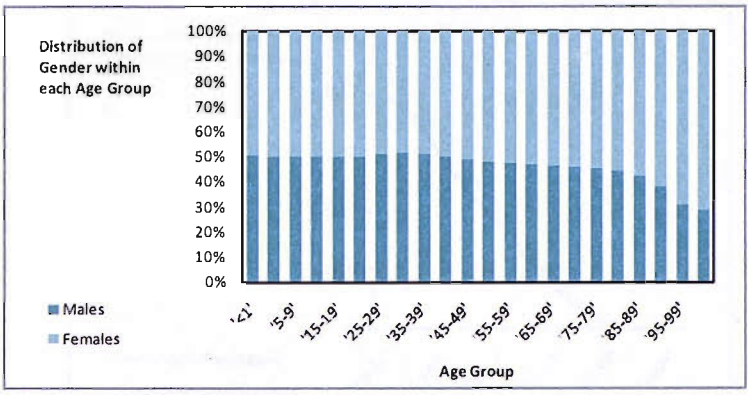


Figure 6.4: The distribution of each gender within each of the population's age groups. Source: 2000 WHO life tables for Zimbabwe [203]

6.3.3 Creating an Individual

Whether creating an adult or a child, generating an individual involves determining various attributes. Table H.3 gives a summary of the attributes an individual carries which were selected because of their ability to impact the individual's pathway through the model. This Section describes how each of the attributes were established. The 2000 WHO life tables for Zimbabwe [203] informed much of the following analysis, a copy of which can be seen in Appendix I.

6.3.3.1 Gender

A person's sex is decided by allowing half the individuals to be male and half to be female. This reflects the fact that when a baby is born there is nearly equal chance that the baby will be a boy or a girl. It also reflects the gender distribution seen in the 2000 WHO life tables for Zimbabwe [203] and shown in Figure 6.4. Figure 6.4 shows that the older a person in the simulation is, the more likely it is that they are a woman. An explanation of how this phenomenon is captured is discussed in the following Section.

6.3.3.2 Age

Once it has been decided whether to create an adult or a child and what gender this person will be, we use the life tables to calculate the likely age of the person. Figure 6.4 shows the distribution of the population amongst the various age groups according to gender. In Figure 6.5 we can see that the likelihood of a person being

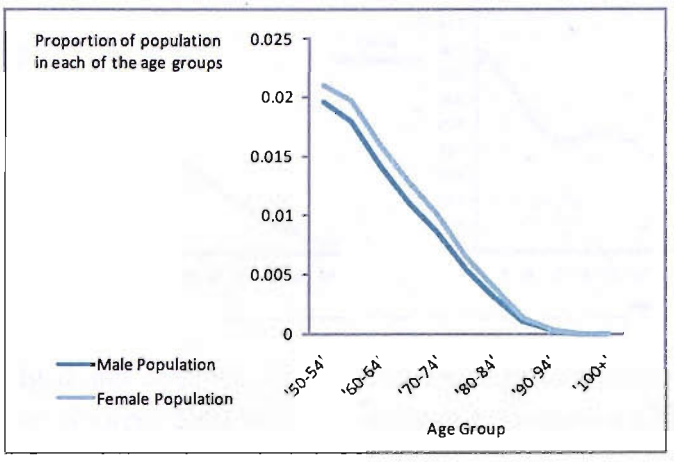


Figure 6.5: The proportion of each gender within each of the population's age groups. Source: 2000 WHO life tables for Zimbabwe [203]

female after age 50 is higher than males and therefore by using these distributions to generate a person's age when initialising the population we ensured this was the case.

6.3.3.3 Time of Natural Death

The time that a person is due to die from natural causes is calculated immediately when a person is created in the model, either when the population is being generated at the start of the simulation or when a person is born once the simulation is running.

When the population is being generated, we know the age of the individual and therefore to calculate the person's time of natural death, we can use the life expectancy data from the life tables. This data is given for both sexes and shows the likely number of years a person has left to live given that they have already survived to a certain age. The life expectancy data is abridged, so we used the Heligman-Pollard method [87] from actuarial science to complete the tables thus enabling us to work with individual ages as opposed to five-year age groups. This made estimates more precise, for example it meant that rather than all males aged between 5 and 9 living for another 38.4 years, they would live for a further 38.7, 37.8, 37.0, 36.0 and 35.1 years respectively. A detailed account of how the Heligman-Pollard method was used to generate complete life tables for both genders is given in Appendix J. Both the abridged and complete life expectancy estimates for males and

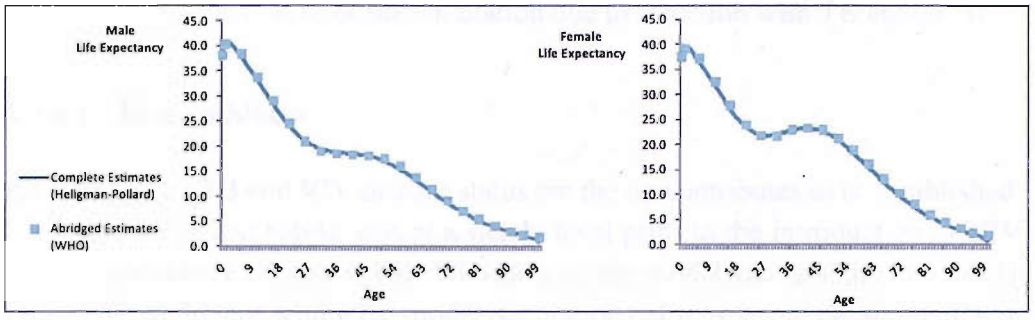


Figure 6.6: Abridged and complete life expectancy estimates for males and females in Zimbabwe. Source: 2000 WHO life tables for Zimbabwe [203]

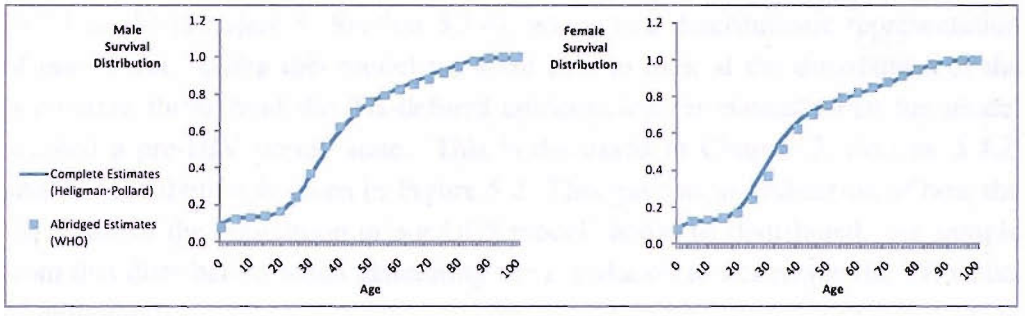


Figure 6.7: Abridged and complete survival distribution for males and females in Zimbabwe. Source: 2000 WHO life tables for Zimbabwe [203]

females in Zimbabwe are shown in Figure 6.6.

When a person is born, their time of natural death is calculated and assigned to the individual as an attribute. In this situation, we use a survival function calculated using the life tables. The abridged life tables for both sexes provide us with the number of deaths in age group x, which we used to determine the probability of death within each age group x. Taking the cumulative probabilities we create a distribution which we can sample from to determine in which age group the baby will die. To be more precise about the age that the baby will die, we again use the completed life tables generated using the Heligman-Pollard method [87] to generate a distribution which can be sampled from to provide an actual age of death, rather than age-group. Both the abridged and complete survival distribution for males and females in Zimbabwe are shown in Figure 6.7.

Each time a baby is created, we sample from the complete survival distributions given in Figure 6.7 to determine, given the baby's gender, at what age this individual will die from natural causes. This person's time of death may be brought

forward during the course of the simulation due to infection with TB and/or HIV.

6.3.3.4 Disease Status

An individual's TB and HIV disease status are the final attributes to be established. TB incidence in Zimbabwe was at a steady level prior to the introduction of HIV into the population (Figure 6.10). We warm up the model and recreate this steady state of TB incidence within the model population before starting the simulation in 1980. As we warm up the model during a time in which HIV is not present, when generating the population, an individual's HIV status is negative. To establish a person's TB status, we use the output from the previously developed SEEINITR DCM model (Chapter 5, Section 5.2.6), which is a deterministic representation of our model. Using this model we were able to look at the distribution of the population throughout the TB-defined epidemiological classes when the model reached a pre-HIV steady state. This is discussed in Chapter 5, Section 5.4.2, and the distribution is given in Figure 5.2. This gave us an indication of how the TB status of the population in our DES model should be distributed. We sample from this distribution when generating the population to determine the TB status of individuals.

Having determined all of a person's attributes, this person can be created and assigned to the current household being assembled. Given that all of the attributes have been ascertained using data from the 2000 WHO life tables for Zimbabwe [203] and from the baseline survey data from Harare (Chapter 2, Section 2.6.1), we have some confidence that the model population correctly captures the characteristics of the study population in Harare. Section 6.3.5 discusses the verification of this.

6.3.4 Births and Deaths

To be in line with the deterministic compartmental models already developed both by us (Chapter 5), and in the literature (Chapter 3), the population size remains stable throughout the simulation. Using a steady environment in epidemiological modelling is useful as it implies a steady influence of the environment on disease dynamics. Thus, when the model is simulated, for instance to investigate the consequences of different interventions for control, we can have confidence that

demographic factors are controlled, and the different outcomes can therefore be explained by purely epidemiological factors [112].

To maintain a stable model population the birth rate in the model is equal to the death rate. To achieve this, each time a person dies, a baby is born into the model's population.

Given that we have a household structure present in the model, the problem arises of which household to assign the new baby to. We want to maintain the distribution of household size in the population as this has been sampled from the Harare baseline data (Section 6.3.1), so the obvious solution is to assign the baby to the household in which the person has just died. This means that a household's size will remain constant throughout the simulation and therefore the distribution of household size is maintained. This method was tried however it raised two concerns. The first was that it created a large number of orphaned or adultless households, which in reality is improbable, as orphaned children would move into care or have family move in with them. The second was that there were a high number of births in households with a high disease burden. This seemed unlikely but more importantly meant that the interaction between TB diseased adults exposing children to infection may not have been captured by the model.

Another method that we tried involved households each having a probability of accepting a baby given their current household size and the household size assigned to them on creation. This method meant that undersized households are more likely to be assigned a baby than households which are oversized. This method was computationally intensive, and ascertaining the appropriate probabilities in order to keep the household structure stable was not possible.

To counteract these issues we assign births in the following way. Given that an individual has just died, its former household has just been reduced in size, meaning that a household of size x has now become size $x - 1$. If we assign the birth to a different household of size $x - 1$, this household now becomes size x. This ensures that the number of households of size x and $x - 1$ remain the same and therefore the distribution of household size is maintained throughout the simulation. So for example, if a person dies in a household of size 4, that household becomes size 3; we therefore assign a baby to a household of size 3, so that it increases its size to 4.

We ensure that a baby is assigned only to a household with adults present and if we are assigning to a house of size zero, then we instruct the simulation to create and assign an adult rather than assigning a birth.

Babies born into the model are assumed to have a "Susceptible" TB status and

a negative HIV status. This assumes that there is no inherited immunity to TB and that there is no mother-to-child transmission of HIV in the model. New adults are also assumed to have a susceptible and negative TB and HIV status because we are not attempting to incorporate the effect of immigration on the disease dynamics of the population. This, however, is an epidemiological complexity that could easily be incorporated into the model by using appropriate data to inform the likely disease status of immigrants.

6.3.5 Validating the Population

Given that the population has been initialised using data from the 2000 WHO life tables for Zimbabwe [203] and from the baseline survey data from Harare (Chapter 2, Section 2.6.1), we can be reasonably confident that the model population correctly captures the attributes and dynamics of the study population in Harare. As a check, we output and investigate various characteristics of the model's population and compare it with the population data for Zimbabwe.

Using the model population, we output the probability distribution of dying at a certain age and considered it alongside the WHO survival distribution given in Figure 6.7. The output displayed in Figure 6.8 is produced by taking 10 runs of the model and examining its population at time point 2000, the year that the WHO survival distribution is taken from. Figure 6.8 compares the average distribution from these runs with the survival distribution and shows that the model, although not exact, is definitely displaying the correct pattern of survival likelihoods. The probability of dying decreases rapidly as a person gets past infancy and then increases again dramatically in the teenage years until the 20s or 30s when it starts to decrease again. This decrease continues into the 50s where it then stays level before dipping after the 80s implying an extremely low probability of surviving to greater than 90 years old. This behaviour is qualitatively similar to that seen in the WHO data for Zimbabwe in 2000, although there seems to be a clear systematic difference from ages 20 to 30. This difference could affect the age distribution of the population and result in a slightly younger model population than is observed in Zimbabwe.

The age distribution of the model's population was also output. We looked at the population once it had been warmed up and during the simulation. When the population is generated at the beginning of the simulation we sample from the age distribution given in the 2000 WHO life tables for Zimbabwe [203] to ensure the model population exhibits the same age distribution (Section 6.3.3.2).

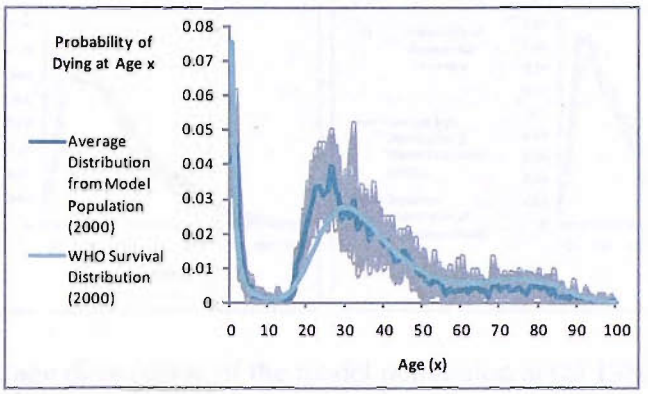


Figure 6.8: The probability that a person will die at age x. The average distribution (blue) from 10 runs (all in grey) of the model's population at year 2000 is compared with the 2000 WHO survival distribution for Zimbabwe

As the model runs however, individuals age, die and are born, and various disease processes act upon them, so we need to ascertain that the age distribution of the model population is still an acceptable representation of the study population in Harare, both once the model has warmed up and during the simulation.

Ideally, as we start the simulation in 1980, when generating the population we would want to sample from an age distribution equivalent to the year 1980, yet the 2000 WHO life table data were the earliest data available. Reassuringly however Figure 6.9a shows that once the model is warmed up, the age distribution of the population exhibits the behaviour we might anticipate for 1980. One of the effects of the HIV epidemic on populations is that it creates a younger population structure, with life expectancies reducing and less people reaching old age. One would therefore expect to see a less positively-skewed age distribution in 1980, than after the HIV epidemic was present. Figure 6.9a shows that the age distribution of the model population in 1980 is less positively-skewed than the 2000 data, meaning that the model population is generally older in 1980 than the Zimbabwean population in 2000. This is an encouraging result and assures us that using 2000 data to initialise the population, but allowing it to warm up for 150 years without HIV present, allows the population time to adjust to a situation more likely for 1980.

We see in Figure 6.9b that the age distribution of the model population in 2000 gives a comparable representation of the Zimbabwean population in 2000. This implies that the processes controlling and maintaining the population within the model are sufficiently accurate and more importantly that we have some confidence that any experiments and conclusions made by the model, are made using a population concurrent to Harare, Zimbabwe.

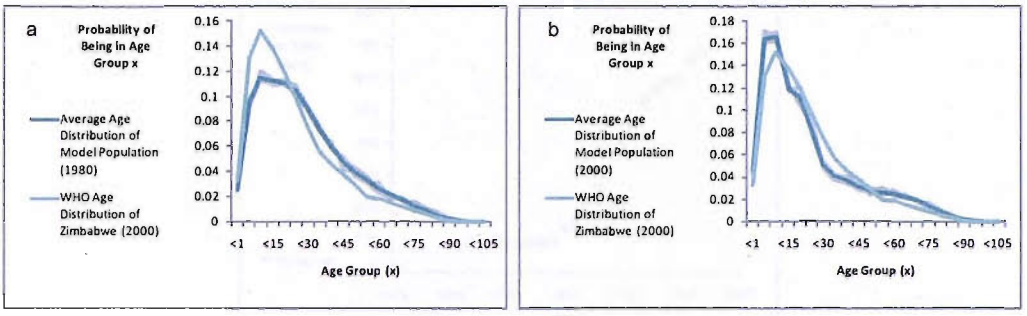


Figure 6.9: The age distribution of the model population at (a) 1980 and (b) 2000. The graphs show the average age distribution (blue) from 10 runs (all in grey) of the model's population once its has been warmed up (1980) and during the simulation (2000). These are compared with the 2000 WHO survival distribution for Zimbabwe

Given the analysis of the model population we can be sure that the methods being used both to generate the population and to progress the population through the simulation give an acceptable result, with the model population accurately reflecting the population in Harare, Zimbabwe.

6.4 Warm Up

The objective of the warm-up process is to obtain the initial population and establish the correct dynamics and disease levels before the start of the simulation. TB incidence data is available for Zimbabwe from 1980 onwards [134] and therefore we want to start the simulation at this time point. Figure 6.10 shows that TB incidence in Zimbabwe is shown to be at a consistently low and steady level prior to the introduction of HIV. This is consistent with the model of TB developed in Chapter 4 which used the observation that even if TB notification rates increase dramatically as a result of the HIV epidemic, the annual risk of infection and the prevalence of TB disease in HIV-negative people may increase only slightly, if at all. This led to us assume that $I^-(t)$, the TB notification rate in HIV-negative people, is constant and equal to $I^-(0)$, the TB incidence observed before the HIV-epidemic began to have an impact on TB. We therefore want to warm the population up to a state where there is no HIV present and the population's TB incidence is steady and consistent with the disease levels implied by the TB incidence data for Zimbabwe.

To establish the length of warm up, initial runs of 200 years were first used to

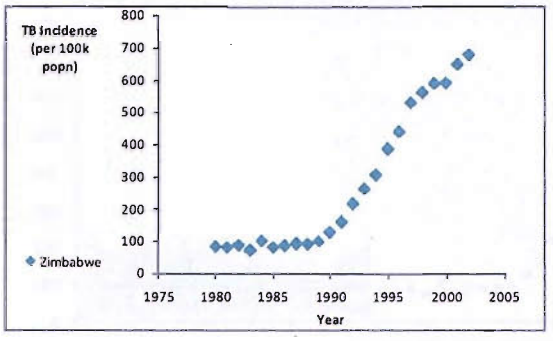


Figure 6.10: TB incidence data for Zimbabwe. The low and steady levels of TB before the HIV epidemic impacted can be seen between 1980 and 1990. Source: 2007 WHO Report [134]

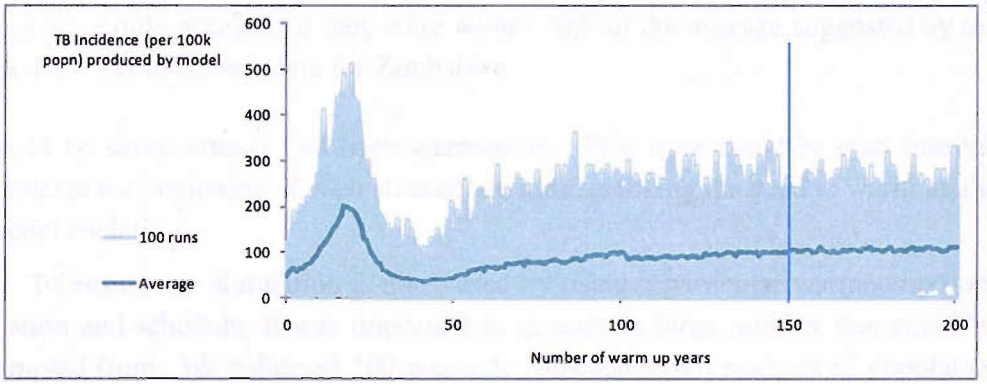


Figure 6.11: Average TB incidence produced by 100 runs of the model over 200 years. HIV is not yet present in the model

observe the TB incidence levels in an HIV-negative population over an extended period. Figure 6.11 shows the average TB incidence per 100,000 members of the population produced by 100 runs of the model. We can see that there is an initial epidemic of TB within the population which dissipates after 40 years. For the next 85 years we see a slow and steady increase of TB until after 125 years of warming up the population, the TB incidence levels off, continuing at that level for the next 75 years. We therefore use a warm up time of 150 years, as this means the population has been in a steady state for 25 years.

We are interested in the behaviour of the epidemic between the years 1980 and 2030 and will only output data for those 50 years. It seems inefficient therefore to run the model for 150 years prior to each simulation run. We developed a "save state" process which meant that the model's state (its population and schedule)

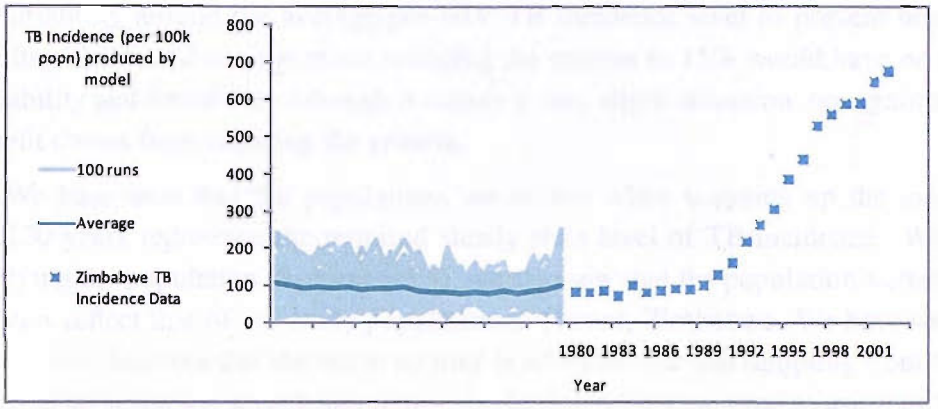


Figure 6.12: Average TB incidence produced in the last 20 years of the 100 accepted warm up runs compared with TB incidence data for Zimbabwe. Warm up runs were only accepted if they were within 30% of the average suggested by the pre-HIV TB incidence data for Zimbabwe

could be saved after it had been warmed up. This state could be read into the model at the beginning of each simulation, thus removing the need to warm up the model each time.

To ensure the simulation is not biased by using a particular warmed up population and schedule, it was important to generate a large number that could be sampled from. We collected 100 warm up runs and saved each set of population and schedules produced by each run. This created an output file with 100 sets of warmed up population and activity lists. This file is read into the model at the beginning of the simulation and each time a run is started, the simulation samples from these sets to decide the particular population and schedule to use.

The average TB incidence produced by the warm ups we collect does relate to the TB incidence data for Zimbabwe, but the variability around the average is currently quite large (Figure 6.11). Although we want a certain amount of variability to ensure there is no bias in the model, saving states with TB incidence over twice that suggested by the data or with no TB present at all would be inefficient as we can be confident that these are not scenarios that were present in 1980. To optimise the scenarios used, we therefore only accept runs if their steady state TB incidence is within 30% of the average suggested by the pre-HIV TB incidence data for Zimbabwe. The results in Figure 6.12 show that the average TB incidence from the scenarios is equal to the pre-HIV TB incidence levels for Zimbabwe, and that the variability is significantly reduced when compared to the variability found in Figure 6.4. We chose 30% because it was felt this allowed a large enough level

of variability around the average pre-HIV TB incidence level to prevent biased results. We looked at what effect reducing the criteria to 15% would have on the variability and found that although it causes a very slight reduction, no significant benefit comes from reducing the criteria.

We have seen that the populations we collect when warming up the model for 150 years reproduce the required steady state level of TB incidence. When verifying the population (Section 6.3.5), we also saw that the population's characteristics reflect that of our study population in Harare, Zimbabwe. We have some confidence therefore that the warm up time is adequate and that sampling from 100 warmed up scenarios at each simulation run both prevents the need to warm up the model each time, whilst reducing bias by keeping the sample size large.

6.5 HIV

Infection with HIV is governed by a static model of HIV, which generates the number of HIV infections to be made each year. The model relates available data on the prevalence of HIV infection to the incidence of HIV infection, a much more difficult quantity to measure.

HIV prevalence data for Zimbabwe is available between 1984 and 2000 [176]. We fit a double logistic equation to this data to describe its behaviour, and obtain complete estimates for HIV prevalence. A double logistic curve was chosen as it is recommended by WHO/UNAIDS [175] and is an established approach to making epidemiological estimates of HIV prevalence in countries with a concentrated epidemic and where there is evidence of a decline in prevalence [111]. A complete description of the double logistic equation and an explanation of the fitting process is given in Appendix K. A double logistic equation is chosen as it allows the initial rate of increase, the peak prevalence, the final steady-state prevalence and the rate of convergence to the steady state, implied by the data to be defined. The double logistic curve obtained can be seen in Figure 6.13 and provides estimates of HIV prevalence for Zimbabwe between 1980 and 2030 which are read into the simulation.

By matching the HIV prevalence within the model with the HIV prevalence estimates, we can ascertain the annual HIV incidence, $I(t)$, as

$$I(t) = (H(t) - N(t) - d(t))(1 + p). \tag{6.1}$$

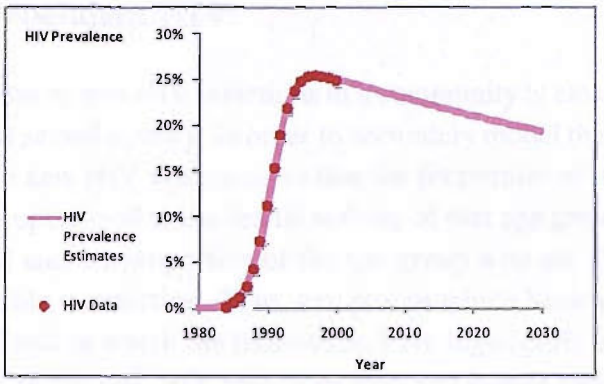


Figure 6.13: HIV prevalence estimates obtained by fitting a double logistic equation to the HIV prevalence data for Zimbabwe. Source: UNAIDS/WHO [176]

This says that the HIV incidence for year t is the number of individuals that need to be HIV-positive in year t in order for the model's HIV prevalence to match the estimates ($H(t)$), minus the current number of HIV-positive individuals in the model ($N(t)$), minus the number of those HIV-positive individuals due to die in year t ($d(t)$). This provides us with the number of individuals that need to be infected with HIV during year t in order for the model's HIV prevalence to match the estimates. We scale this number by the probability that a newly infected person will die within a year, p, to give the final annual HIV incidence, $I(t)$.

HIV incidence is calculated before the beginning of each simulation year. The annual HIV incidence instructs the model how many individuals to infect with HIV throughout that year. This number of HIV transmission events are then generated and scheduled throughout the year according to a uniform distribution.

It is assumed that HIV transmission events can only occur in adults and therefore only individuals over 15 years old are chosen to be infected. It is also assumed that there is no mother-to-child transmission of HIV present, therefore children can not have an HIV infection in the model.

HIV transmission is age-dependent and so a transmission event contains instructions as to which age group the infection should occur. An individual is then chosen randomly from that particular age group to be infected. How age-dependency is incorporated and how the distribution of infection amongst the age groups is determined is discussed further in the following Section.

6.5.1 Age-Dependent HIV

The age distribution of new HIV infections in a community is closely related to the age distribution of sexual activity. In order to accurately model the age dependence of HIV, we assign new HIV infections so that the proportion of new infections in an age group is proportional to the sexual activity of that age group, based on data from the UK [95] and the proportion of the age group who are HIV-negative and therefore susceptible to infection. Thus, age groups which have a high proportion of HIV-negatives and in which the individuals have high levels of sexual activity will receive more of the new HIV infections than age groups with lower numbers of HIV-negatives and lower levels of sexual activity.

The UK sexual behaviour survey [95] contains data on the occasions of heterosexual sex in the past 4 weeks and the number of new sexual partners in the past year, specified according to age group. Such detailed data is not available for Zimbabwe and so we have to make the assumption that the sexual behaviour of British individuals can be used to predict the sexual behaviour of Zimbabweans. This may be an unreasonable assumption however it is made because such detailed sexual behaviour surveys have not been carried out in sub-Saharan Africa. Given that we know the number of times that a person has sex in a 4 week period we know how many contact events there are in a year for one person (a_i, where i represents each age group). We also know how many partners that person is likely to have had in the past year (b_i). We use these figures to work out the relative number of contact events within each age group. Multiplying these figures together gives us a relative contact rate for each age group which reflects each of their risk behaviours, and therefore we can assume that the contact rate c_i is proportional to $a_i b_i$ with the same constant of proportionality (α) applied to each age group. The relative contact rate for each age group i is given by

$$c_i = \alpha a_i b_i. \tag{6.2}$$

We assume that those infected with HIV are as likely to have sex as those who are HIV-negative. Thus, at time t, $p_i(t)$ of the sexually active population in age group i will already be HIV-positive, and so could not be infected. The remaining $(1 - p_i(t))$ are susceptible to infection, where $p_i(t)$ is the proportion of age group i who are HIV-positive.

As infection can only take place within the susceptible population, the contact rate is only effective for a certain proportion of the population. We multiply the contact rate, c_i, by the proportion of the age group that are susceptible, $1 - p_i(t)$, to

give us the effective contact rate amongst susceptibles, $\beta_i(t)$, for each age group, at time t. Therefore the effective contact rate amongst susceptibles is given by

$$\beta_i(t) = (1 - p_i(t))c_i \,. \tag{6.3}$$

Given that we now know the relative effective contact rates in each age group we can work out the proportion of transmissions, $r_i(t)$, that should occur within each age group, at time t. We assume that this is given by dividing the number of effective contacts for an age group by the total number of effective contacts for that year. We assume that this provides us with a distribution describing the spread of transmission events amongst the age groups. Therefore the proportion of transmissions that will occur within each age group is given by

$$r_i(t) = \frac{\beta_i(t)}{\sum_i \beta_i(t)} = \frac{(1 - p_i(t))a_i b_i}{\sum_i ((1 - p_i(t))a_i b_i)}. \tag{6.4}$$

At time t, therefore, the HIV incidence or number of new HIV infections that occur in age group i ($I_i(t)$) is given by

$$I_i(t) = r_i(t)I(t), \tag{6.5}$$

where the overall HIV incidence, $I(t)$, is multiplied by the age group distribution of transmission, $r_i(t)$ and where

$$I(t) = \sum_i I_i(t). \tag{6.6}$$

6.5.2 Validation

In order to justify this approach, we can compare the age-specific HIV prevalence of the model population with data from a similar setting. It has been observed by Williams *et al.* [198] that the shape of the age-specific prevalence of HIV curves remains constant for different scales of the HIV epidemic. It is possible to output from the model the proportion of HIV infections across the age groups and so compare the model output with the age-dependent HIV prevalence data provided by UNAIDS for South Africa [174] to see whether this pattern is observed.

Figure 6.14 shows the distribution of HIV infection across the age groups for the model population in 2005 and the data for South Africa. Comparison with the

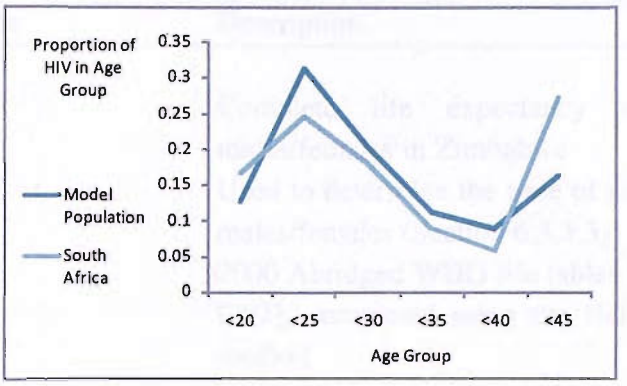


Figure 6.14: Comparison of the distribution of HIV through out the age groups of the model's population and South Africa (Williams 2000 [198])

model data shows that it is slightly overestimating the proportion of HIV infection amongst the 25-30 age group and underestimating amongst the 40+ age group. However, the overall distribution of HIV throughout the age groups is acceptable, especially when considering that the observations of Williams *et al.* [198] came from data from 1998 when the HIV epidemic was still thriving. The authors acknowledged that, as the HIV epidemic saturates and starts to decline (as is the case with the data we are observing in 2005), the conclusions about a stable pattern of infection will be less valid, as the pattern of infection is likely to change.

6.6 Inputs and Outputs

The following Section is a summary of the inputs and outputs of the model and collects together the data sets that have been discussed in this Chapter.

6.6.1 Inputs

The simulation uses various data to inform the model processes. Much of this data and how the simulation uses it to generate the population, for example, has already been discussed in this Chapter. The model reads in the comma separated files and stores them in structures such as vectors which are kept and maintained by the DataFile class (Appendix H.1). Table 6.1 gives a brief description of the input data files, how they are used in the simulation and their source.

Input File Name	Description
MaleLE **FemaleLE**	Complete life expectancy estimates for males/females in Zimbabwe Used to determine the time of natural death of males/females (Section 6.3.3.3)
Source:	2000 Abridged WHO life tables for Zimbabwe [203], completed using the Heligman-Pollard method
MaleSurvivalDistribution **FemaleSurvivalDistribution**	Complete survival distribution estimates for males/females in Zimbabwe Used to determine the time of natural death of a male/female born into the model (Section 6.3.3.3)
Source:	2000 Abridged WHO life tables for Zimbabwe [203], completed using the Heligman-Pollard method
HIVPrevalence	HIV prevalence estimates for Zimbabwe between 1980 and 2030 Used to determine the HIV incidence in each year (Section 6.5)
Source:	Established by fitting a double logistic equation to UNAIDS HIV prevalence data for Zimbabwe [176]
TBIncidence	TB incidence data for Zimbabwe between 1980 and 2002 Used to evaluate the fit of the model's TB incidence
Source:	WHO Global Report [134]

Table 6.1: Simulation Input Files

6.6.2 Outputs

The model provides thorough information on a wide range of simulation details regarding individuals and households in the population, disease statistics and intervention efficacy. This information is output into comma separated files which can be analysed using standard statistical methods to quantify the dynamics of household transmission of TB, to examine the implications of co-infection with HIV and to evaluate a particular intervention, or to provide guidance on the importance of household transmission of TB in preventing the spread of the disease. Data is collected and output for each simulation run. Table 6.2 gives a brief description of the output files generated by the model.

Output File Name	Information Output
Ages	A list of the ages of all the individuals in the population
CasesAvertedX	The total number of TB cases averted each year over the period intervention x was being implemented
CasesFoundX	The total number of TB cases found each year over the period intervention x was being implemented
ClusterCoefficients	The HIV and TB coefficients for each household. A disease coefficient is a value between 0 and 1 which gives a measure of the proportion of the household infected.
HIVAges	A list of the ages of all the HIV infected individuals in the population
HIVIncidence	The yearly incidence of HIV in the population
HIVPrevalence	The yearly prevalence of HIV in the population
HouseholdDetails	For each household, each of its occupants with details of their age, gender, HIV and TB status
HouseholdSize	The size of each of the households

Output File Name	Information Output
HouseholdSizeList	For each household size, a list of the households of this size
InterventionX	The total number of households visited, TB cases found, TB deaths and TB transmissions in the model when intervention x was being implemented
People	For each of the individuals, its ID, age, gender, householdID, HIV and TB status
PreHIVTBLevel	The average TB incidence for the 10 warm up years prior to 1980
PrevInfUnder5	The prevalence of TB infection in children under 5 years olds living in households with and without confirmed cases of TB
SaveState	Two files containing the complete details of the population and schedule at a certain time point
TBEpiStart	The year that the TB epidemic begins
TBIncidence	The yearly incidence of active TB per 100,000 population
TBIncPeak	The peak TB incidence value
TBModelFit	A sum of squares value for the fit of the model's TB incidence to the TB incidence data for Zimbabwe between 1980 and 2002 [134]
TBPrevalence	The yearly prevalence of active TB in the population

Table 6.2: Simulation Output Files

6.7 The Random Number Generator

To generate the random numbers used throughout the simulation to incorporate stochasticity into the model, two random number generators are used. The Mersenne Twister (Matsumoto and Nishimura (1998) [114]) and RANROT (Fog (2001) [40]) are considered by experts to be excellent random number generators and a combination of the two methods was used in our simulation. The RANROT generator is similar to the additive or lagged Fibonacci generators, but with extra rotation or swapping of bits. The Mersenne Twister is a pseudorandom number generating algorithm which considers the flaws of various existing generators, and has a far longer period and far higher order of equidistribution than any other implemented generator. It is also fast and makes efficient use of memory. A combination of the two random number generators was used as this generally performs better than either of the two alone.

Chapter 7

The Model Parameters

In this Section we describe the parameterisation of the discrete event simulation (DES) model designed in Chapter 6, which will be used to evaluate the effects of more intensive case-finding strategies for TB control in a high HIV prevalent setting.

The first stage in developing the DES model of TB transmission in Harare was to design the simulation itself, which was discussed in the previous Chapter. The second stage, which we describe in this Chapter, is to incorporate the various aspects of the natural history of TB into the model and to parameterise it so that it corresponds with earlier deterministic models of TB and HIV. This Section will also discuss how the model was validated using country-wide statistics for Zimbabwe and findings from previous studies of the distribution of TB amongst populations.

7.1 An Overview of the Model Structure

The model of TB has the same structure as the deterministic SEEINITR model discussed in Chapter 5, Section 5.4 but is a discrete event simulation model which allows the population to belong to seven epidemiological classes, dependent on TB status. A schematic of the model's structure can be seen in Figure 7.1. The structure was chosen because it allows the various aspects which are considered important in the epidemiology of TB in HIV prevalent settings, to be incorporated and because it allows for all of the different epidemiological routes experienced by individuals infected with TB to be recognised. A more detailed discussion on why this structure was selected is given in Chapter 5, Section 5.3.

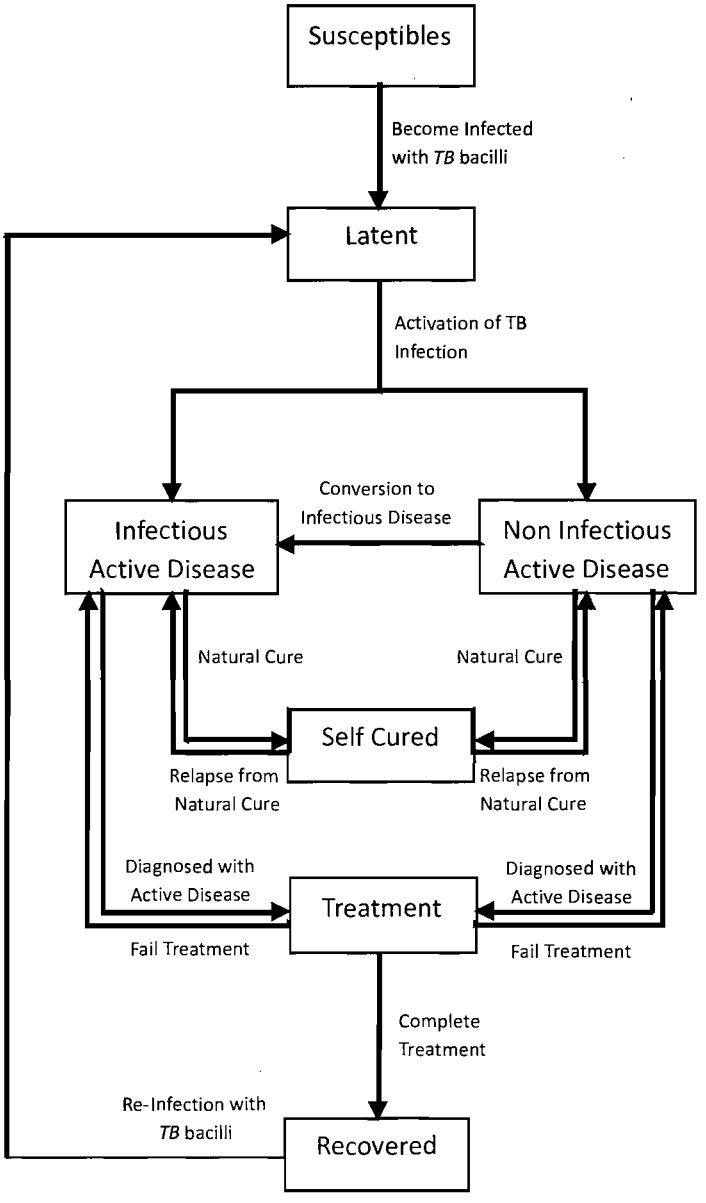


Figure 7.1: Schematic of progression through the TB disease states in the Discrete Event Simulation Tuberculosis model. Death may occur from any state, but death rates are higher from the active disease states.

Movements through the pathways of the model are determined by an individual's attributes (age, gender, HIV status), with the distributions used to describe progression currently based on the literature (Chapters 3), and data from the Harare survey (Chapter 2, Section 2.6.1).

Death can occur from any state of the model, and death rates are higher for individuals with active disease. We assume that the TB disease progression parameters for HIV-positive individuals only change when the individuals enter either early-stage HIV or late-stage HIV. Early-stage HIV or acute HIV refers to the first few months after a patient is infected with HIV. During this time we assume the patient has a period of immunosuppression and will therefore have an increased susceptibility to developing active disease after initial infection with TB (Section 7.2. Late-stage HIV is defined to be the World Health Organization's HIV Disease Stage 3 and above, approximately 6 years after infection [118] [130]. In these stages, individuals are assumed to be more susceptible to reactivation of their TB infection and exhibit higher progression and death rates from TB.

Susceptible individuals are not infected with TB. When they become infected, they enter the latent infection class. Individuals will then follow one of two routes: a) develop active disease within 5 years, termed fast progression to active disease (often referred to as "fast latent"); or b) retain a latent TB infection, only progressing to active disease if their immune system is significantly weakened.

Active TB disease can be infectious or non infectious. When an individual develops active disease they will either be treated, and move to the "Treatment" class, or they will die. For those with infectious TB disease, the time until death, self cure or treatment determines an individual's duration of infectiousness, which in turn determines how many people the individual is likely to infect. Transmission is currently modelled under the assumption that the majority of transmissions will occur outside of the household. When a transmission event is generated, an individual will be selected at random to become infected. If transmission is within household, the individual to be infected will be selected from within the household, and if the transmission is outside the household, any member of the model population could become infected. Persons who have previously recovered from TB, or already have a latent infection, can be reinfected. Persons with active disease or undergoing treatment cannot be reinfected and we assume that no transmission will take place if one of these individuals is chosen. To ensure a transmission event still occurs, different individuals are selected until a transmission is successful. If it is not possible to infect anyone within a household the transmission will be made outside of the household.

An individual will either fail or successfully complete treatment. Those that successfully complete the treatment course enter the "Recovered" class and those that fail will return to active disease. When a person has recovered they are again susceptible to reinfection from an infectious person, although they have an increased immunity compared to the "Susceptible" population.

This brief overview of the model's structure demonstrates how accurately the model captures all of the available pathways and various characteristics of the natural history of TB, which was given in full in Chapter 2, Section 2.3. The rest of this Section will give a description and justification of the parameters and distributions used throughout the model.

7.2 Risk Factors in TB Epidemiology

The main factors affecting an individual's pathway through the model are its household size, its age, its previous TB history and its HIV status. These will determine the likelihood of being infected with TB and if infected, will affect a person's length of latency, the type of active disease developed, their period of infectiousness, the number of transmissions they are likely to generate and their susceptibility to reinfection.

In order to fully understand how an HIV infection affects an individual's disease characteristics, we firstly need to explain how we propose to model an individual's HIV infection.

Patients can be HIV-negative or HIV-positive. HIV-positive individuals can be in the early, middle or late stages of the disease. Early-stage or acute HIV refers to the first few months after a patient is infected with HIV. During this time the patient has transient but sometimes quite profound immunosuppression and we therefore assume that those with acute HIV have an increased susceptibility to developing active disease after initial infection with TB. Because the window of immunosuppression is so brief, there is insufficient time for a previous latent infection to reactivate and therefore we assume that persons with acute HIV have no increased risk of reactivation. The middle stage of HIV refers to the time after recovery from acute HIV until a patient reaches the late-stage (WHO stages 3 and above [130]). We assume that during the middle stage there is no increased risk of reactivation of a TB infection. Patients with late-stage HIV are assumed to be more susceptible to activation of their TB infection and exhibit higher progression and death rates from TB [118].

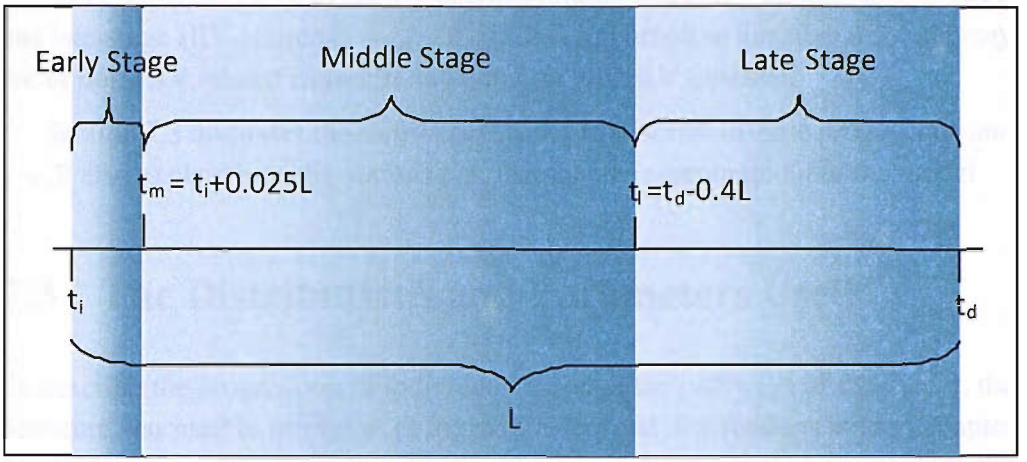


Figure 7.2: How an individual's HIV timeline is generated. t_i, time individual is infected with HIV; L, HIV survival time generated using a Weibull distribution with mean length 10 years; t_d, time individual will die of HIV; t_m, time individual becomes middle-stage HIV; t_l, time individual becomes late-stage HIV

In order to generate an individual's HIV timeline which means to establish the different stages of an individual's HIV infection, we determine the length of infection (L) and then assign each stage according to a proportion of the time they are infected. The process of generating the HIV timeline is illustrated in Figure 7.2. In order to ascertain an individual's length of HIV infection, their survival time is calculated at the point of infection, t_i. This means that when an individual is infected, we sample from a distribution of survival times for HIV infected individuals to give the number of years the individual is likely to survive and therefore the time the individual will die from HIV, t_d. The HIV survival distribution is represented by a Weibull distribution with α and β values of 1.6 and 11.185 respectively; meaning HIV individuals live on average for 10 years. This distribution was used by Salomon in 2001 [146] and has been widely accepted and used by other TB and HIV modellers. The length of time this individual will be infected with HIV is therefore given as the difference between the time of their infection and the time of their death, $L = t_d - t_i$.

Having determined an individual's infection period, we then need to allocate the proportion of time that the individual will be in the various stages. Expert opinion suggests that an individual will be early-stage and late-stage HIV for a period of 3 months and 4 years respectively [118] [130]. As the individual will be HIV-positive for an average of 10 years, this corresponds to being early-stage for the first 2.5% of their infection and late-stage for the last 40% of their infection. We

therefore schedule the individual to become middle stage HIV at time $t_m = 0.025L$ and late-stage HIV at time $t_l = t_d - 0.4L$. It is important to note that a person may die of non-HIV related causes at any point on this HIV timeline.

Section 7.3 discusses the distributions used to describe disease progression and it will also explain how the various risk factors are incorporated into the model.

7.3 The Distributions and Parameters Used

To describe the progression of individuals through the pathways of the model, the literature was used to inform us of the parameters and distributions to use (Chapter 3 and Appendix G). Sampling from various statistical distributions enables the model to capture the stochastic nature of individual disease evolution.

7.3.1 TB Infection

When an individual is infected with TB they enter the latent infection class. The length of time an individual spends in this class depends on their immune response to the infection. An effective immune response means the individual will retain a latent infection, usually for the rest of their lives. An ineffective immune response means the individual will progress to active disease within 5 years, which we call "fast latent". The response the individual will have, depends on their age, HIV status and previous TB history.

Figure 7.3 shows how an individual's attributes affect the likelihood of an individual progressing to active disease within 5 years, p. A p value is established for each individual by asking questions about its age, HIV status and previous TB history. To decide which immune response this individual will have, we generate a random number which if less than the individual's p value indicates an ineffective immune response.

Using parameter values from previous modelling literature we assume that 67% [56] of those with an early- or late-stage HIV infection will progress to active disease, otherwise their p value is determined according to their age.

7.3.1.1 Age Dependency

We incorporate age dependency using the approach developed by Vynnycky and Fine [181]. We calculate the likelihood of an individual progressing to active dis-

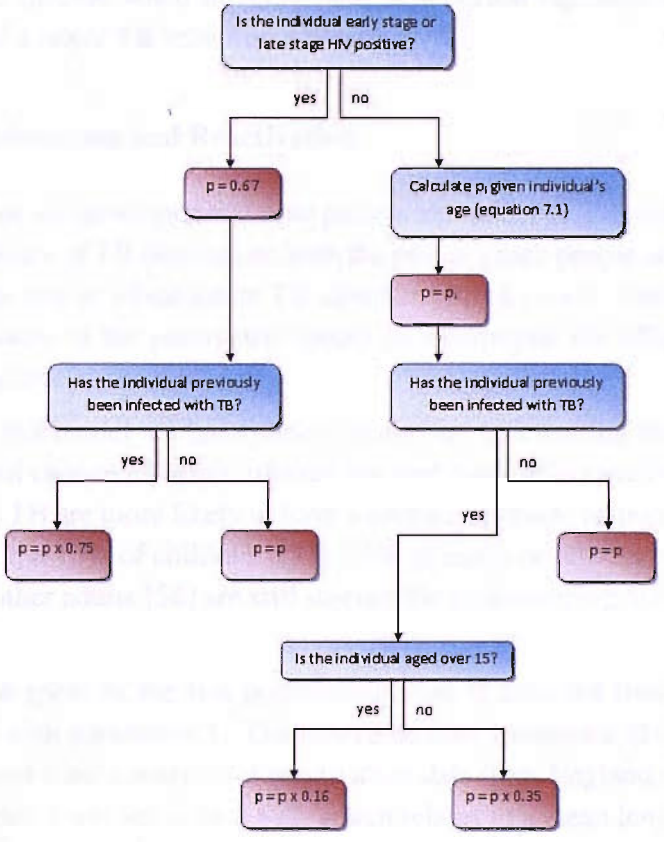


Figure 7.3: Determining an individual's likelihood of progressing to active disease within 5 years, p

ease within 5 years by calculating p_i (an individual's p value at age i). Given that $p_{10} = 0.0406$ and $p_{20} = 0.138$ [181], we work out the individual's p value using equation 7.1.

$$p_i = i\frac{(p_{20} - p_{10})}{10} + 2p_{10} - p_{20} \tag{7.1}$$

Under these assumptions, children under ten years of age are unlikely to develop active disease when infected; and as a person becomes older the risk of activation of a latent TB infection increases.

7.3.1.2 Reinfection and Reactivation

We saw when we developed a simple parametric model of TB and HIV in Chapter 4, that incidence of TB depends on both the rate at which people acquire new infections and the rate at which latent TB infections break down. There were therefore two expressions in the parametric model to incorporate the effect of reinfection and reactivation.

In our DES model we incorporate reinfection by ensuring that all individuals have an equal chance of being infected but that those who have had a previous infection with TB are more likely to have a stronger immune response to reinfection. We assume that 16% of children [181], 75% of early- or late-stage HIV adults and 35% of all other adults [56] are still susceptible to developing active disease when reinfected.

The time spent in the fast progression state is sampled from an exponential distribution with parameter λ. The choice of the exponential distribution follows Vynnycky and Fine's analysis of reactivation data from England and Wales [181]. The parameter λ was set to be 0.872, which relates to a mean length of 1.15 years.

We incorporate reactivation of a latent infection by allowing those that had an effective immune response the opportunity to progress to active disease in the future. The exponential distribution was also used to calculate this time of active disease. In most cases these individuals will retain a latent infection for the rest of their lives and it will fail to reactivate. For this situation the parameter λ was found to be 0.001 [164] [92] [20] [161] [180] [181] and 0.6 for individuals with late-stage HIV (Section 7.4), which corresponds to an average of 1000 years and 1.7 years respectively.

7.3.2 TB Disease

When a person infected with TB progresses to active disease, we determine which type of disease the person is likely to develop. We assume that 73% of individuals with early- or late-stage HIV, and 54% of those without, will develop non infectious disease [197].

Once suffering from active disease we calculate when the individual is likely to die from TB, when they are likely to be diagnosed, when they are likely to self cure, and if non infectious, when they are likely to convert to infectious TB. The times of these events can then be compared to determine which event is most likely to occur first, therefore establishing the time spent in the active disease class and the method of leaving it (by receiving treatment, by self curing etc).

7.3.2.1 Time of TB Disease Induced Death

Once suffering from active disease we calculate when the individual is likely to die from TB using a Weibull distribution with mean survival time of 3.3 and 5.0 years for infectious and non infectious TB respectively, and 3.1 and 2.1 years for individuals with early- or late-stage HIV and infectious or non infectious TB (Table 7.1).

Disease	α	β	Mean (years)	Reference
Infectious TB	2.999	3.696	3.3	[55]
Infectious TB (early/late-stage HIV)	0.298	0.033	0.32	[197]
Non Infectious TB	2.499	5.635	5.0	[55]
Non Infectious TB (early/late-stage HIV)	0.291	0.046	0.49	[197]

Table 7.1: Table showing the parameters of the Weibull distributions used to sample an individual's time of TB disease induced death

7.3.2.2 Time of Treatment

The mean time until an individual receives treatment is 2 years. This value is taken from current TB literature [56] but a suitable distribution to describe the times is unknown due to a lack of data. We use a normal distribution with mean and vari-

ance 2 years, to incorporate some stochasticity into the rate at which individuals are diagnosed and the diagnosis rate we use takes into consideration that not all people will be diagnosed correctly, as diagnosis of TB is quite often difficult and prone to human error. The duration of treatment is deterministic, taking 6 months. This is the minimum duration of a TB treatment regimen as recommended for administration by the WHO.

7.3.2.3 Time of Self Cure

Using parameter values from literature [197], we model the time it takes for an individual to self cure from infectious or non infectious TB using an exponential distribution with mean 2.5 and 5 years respectively. We assume that HIV-positive individuals cannot self cure.

Individuals relapse to active disease at a rate of 0.21 [197]. We model this using an exponential distribution with mean 4.8 years. We assume that if individuals do not relapse within 7 years, they will not relapse [93] and will only progress to active disease if they are reinfected, their infection is reactivated at the usual background rate of 0.001, or their infection is reactivated due to infection with HIV (Section 7.3.1.2).

7.3.2.4 Time of Conversion

If an individual has non infectious TB, there is a risk that the disease will become infectious. The rate at which non infectious disease converts to infectious disease has been estimated by previous modellers to be 0.015 [197], which is equivalent to 67 years. We model this using an exponential distribution.

7.3.3 TB Transmission

As previously discussed, the end goal of the modelling is a discrete event simulation (DES) model of TB transmission in Harare, Zimbabwe which will allow a full assessment of the role of household versus community transmission of TB. Differentiating between these two modes of transmission is important, as household transmission of TB has long been recognised as having very different dynamics from that in the wider community. Household exposure to infectious TB carries a high risk of infection (25 to 50% per year) [143] while random transmission occurs at a lower background rate of approximately 0.5 to 1% per year [62].

As we have seen, when a person becomes infectious, the time that they will leave the active disease class by either death, self curing or getting treatment is generated. This generates the individual's duration of infectiousness. This is then used to determine how many effective transmissions the individual will make.

Individuals will transmit to an average of 10 people per infectious year [56] with the majority of transmissions occurring outside of the household. We determine the number of transmissions a person will make using a Poisson distribution and the length of time the individual is infectious. We allocate the 10 transmission events according to whether they will infect household or casual contacts using a ratio of 9:1. This means that the number of household or random transmission events caused by an infectious person per year are generated with Poisson distributions with means $\lambda_H = 9$ and $\lambda_R = 1$ respectively. Justification of the ratio is discussed in Section 7.4.

Equation 7.2 shows the Poisson distribution used to generate a random variate which defines the number of household transmissions, n_H and the number of random transmissions n_R an individual will generate during their length of infectiousness, d.

$$f(n_i; \lambda_i) = \frac{e^{-\lambda_i d}(\lambda_i d)^{n_i}}{n_i!} \qquad i = H, R \tag{7.2}$$

We assume that the contact rates are the same even when a person has early- or late-stage HIV. Although this is unlikely, the effect of advanced HIV infection on the number of household or random TB transmission events a person might create is not understood and data is not available in order to make sensible estimates.

When the number of transmissions an individual will make during their infectious period has been determined, the times of these events are established using the uniform distribution. We sample from the uniform distribution to distribute the $n_H + n_R$ transmission events throughout the individual's period of infectiousness.

7.4 Determining Unknown Parameters

The previous Section showed how the literature was used to derive the parameters and distributions to use to describe the progression of individuals through the pathways of the model (Chapter 3 and Appendix G). There are however two parameter values that have either not been estimated before, or estimated but less widely accepted or validated. These two parameters are the transmission ratio, r,

which determines how the transmission events are allocated between household or casual contacts, and the HIV reactivation rate, v_{HIV}, which determines the rate at which those with both a late-stage HIV infection and a latent TB infection will progress to active TB disease.

As no previous studies have modelled household versus random transmission of TB using discrete event simulation, there is no estimate present in the literature of the ratio of household to random contact events. Similarly, for the rate at which latent infections in late-stage HIV individuals become TB cases by endogenous reactivation, there is only one study which suggests a value for the parameter. Schulzer *et al.* [153] suggest the rate of endogenous breakdown is 0.17 in late-stage HIV individuals which corresponds to individuals taking a mean of 5.88 years to reactivate. However, using this parameter value in the model meant that the TB incidence was barely affected by the HIV epidemic, which meant that the effect of HIV on the TB epidemic was not amplified enough.

To obtain suitable estimates of both these parameter values we create a matrix to investigate the fit of the model as both parameter values vary.

We assume that the HIV reactivation rate, v_{HIV}, can take the values of 0.0 to 1.0, with 0.0 suggesting late-stage individuals with a latent TB infection are at no increased risk of reactivating, in which case HIV will have no effect on the TB epidemic; and 1.0 suggesting that they reactivate within a year of becoming late-stage HIV, which would indicate the unlikely case that all late-stage individuals with a latent TB infection would progress to active TB disease. We would expect the correct value to lie somewhere in between these two extremes.

We know that individuals will transmit to an average of 10 people per infectious year [56] and we assume that the transmission ratio, r, or the split of random and household transmissions will be either 8:2, 9:1 or 10:0; with 10:0 representing a random contact rate of 10 per infectious person, per year, and a household contact rate of zero. Research in low HIV prevalent settings suggest that transmission events between casual contacts greatly outnumber household transmission events [143] hence the choice of transmission ratios which are biased toward allocating random transmission events.

Applying the range of values for both parameters to the model, we investigate their effect on the fit of the model by comparing the TB incidence output with country-wide statistics for Zimbabwe. We use least squares analysis [1] to provide

[1]This method simply calculates the difference between the model's average predicted values and the values observed in the data. The sum of the residuals squared gives our 'sum of squares' value, which allows the fit of different models to be compared.

r / v_{HIV}	8:2	9:1	10:0
0.0	2085604	1817977	1763994
0.1	1363930	1026468	928116
0.2	894658	559317	543072
0.3	649361	349288	346162
0.4	473650	253986	270717
0.5	397803	250095	257280
0.6	424352	**226252**	304397
0.7	344547	229447	392820
0.8	346427	345545	437681
0.9	300743	279225	410879
1.0	325325	369340	444247

Table 7.2: Table showing the matrix of the sum of squares value the model produces for each pair of possible parameter values for the transmission ratio and the HIV reactivation rate. The combination which produces the best fit is shown in bold

a 'sum of squares' value which gives an indication of how accurate a fit the model is. By exploring all possible pairs of parameter values over the full range for each parameter we generate a matrix of sum of squares values which can be examined to find the combination of parameter values which produce the best fit. The combination of transmission ratio and HIV reactivation rate which minimises the sum of the residuals (the difference between the predicted and observed values) squared thus produces a model which most accurately reproduces the observed TB epidemic.

Table 7.2 and Figure 7.4 show the sum of squares values produced by the model under each pair of possible parameter values for r and v_{HIV}. It shows that using a transmission ratio of 9:1 and an HIV reactivation rate of 0.6 produces the smallest sum of squares value and therefore the most accurately fitting model.

In order to get a more precise estimate of v_{HIV} we effectively zoom in on the minimum of the curve produced by investigating the HIV reactivation rate at a transmission ratio of 9:1. The minimum is found when v_{HIV} = 0.6 and so we investigate the curve more closely by increasing the precision of v_{HIV} and investigate the model as the parameter ranges between the values of 0.55-0.70.

Analysis of the curve produced by investigating the HIV reactivation rate at a transmission ratio of 10:0 shows that the minimum is found when v_{HIV} = 0.5. The minimum of the 10:0 curve produces similar sum of squares values to the

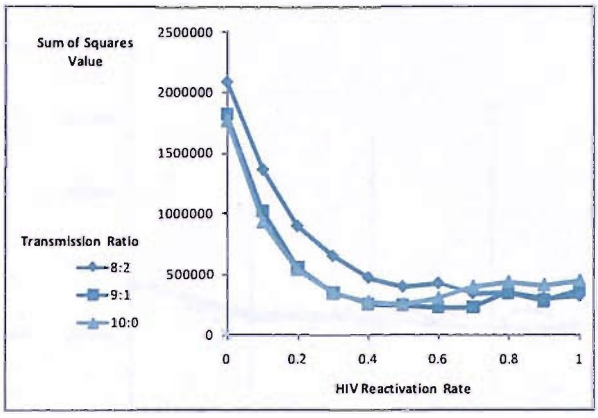


Figure 7.4: Graphical representation of the matrix in Table 7.2 which shows the sum of squares values produced by the model under each pair of possible parameter values for the transmission ratio and the HIV reactivation rate

minimum of the 9:1 curve as can be seen in Figure 7.4. This indicates that the model produced when r = 9:1 and v_{HIV} = 0.6, is similar in accuracy to when r = 10:0 and v_{HIV} = 0.5. In order to ensure the overall minimum doesn't lie on the 10:0 curve, we also investigate the minimum of the 10:0 curve more closely by increasing the precision of v_{HIV} and investigating the model as the parameter ranges between the values of 0.35-0.55.

Figure 7.5 shows the minimum of the 9:1 and 10:0 curves, with the HIV reactivation rate investigated with more precision. We can see that the overall minimum still remains at r = 9:1 and v_{HIV} = 0.6 and therefore we apply these parameter values to the model as they produce a model which most accurately reproduces the observed TB epidemic. The final fit of the model to the TB incidence data for Zimbabwe can be seen in Figure 7.11. Concluding that these are the values the parameters should take implies that the number of transmissions an infectious individual will cause in a year is modelled using a Poisson distribution with mean 9 for random contacts and 1 for household contacts; and that when an individual with a latent TB infection moves into late-stage HIV, they will reactivate within on average 1.67 years. These values have been corroborated as realistic by expert opinion and they also concur with conclusions implied by previous studies that random transmission events will outnumber household transmission events [143], and that individuals with late-stage HIV and a latent TB infection are at a high risk of endogenous reactivation.

Although we have found a combination of the two parameter values that give a model which outperforms models which use any other combination of parameter

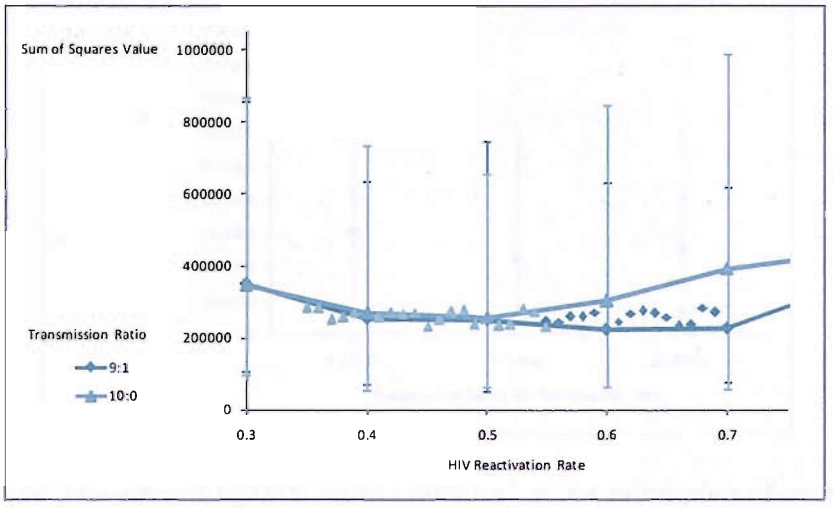


Figure 7.5: Graph showing the minimum of the 9:1 and 10:0 curves from Figure 7.4, with their 90% confidence intervals. The graph also shows the sum of squares values produced by the model under each transmission ratio when using more precise HIV reactivation rate values around the minimums of 0.6 and 0.5 respectively

values, there is some argument to suggest that this could be due to variability. As discussed, and can be seen in Figure 7.5, the minimum of both the 9:1 and 10:0 curves, produce similar model fits. In other words, when the model uses r = 9:1 and v_{HIV} = 0.6, it produces a better but similarly optimal fit to when r = 10:0 and v_{HIV} = 0.45. In order to understand the reliability of our chosen parameter values and to appreciate how variability is likely to be influencing our conclusions, we produced 100 runs of the model at each transmission ratio where the sum of squares was minimised. This means that for each transmission ratio we take the HIV reactivation rate which produced the best fitting model and compare their performance, therefore when r = 8:2, 9:1 and 10:0, v_{HIV} = 0.9, 0.6 and 0.45 respectively. Figure 7.6 shows the resulting average sum of squares value and 90% confidence intervals obtained from 100 runs of the model. We can see that our initial conclusion that r = 9:1 and v_{HIV} = 0.6 still holds as this produced the lowest sum of squares value and therefore the best fitting model, but because of the wide 90% confidence intervals, it would be valuable to assess the effect of setting r = 10:0 and v_{HIV} = 0.45, which produce a close to optimal model. This will be done in Chapter 8, Section 8.4.1, which will look at analysing how sensitive the model is to using alternative values for r and v_{IHV}.

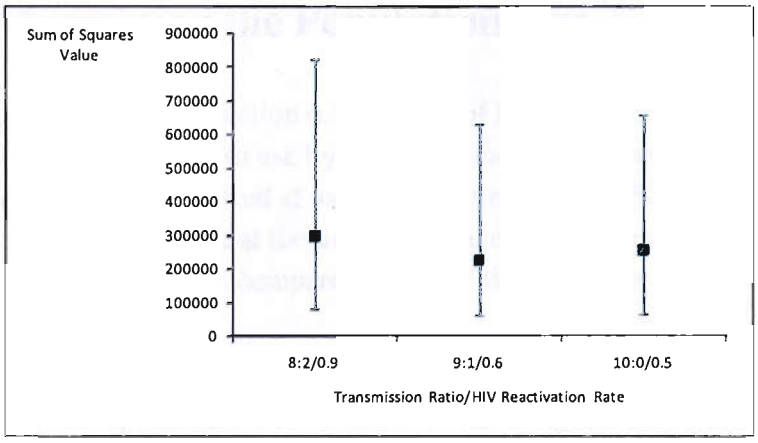


Figure 7.6: The sum of squares values obtained at the mimimum of each of the three curves from Figure 7.4 with their 90% confidence intervals. For each transmission ratio we take the HIV reactivation rate which produced the best fitting model and compare their performance

7.5 Uncertainty Analysis

Given that all of the other parameters have been taken from the Harare baseline data, previous modelling literature and expert opinion, we can have some confidence in their values. We have however, considered the impact of changing some of the more uncertain parameters such as the length of early- and late-stage HIV, the HIV survival rate and the average household size.

Understanding the effect of changing these parameter values gives us insight into how the various aspects influence the model and the TB epidemic it can produce. Appendix L gives full details of the analysis, and documents the conclusions regarding the sensitivity of the model to 3 particular parameters. It concludes that changing the duration of late-stage HIV, and the survival distribution of HIV individuals, affects the timing of the TB epidemic and the amplifying effect of HIV on the average TB epidemic produced by the model. It also concludes that changing the average household size from 3.99, which was implied by the Harare baseline data, to 5.5, as suggested by literature, does not affect the TB epidemic produced.

7.6 Determining the Population Size

As discussed in Chapter 6, Section 6.3, the size of the model population is 10,000. We established the best size to use by analysing the model when different population sizes were used. We looked at using population sizes of 5000, 10000, 15000, 20000 and 25000 and looked at the average TB epidemic produced by the models between 1980 and 2007. We compared the fit of the models to TB incidence data for Zimbabwe.

It was found that regardless of the population size used, the models all produced similar TB epidemics, however there was a significant difference in the variability of the estimates with a reduction in variability seen as the population size got larger. Figure 7.7 shows the fit of the model to the TB incidence data for Zimbabwe as the population size varied. It shows that the 90% confidence intervals around the average TB epidemic produced by the model are wider if the population size is smaller, but that the fit of the model looks unchanged. Figure 7.8a shows the average fit of the model to the TB incidence data for Zimbabwe, with its 90% confidence intervals, by means of a sum of squares value. This is the sum of the residuals (the difference between the model's predicted values and the observed data) squared. Figure 7.8b shows the width of the 90% confidence intervals from Figure 7.8a which illustrates how much variability exists in each model's estimates of the TB epidemic. We can see that the fit is significantly improved and the variability significantly reduced as we move from using a population of size 5000 to size 10000, however, as one moves upwards from 10000, the improvement in the accuracy of the model and the reduction in the variability of the model's estimates is less significant.

As we increase the size of the population there is a pronounced increase in the cost of running the model, as the time taken to obtain 100 runs of the model increases considerably. We have seen that there is clear justification for using a population of size 10000 over one of size 5000 as the improvement in the fit and variability of the model is significant, therefore the increase in cost justified. Figures 7.7 and 7.8 give little justification for using a population of size 20000 or 25000 over 15000 as the fit and variability of the model estimates are only slightly improved, compared to the cost, which would be dramatically increased.

It is unclear whether the estimates of TB incidence and the variability of these estimates produced by a model with a population size of 15000 as opposed to 10000 would warrant the increase in cost. We investigated this further by recording the time taken to collect 100 runs from both a model with a population of size

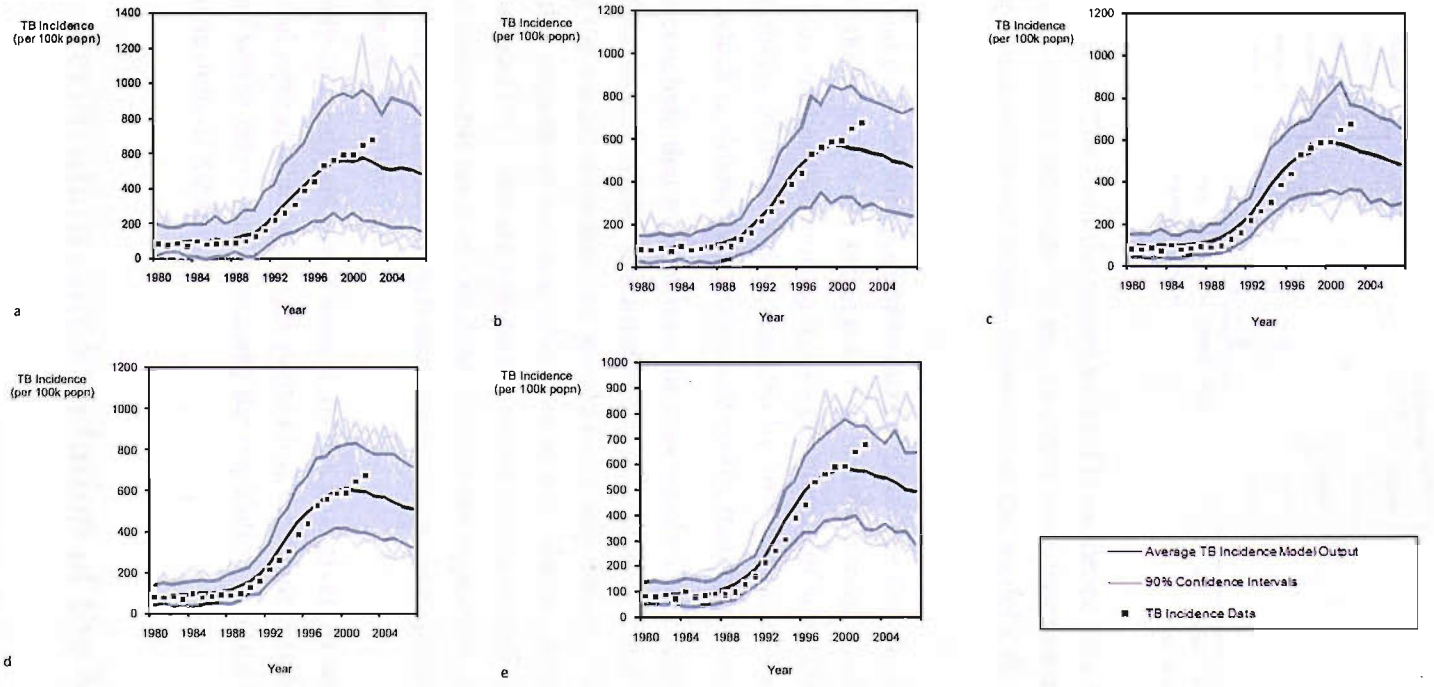


Figure 7.7: Model estimates of TB incidence in Zimbabwe when the model population is of size (a) 5000 (b) 10000 (c) 15000 (d) 20000 and (e) 25000. Each graph shows the observed TB Incidence data from Zimbabwe and the output from 100 runs of the model, along with the average result and 90% confidence intervals for each scenario

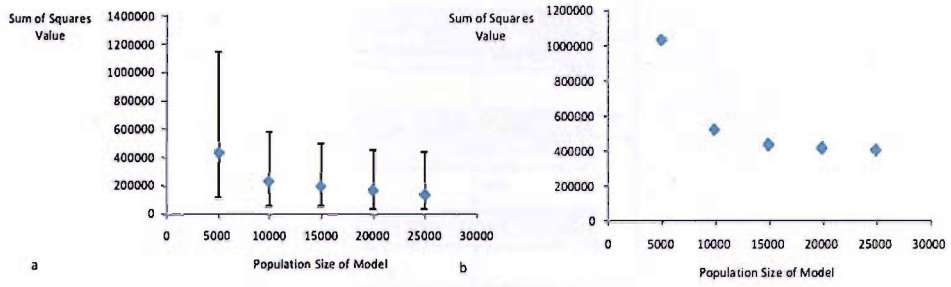


Figure 7.8: (a) The fit of the model to the TB incidence data for Zimbabwe as the population size of the model varies. The 90% confidence intervals are also shown. (b) The width of the confidence intervals on the model's fit, for each population size.

10000 and a model with population size 15000. The first model, with a population of size 10000, took 1491 seconds to run, and the second model took 4274.17 seconds. This implies that moving from a population of size 10000 to 15000 reduces the variability of the model's estimates by 16% but increases the running cost by 287%, which is almost three time more costly, making this move uneconomical.

We conclude that to minimise the cost involved in increasing the population size of the model, whilst also minimising the variability and the loss in the accuracy of the model estimates, we should use a population of size 10000. Interestingly, this population size was also used in a similar stochastic simulation model of influenza [191]. We are satisfied that this sized population produces estimates with an acceptable level of variability, which are an accurate representation of the TB epidemic in Harare, Zimbabwe and which can not be justifiably improved upon given the cost involved.

Analysing different population sizes has also given us some assurance that if the model were applied to a larger population, the results would not be significantly changed which infers that increasing the population to a real size will not change the conclusions of the study.

7.7 Verification and Validation of the Model

We have followed the seven step approach for conducting a successful simulation study as set out by Law and McGomas (2001) [104] and shown in Figure 7.9.

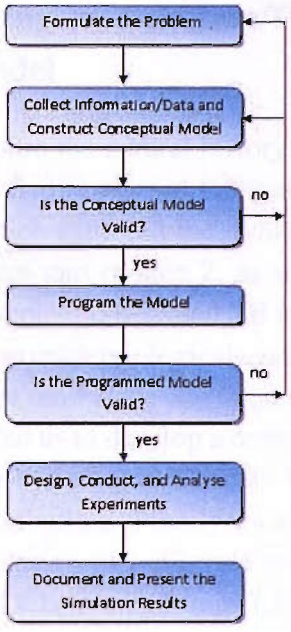


Figure 7.9: A seven-step approach for conducting a successful simulation study. Source: Law and McComas (2001) [104]

7.7.1 Step 1: Formulating The Problem

At the beginning of the project we attended a meeting in Zimbabwe which involved the collaborators, DETECTB, and many subject matter experts such as epidemiologists, mathematical disease modellers and clinicians. The overall objectives of the study were established along with the specific questions to be answered. These were discussed in Chapter 1, Section 1.1. Which type of model to develop was also discussed, as were issues surrounding questionnaire design. This ensured that the data collected would provide maximum assistance in allowing the model to address the specific study objectives. The time frame of the study was also considered and it was agreed that the development of the model and initial conclusions would be addressed by a three year PhD and a further 3 year post-doctorate post would incorporate further complexities such as geospatial aspects, poverty and cost effectiveness analysis. Information on the planned extensions to this study are given in Chapter 10, Section 10.3.

7.7.2 Step 2: Collecting Information/Data and Constructing a Conceptual Model

Chapters 2, 3 and 4 looked into the natural history of TB, the present TB control strategies, previous infectious disease and tuberculosis modelling literature and development of a model which explored the dynamic relationship between HIV and TB. These Chapters form part of step 2, as we were collecting information on TB epidemiology, the interaction between TB and HIV, and the way in which models have represented these epidemiological systems in the past in order to fully understand the 'system' to be modelled. Having a comprehensive appreciation of the processes involved enabled us to develop a deterministic compartmental model in Chapter 5 and to identify literature in Chapter 3 to help us specify model parameters and probability distributions. Finally we were able to create a conceptual model which specified the design of a discrete event simulation model with its assumptions, algorithms and schematic (Figure 7.1) defined.

7.7.3 Step 3: Is the Conceptual Model Valid?

To determine whether the conceptual model was valid, a structured walk-through of the model to audiences including DETECTB and experienced infectious disease modellers was performed and errors in epidemiological understanding and omissions in epidemiological complexities were corrected and updated.

This process of *conceptual model validation* is discussed by Sargent (1991) [149] and involves "determining that the theories and assumptions underlying the conceptual model are correct and that the model representation of the problem entity is reasonable for the intended purpose of the model" [150].

The modelled theory of the natural history of TB was put together after consideration of the literature on infectious disease and TB modelling, and the structure of this simulation model is comparable to many of the previous studies with the schematic being almost identical to a recent study by the WHO which looked at the impact of HIV on the control of TB in India [197]. Additionally, the face validity of the theories and assumptions of the model were confirmed by experts in the field.

It was established that the simulation model described in this thesis simulates the effects of household versus random transmission of TB accurately enough such that the relative effectiveness of household interventions in controlling TB and the effect of different active case-finding methods may be compared.

7.7.4 Step 4: Programming and Verifying the Model

The fourth step of conducting a successful simulation study involves programming the model and carrying out a comprehensive verification. Validation and verification can easily be confused, however validation is ensuring you built the right model and verification is ensuring you built the model right.

As discussed in Chapter 6, Section 6.2, the model was implemented using Object-Orientated Programming techniques with the C++ programming language and the .Net framework. In order to adequately verify the simulation code and ensure it accurately and reliably represents the conceptual design, the following methods were employed.

We used program design and development procedures from software engineering to ensure the correct computer program is derived. In particular we used an object orientated design with structured programming and program modularity. For example, a separate program module and/or object was used for each simulation function (e.g. random number and random variate generators, scheduling transmission events, and implementing intervention routines).

To test the program for correctness and accuracy, firstly the simulation functions were tested. This was done by examining classes as standalone modules outside of the main simulation. Many outputs from the processes within these functions were produced and read into Excel worksheets in order to compare them with expected distributions and results. This was done for each of the functions which establish the time an individual takes to progress through the various epidemiological stages, as well as to test assumptions about mortality, TB transmission, the distribution of household size, the age distribution of the population, and age dependent HIV incidence.

Secondly, we used both static and dynamic testing of the model and its submodels to ensure they were correct. Static testing involved explaining the computer code statement-by-statement to other members of the development team. Dynamic testing involved executing the model under different conditions to determine whether the implementation of the computer program was correct. We ran the code under simplified conditions (for example, with all individuals HIV-positive, or with very low and/or high TB case detection rates) and stopped the code when specific simulation functions were called to ensure the algorithms were correct, the modules interacted correctly and that events were being correctly managed.

Finally, we used tracing throughout the development of the model. This involved stepping through the code one step at a time and noting down the values

assigned and details of the events scheduled to ensure consistency within the code. Tracing also allowed individuals to be followed through the simulation to ensure that their disease progression was correct and happened at appropriately sampled times. It also enabled us to see whether other factors acting upon them, such as their HIV status, were acting correctly.

7.7.5 Step 5: Is the Programmed Model Valid?

Validating a simulation model involves ensuring that it is sufficiently accurate to answer the questions that it has been designed to address [150]. The focus of our simulation study is an investigation of different modes of transmission of TB in a high HIV setting, leading to an evaluation of different strategies for active case finding for TB. Therefore, the most important aspects of the model are the transmission of TB and the effect of HIV on TB. In order to have credibility in the medical area, it is also important that the model is able to reproduce historical data of TB incidence and HIV prevalence.

Operational validity is concerned with whether the outputs of the simulation model have the required accuracy for the purpose of the analysis [150]. We have used graphical comparison of the data from the model and available observed data as an approach to operational validation. Graphs enable us to make a subjective judgment on whether a model possesses sufficient accuracy and also enable subject matter experts such as DETECTB and the team of infectious disease modellers, to make subjective judgments on whether the model is sufficiently accurate for its intended purpose.

We have historical data for the prevalence of HIV infection among pregnant women attending antenatal clinics across Zimbabwe between 1984 and 1999, and the TB incidence per 100,000 members of the population for active TB disease between 1980 and 2002, also for Zimbabwe. More information regarding the data is given in Chapter 2, Section 2.6. Figure 7.10 shows the observed TB incidence and HIV prevalence in Zimbabwe from 1980 onwards. HIV prevalence is increasing in Zimbabwe from 1984 and the number of TB cases shows a corresponding increase after a time lag of about 6 years, the time from initial HIV infection to developing late-stage HIV.

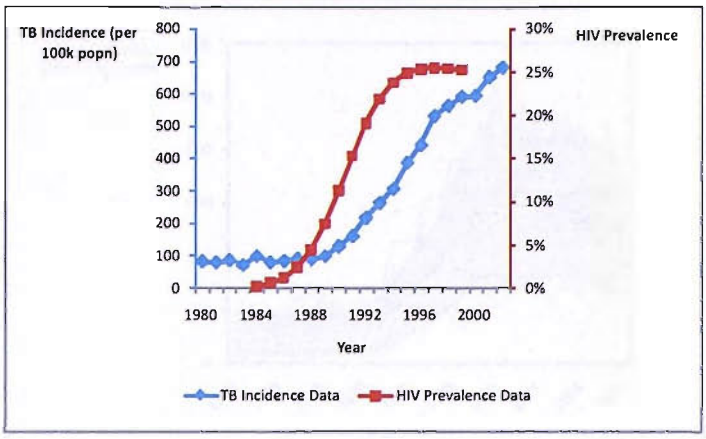


Figure 7.10: TB incidence and HIV prevalence in Zimbabwe. Source: 2007 WHO Report [134] and UNAIDS [176]

7.7.5.1 TB Incidence Model Output

The graphical comparison of the historical data with the model output of TB incidence is given in Figure 7.11. It shows the average model output for TB incidence with 90% confidence intervals, along with the output from 100 runs. It suggests that the model is able to reproduce TB incidence rates in the absence of HIV disease and shows a similar increase in TB cases following the start of the HIV epidemic. This suggests that the model provides a good description of HIV-negative TB and has captured most of the interactions between TB and HIV. However, the model does not produce as sustained an increase in TB cases as the historical data, perhaps suggesting that the model may be underestimating the effect of HIV on TB disease as the HIV epidemic matures.

Data which became available very late in the research on TB case-notification rates in Harare (Appendix M) were not able to clarify whether the TB epidemic is still growing. Our model predictions suggest that the TB epidemic is slowing, which is what we would expect given that HIV prevalence is now declining, and is in line with the opinion of experts [50]. Furthermore, despite experimentation with the model's various parameters it was still not possible for the model to create a TB epidemic which both imitated the observed TB incidence between 1980 and 2000 and continued to grow after 2001. We can be less concerned therefore, about the model's underestimation of TB disease in 2001 and 2002 as these two data points suggest that the TB epidemic is still increasing dramatically which seems unlikely [50]. However, we still address this possibility by investigating the implications of introducing interventions into a growing epidemic as opposed to one which is

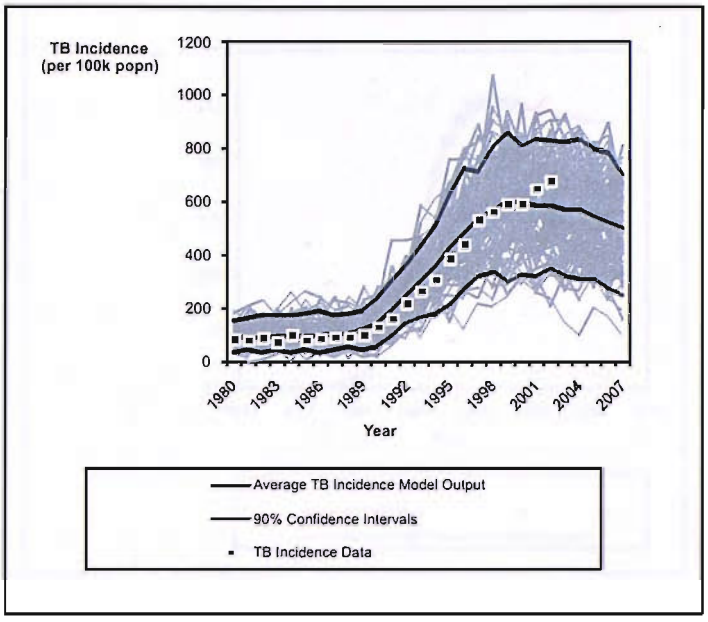


Figure 7.11: Model estimates of TB incidence in Zimbabwe. The graph shows the observed TB incidence data from Zimbabwe and the output from 100 runs of the model, along with the average result and 90% confidence intervals

maturing. This analysis is given in the sensitivity analysis in Chapter 8, Section 8.4.3.

7.7.5.2 TB Transmission Dynamics

Validating the modelling of TB transmission is more difficult as there are few data available. Most important to our investigation is having the correct balance between transmission occurring within the household and that occurring within the wider community. Due to the lack of data available on the proportion of transmissions in these different settings, we used findings from an old study in India [124] which investigated the distribution of persons infected with TB in households with and without confirmed cases. The study found that in children aged under 5 years, there was a prevalence of infection of 12% in households with bacteriologically confirmed cases as compared to 2% in households with no confirmed cases, giving an "infection intensity" (i.e. the ratio of percentage of contacts aged under 5 years infected among members of households without a case to the percentage of contacts in the same group infected among members of households with a case) [88] of 6. The average output from 100 runs of the model suggested that the "infection

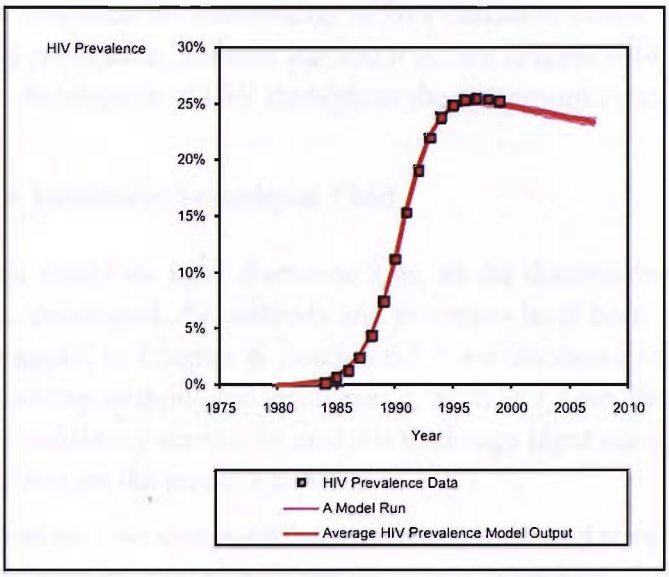


Figure 7.12: Model estimates of HIV prevalence in Zimbabwe. The graph shows the observed HIV prevalence data from Zimbabwe and the output from 100 runs of the model, along with the average result. Source of HIV Data: UNAIDS/WHO [176]

intensity" within children under 5 years in our model population was 5.2. This implies that transmission dynamics are being adequately captured by the model however, having further data on the balance between household and random transmission would allow a more rigorous validation.

7.7.5.3 The HIV Model

The graphical comparison of the historical data with the model output of HIV prevalence is given in Figure 7.12. It shows the average model output for HIV prevalence, along with the output from 100 runs, and the HIV prevalence data for Zimbabwe [176]. The output suggests that the model is accurately representing the HIV epidemic with little variability between runs.

In Chapter 6, Section 6.5 we discussed the design of the simple age-specific HIV model and how it was implemented into the main simulation model. This particular aspect of the modelling could be questioned and therefore we used other data to validate it and increase its credibility in the medical area. To justify the approach, we compared the age-specific HIV prevalence of the model population with data from a similar setting. Chapter 6, Section 6.5.2 discusses the validation

process which compared the distribution of HIV infection across the age groups from the model population and data for South Africa (Figure 6.14), it concluded that the overall distribution of HIV throughout the age groups is acceptable.

7.7.5.4 Other Validation Techniques Used

Throughout this thesis we have discussed how, as the discrete event simulation model has been developed, the methods and processes have been constantly validated. For example, in Chapter 6, Section 6.3.5 we discussed validation of the population modelling methods and in Chapter 7, Section 7.5 we discussed how we used parameter variability-sensitivity analysis to change input parameter values to determine the effect on the model's behaviour.

In this Section we have discussed the various methods used to verify the model's code and the techniques used to validate the output behaviour of the model. We have explored the operational validity of the model which has allowed experts to determine that the model is sufficiently accurate to enable the effect of different active case-finding methods to be compared. The model almost perfectly imitates the observed TB epidemic with the TB incidence produced by the model accurately capturing both the pre-HIV situation and the appropriately timed and amplified impact of HIV. Comparison with a household survey in India has allowed us to ascertain that the transmission dynamics are being accurately captured and finally, we have seen that the simple age dependent HIV model implemented, allows the model to reproduce the correct HIV prevalence levels with infection appropriately distributed throughout the age groups.

Although we have mostly relied upon validation techniques such as face validity and historical data validation, we have also applied other methods such as those described by Sargent (1998) [150] as degenerate and extreme condition testing, and operational graphics. We experimented with parameter values (for example, a zero TB case detection rate) to test the degeneracy of the model's behaviour and similarly applied unlikely combinations of factors into the model to test that the outcomes were plausible. We also displayed performance measures such as TB incidence and the number of individuals in each epidemiological state, as the model progressed, which enabled us to study the dynamic behaviour of the model and its performance measures as it moved through time.

Given that we have conducted a process of verification and validation we now have some confidence in the theories and assumptions of the model. We know that the simulation is an acceptable representation of the 'system' being modelled and

that the model's output behaviour is sufficiently accurate to allow it to address the questions being raised in the study. We are now in a position to use the model to investigate various intervention scenarios and therefore we need to design and execute various simulation experiments (step 6) and collect and assimilate the results (step 7).

7.7.6 Steps 6-7: Designing, Making and Analysing Simulation Experiments and Presenting the Results

This Chapter has described a discrete event simulation model of TB and HIV that will be used to assess the effectiveness of different control strategies for TB. The provisional results output by the model are encouraging and show that the model is accurately reproducing the HIV and TB epidemics and the interactions between them. Experimentation within the model will focus on determining strategies for reducing TB transmission by targeting case-finding for TB disease. The following Chapter (Chapter 8) will discuss how the simulation experiments were designed and applied and will give a thorough analysis of the results. Chapter 9 will investigate further model scenarios and in both cases the results will be clearly shown using graphical methods. The final step in conducting a successful simulation study, step 7, involves documenting and presenting the simulation model results, which is the purpose of this final PhD thesis which documents the model, describes the computer program, and discusses the study results along with the validation process to promote model credibility.

Chapter 8

Results

We have described the design and development of a discrete event simulation model of TB and HIV that will be used to assess the effectiveness of different control strategies for TB. We have implemented a simple method for transmission, differentiating between transmission inside the household and transmission in the general community. This will allow us to determine the significance of household versus community transmission and to analyse interventions targeted at household contacts of TB patients.

In this Chapter we describe the design and the results of the simulation experiments aimed to investigate the effect of various intervention scenarios. This Section will also discuss a sensitivity analysis of the output and conclusions of the simulation experiments.

8.1 Experimental Design

The discrete event simulation model of TB transmission in Harare, Zimbabwe allows us to consider various TB case-finding strategies and examine their impact. A few carefully selected case-finding scenarios were chosen as these represented extensions to the current TB control methods which allowed the relative effectiveness of household interventions to be compared.

The current policy of TB control in Zimbabwe is that of passive case-finding, where the rate at which individuals present themselves with TB is captured in the model by the diagnosis rate (see Chapter 7, Section 7.3.2.2). The intervention scenarios that we define assume that this background diagnosis rate continues, in other words the active case-finding policies applied to the model are seen as a

supplement to the passive case-finding methods already in place in the population. We assume that no improvements are made to the current passive case-finding control policy in order to create a baseline simulation scenario to which all other policies can be compared.

We are interested in the impact of different strategies for active case-finding on the magnitude of the TB epidemic, where this is measured by the number of TB cases found and the number of TB deaths and TB cases averted between the start of the interventions in 2008, and 2028. The active case-finding strategies we consider are:

1. Investigating each member of a household in which there is a newly detected TB case or a death from TB, to see if there are any other undetected TB cases or members with a TB infection in the household
2. Investigating all members of the same number of households, where the households are selected at random
3. Investigating all members of the same number of households, where the households contain at least one HIV-positive person
4. Investigating each member of the household of all persons entering late-stage HIV to see if there are any undetected TB cases or members with a TB infection in the household (we assume that it is around this time that an HIV-positive individual will approach the health services)
5. Investigating the same number of random households as in strategy 4

Strategies 1, 2 and 3 can be compared directly with each other as they require the same level of effort in terms of the number of households that need to be visited (Table 8.1 gives details of the number of households visited as part of each intervention). Strategies 4 and 5 can be compared with each other and will provide an insight into whether TB disease is clustered in HIV-positive households. We assume that interventions 2 and 5, in which random households are visited, can be used to approximate an untargeted or community-wide intervention, as these interventions involve individuals being randomly selected from the population to be tested.

We also consider a sixth strategy which investigates the effect of doubling the rate at which people with TB present for treatment. This strategy is not an active case-finding strategy and relies on patients coming forward for treatment earlier.

We have no suggested mechanism for doing this, however it acts as a comparison to the other scenarios and shows the effect of doubling the diagnosis rate.

With the exception of the 6th strategy, the interventions involve a particular household being visited either because a member has presented themselves for TB treatment, a member has died from TB, a member is HIV-positive, a member is late-stage HIV or the house has been selected at random. Visiting the household involves testing each member for TB infection and disease.

Detection is assumed to be done using tuberculin skin testing (TST) and sputum microscopy. If an individual is found to have TB disease they will receive treatment and if found to have a TB infection, they will receive isoniazid preventive therapy (IPT). Preventive therapy is given to a person with a latent TB infection to prevent it developing into active disease and only provides protection against progression whilst the individual is on therapy. On the recommendation of medical professionals in CREATE and TB epidemiology experts belonging to the WHO [52], we assume that IPT is administered for the standard 6 months and is only effective in 50% of cases to account for the incomplete protection that has been observed in trials, and the combined effects of isoniazid resistant TB strains and imperfect compliance.

We run 100 iterations of the simulation model to compare the strategies, finding the difference between the average number of TB deaths averted between the base case and the six strategies for each iteration. Other performance measures recorded include the average number of TB cases found and the average number of cases prevented through IPT. These performance measures are collected as a time series in order to see how efficacy changes over the course of the intervention. Finally, we output the TB incidence expected under each scenario in order to understand the impact of the interventions on the TB epidemic.

8.1.1 Variance Reduction

To minimise the variability, we use the same start point in 2008 for each scenario. This means that at the start of the trial period the simulation will have the same set of events scheduled and identical individuals with identical household structures for each strategy.

We use the same "save state" method developed in Chapter 6, Section 6.4 in order to save each iteration's population and schedule set at the end of 2007. These sets are then read in systematically to form a starting point for each iteration of each trial enabling each trial to be directly compared.

8.2 Results

Figure 8.1 shows the model estimates of the TB incidence in Zimbabwe expected from 2008 onwards under the baseline and intervention scenarios. It shows that strategies 1, 2 and 3 have a relatively small effect on the TB epidemic which is due to the small number of houses that are visited as part of this intervention. This is in line with conclusions from a study in Nepal which found contact investigations can be an inefficient activity [189]. In this respect, strategies 4 and 5 perform much better. Table 8.1 shows the number of households visited per 10,000 households as part of each 20 year active case-finding intervention. It shows that 2972 households are visited over 20 years in interventions 1-3 but that over 6.5 times more households are visited through interventions 4 and 5. Despite strategies 4 and 5 performing better than interventions 1-3, more than six times more households must be visited to complete these strategies and they would be much more costly.

The relative benefits of employing strategy 1 (visiting households in which a person has been diagnosed with TB disease) versus strategies 2 (visiting the same number of random households) and 3 (visiting the same number of households with at least one HIV-positive) depend on the level of clustering of TB cases within households and how this is affected by the presence of an HIV-positive member.

Figures 8.2, 8.3 and 8.4 show that visiting households which contain at least one HIV-positive member mean more TB cases are found and more TB cases and deaths are averted than visiting the same number of TB-diseased or random households. The results also suggest that not targeting TB-diseased or HIV-infected households and just visiting households randomly would be the least effective strategy.

The analysis shows that although none of the first 3 interventions significantly impact the TB epidemic, strategy 3 (visiting households which contain at least one HIV-positive member) is the most effective intervention for finding cases, suggesting that there is some clustering of TB cases and HIV infection. The probability of a household with a member about to commence TB treatment having a second person suffering from active TB disease is 0.010, whilst for a household with at least one HIV-positive member, the probability is 0.013 and for a general household the probability is 0.008. This explains why strategy 3 appears to perform better than the other two strategies, although the variability in the results is very high because of the small number of households being visited.

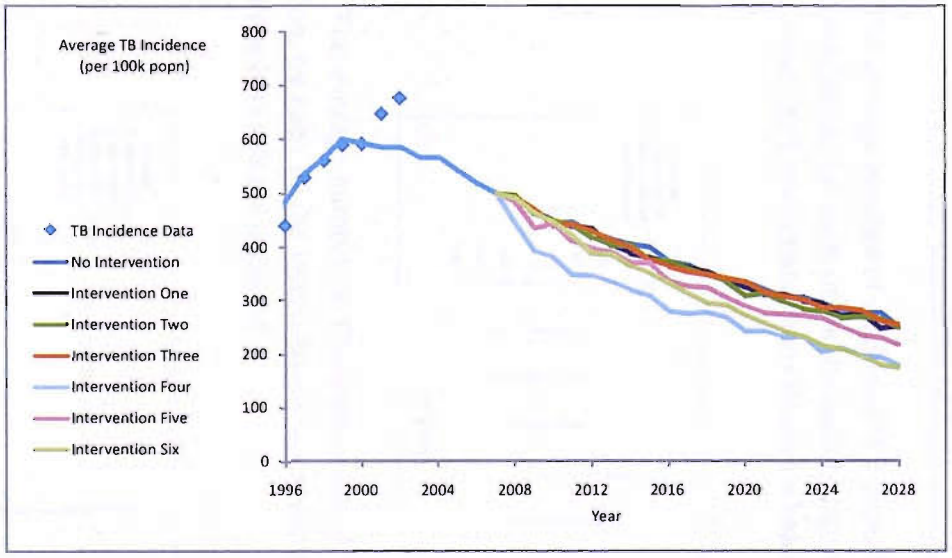


Figure 8.1: Model estimates of the TB incidence in Zimbabwe under six intervention scenarios. The graph shows the observed TB incidence data from Zimbabwe and the average TB incidence produced by the model from 100 runs of the baseline scenario and 100 runs of each intervention scenario when implemented in 2008

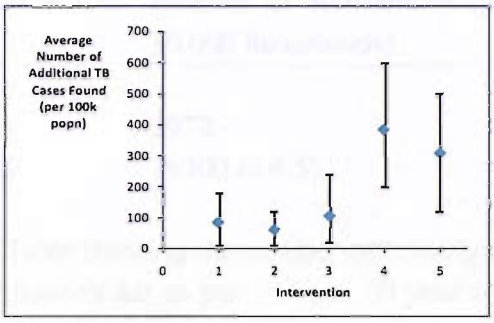


Figure 8.2: The average number of additional TB cases found, per 100,000 members of the population, by each of the household interventions when compared with the base case; 90% confidence intervals are included

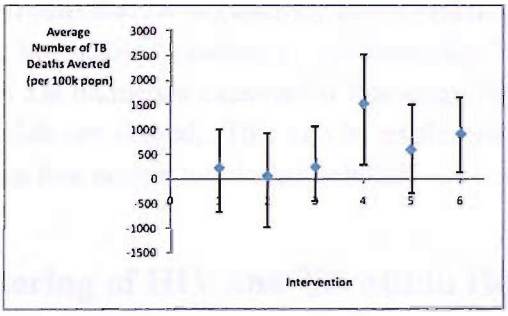


Figure 8.3: The average number of TB deaths averted, per 100,000 members of the population, by each of the interventions when compared with the base case; 90% confidence intervals are included

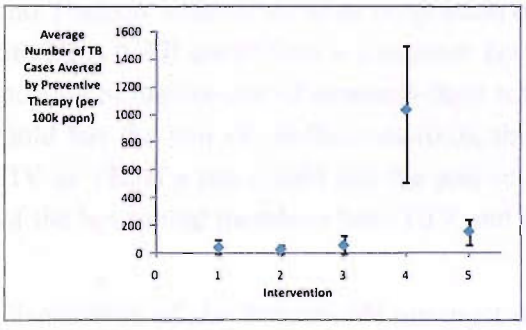


Figure 8.4: The average number of TB cases averted, per 100,000 members of the population, by using IPT in each of the household interventions when compared with the base case; 90% confidence intervals are included

Intervention	Number of Households (per 10,000 households)
1-3	2972
4-5	19300 (x 6.5)

Table 8.1: Table showing the number of households visited per 10,000 households as part of each 20 year active case-finding intervention

The relative benefit of employing strategy 4 over strategy 5 is clear. Figures 8.2 and 8.3 show that visiting late-stage HIV households (households with a member with late-stage HIV) means that more cases are found and more TB deaths averted compared to visiting households randomly. Additionally, Figure 8.1 shows the dramatic reduction in TB incidence expected if late-stage HIV households rather than random households are visited. This can be explained by the clustering of HIV and TB infections that occurs within households.

8.2.1 The Clustering of HIV and TB within Households

We have examined the model for the clustering of HIV and TB which has shown that the large majority of households containing one or more TB cases, contain HIV-infected individuals.

We assign each household an HIV and TB coefficient. An HIV coefficient is a number between 0 and 1 which informs us what proportion of the household are currently infected with HIV. A TB coefficient is a number between 0 and 1 which informs us what proportion of the household currently have active TB disease. For example, if a household has the pair of coefficients (0,0), this implies no-one in the household has HIV or TB; if a household has the pair of coefficients (1,0.5), this implies that all of the household members have HIV and half of the members have active TB.

We look at the distribution of the households amongst each of the possible combinations of coefficients and concentrate on only those with non-zero TB coefficients. Figure 8.5 shows the extent of HIV infection within TB diseased households. The diagram shows the disease coefficient of each household that has TB. If there was a perfect correlation between TB and HIV in which everyone with TB

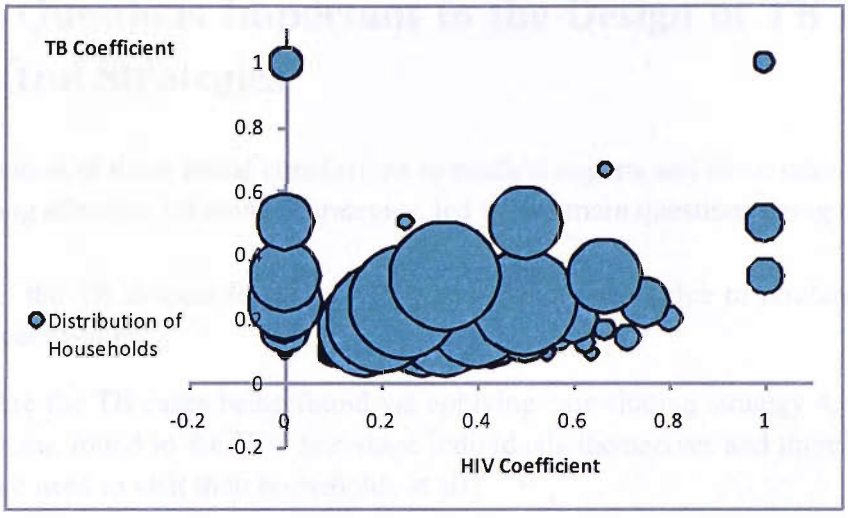


Figure 8.5: The TB coefficient (proportion of the household with TB disease) and the HIV coefficient (proportion of the household infected with HIV) of each household. The diameter of each circle represents the proportion of households within the model, with that particular combination of coefficients. Only households with non-zero TB coefficients are shown

has HIV and vice versa, you would expect all the points to lie on a straight positive diagonal line. The size of each point shows the proportion of households with that particular set of disease coefficients, for example the largest point is at (0.25,0.25) which means that the majority of TB diseased households have a quarter of their household infected with HIV, whilst a quarter of its members have active TB. The diagram shows that the majority of households have 10 to 40% of its members with active TB disease and 15 to 55% of its members HIV positive. Most significantly, the diagram shows that the majority (86%) of households containing a TB case, contain HIV-infected individuals. This relationship between the number of HIV and TB infections within a household suggests that if TB is present in a household it is likely that HIV is too. It therefore follows that in order to find TB cases, it is more efficient to look in households with HIV infections than visiting households randomly.

8.3 Questions Important to the Design of TB Control Strategies

Presentation of these initial conclusions to medical experts and those interested in designing effective TB control strategies, led to two main questions being raised.

1. Is the TB disease found in HIV-positive individuals due to reinfection or reactivation?
2. Are the TB cases being found via applying case-finding strategy 4, mostly being found in the HIV late-stage individuals themselves and therefore do we need to visit their households at all?

It was felt that addressing these questions would both contribute to our understanding of HIV-related TB and further explore the efficiency of the intervention which the model suggests would be most effective in reducing TB incidence in Zimbabwe; that of visiting households with individuals entering late-stage HIV.

8.3.1 Is the TB disease found in HIV-positive individuals due to reinfection or reactivation?

The epidemic of HIV has led to a dramatic resurgence of TB and although it is accepted that the increase in TB incidence is due to HIV-associated TB, it is not clear what the relative contribution of reinfection or reactivation is. Using the model we were able to output whether an HIV-positive individual gets active TB disease though primary infection, reinfection or reactivation.

Figure 8.6 shows the percentage of TB disease in HIV-positive individuals attributable to either primary infection, reinfection or reactivation for 40 years after the beginning of the HIV associated TB epidemic. It shows that on average 87% of HIV-associated TB is due to endogenous reactivation of a latent infection, with only 2% being attributable to reinfection and 11% to primary infection. This implies that the majority of HIV-associated TB is due to reactivation rather than the HIV-positive individual's inability to evoke an effective immune response to a new TB infection.

This insightful result helps to explain why administering IPT to HIV-positive households (households with HIV-positive members) in interventions 3 and 4, means more cases are averted when compared to visiting the same number of

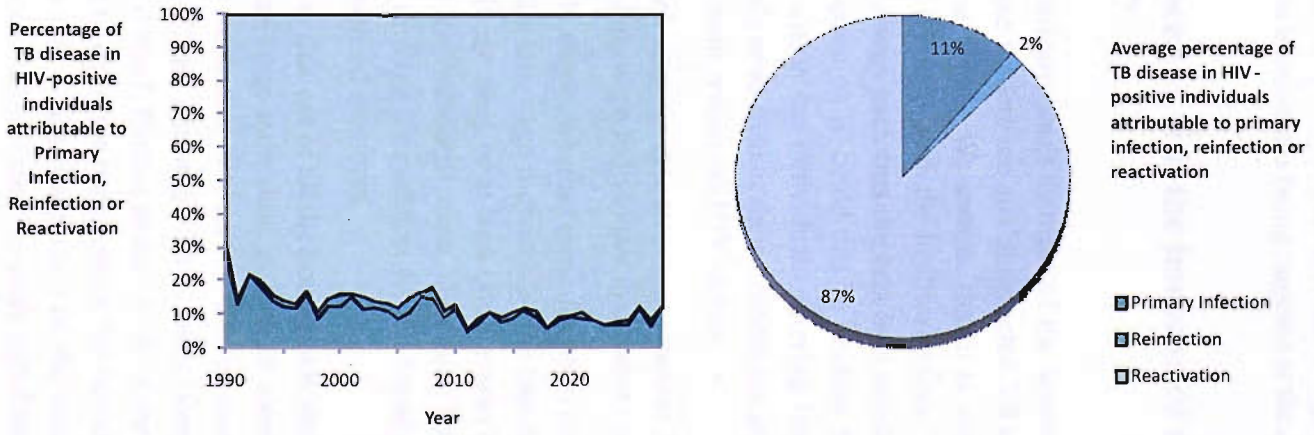


Figure 8.6: The percentage of TB disease in HIV-positive individuals attributable to either primary infection, reinfection or reactivation, over time. The pie chart gives the average result over the 40 year period

random or TB infected households. These interventions ensure HIV-positive individuals, who are at an increased risk of a latent infection reactivating, receive IPT and therefore more cases are averted than if IPT were administered to other individuals. Visiting HIV-positive households has a larger effect on the TB epidemic (Figure 8.1) because treatment is being targeted at the cause of HIV-associated TB.

8.3.2 Do we need to visit the households of late-stage HIV individuals?

Intervention 4 investigates each member of the household of all persons entering late-stage HIV to see if there are any undetected TB cases or members with a TB infection in the household. We assume that it is around this time that an HIV-positive individual will approach the health services. The results have shown that this case-finding strategy performs the best as it averts the most TB cases and TB death events. The question is, could this be because the TB disease and infection we are finding is within the individuals entering late-stage HIV themselves, in which case we don't need to visit their households and could be carrying out TB detection and treatment within the HIV clinics.

Using the model we were able to output whether TB disease and infection are being found in the late-stage individuals themselves or their household members.

Figure 8.7 firstly shows whether those found to need treatment for TB disease or infection were the late-stage individuals, the late-stage individuals and at least one other household member, or at least one member of the late-stage individual's household but not the individual itself. Figure 8.7 secondly shows whether the cases of TB found through intervention 4 were found in the late-stage individuals, or one of their household members.

The graphs show that out of all the households visited, treatment was administered to just the late-stage individual in 10% of cases, however other household members required treatment in a further 5% of cases, and 23% of households required treatment of the late-stage individual's household members but not the late-stage individual itself. Further to that, 69% of the TB cases found in intervention 4, were found to be in the members of the late-stage individuals' household. This shows clearly that a large proportion of the benefit of intervention 4 comes from visiting the households of the late-stage individuals and that if we just treated late-stage individuals in HIV clinics, we could only expect to find 30% of the individuals with TB disease and infection that would be found if their household members were also investigated.

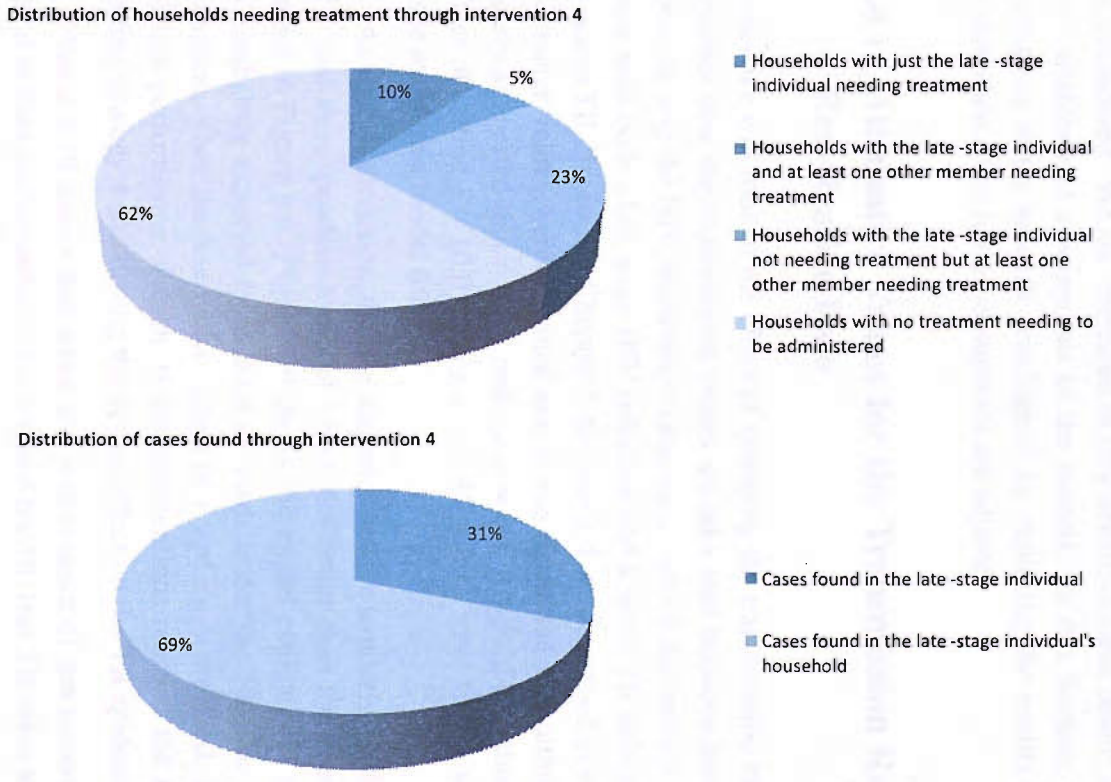


Figure 8.7: Pie charts to show in which households TB treatment is administered and in which members TB cases are found, when visiting late-stage households through intervention 4. Intervention 4 investigates each member of the household of all persons entering late-stage HIV to see if there are any undetected TB cases or members with a TB infection in the household

8.4 Sensitivity Analysis

The results of the modelling suggest that contact-tracing is relatively ineffective when the contacts are restricted to the household members of a person with active TB disease, but very effective when contacts of individuals entering late-stage HIV are considered. We are interested in how sensitive these results are to changes in key variables and assumptions of the models. In this Section, we aim to build confidence in the model's conclusions by exploring the results produced by the model when a few main assumptions are adjusted.

8.4.1 Alternative Values for the Transmission Ratio and HIV Reactivation Rate

Firstly we will look at the effect of changing the transmission ratio, r, which determines how the transmission events are allocated between household or casual contacts; and the HIV reactivation rate, v_{HIV}, which determines the rate at which those with both a late-stage HIV infection and a latent TB infection will progress to active TB disease. In Chapter 7, Section 7.4 we described how values for these parameter values were obtained and it was discussed that although the analysis clearly suggested that r = 9:1 and v_{HIV} = 0.6, it would be valuable to assess the effect of setting r = 10:0 and v_{HIV} = 0.45, which were values which produced a close to optimal model fit.

As already discussed, the two scenarios behave similarly but the original values of r and v_{HIV} produce a model which fits better than the alternative values as shown in Figure 8.8. Figure 8.9 shows the model estimates of the TB incidence in Zimbabwe expected from 2008 onwards under the baseline and intervention scenarios, when the alternative values of r and v_{HIV} are used. It shows that the relative performance of each of the interventions remains the same, with case-finding strategy 4 still having the largest effect on the TB epidemic.

Figure 8.10 shows that when the performance of the interventions are compared to their performance in the original model, less TB cases are found and less cases averted via IPT. This result is expected, as using household interventions for controlling TB when the transmission ratio is 10:0 and there is therefore no household transmission of TB, will be a less effective method than when household transmission is present.

The important result from this analysis is that although household interventions become less effective in this alternative model, we can be confident that if

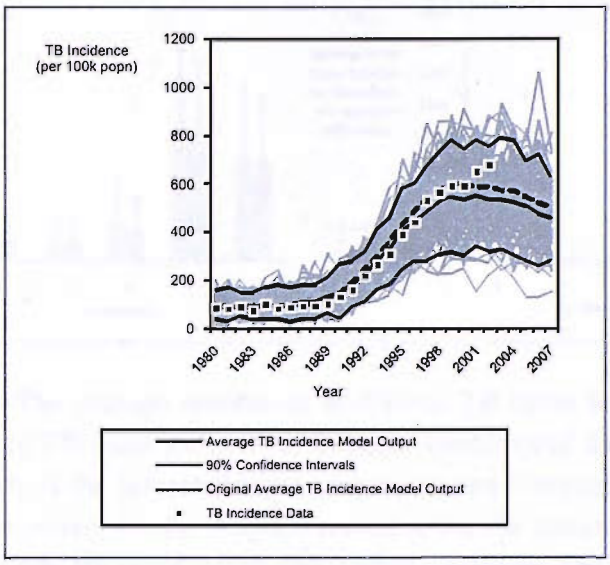


Figure 8.8: Using a model with alternative transmission ratio and HIV reactivation rate values: Model estimates of TB incidence in Zimbabwe. The graph shows the observed TB incidence data from Zimbabwe and the output from 100 runs of the alternative model, along with the average result and 90% confidence intervals. The average result of the original model is included

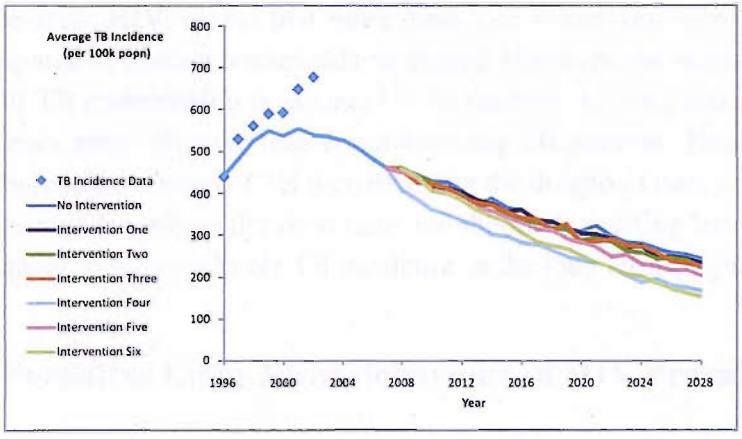


Figure 8.9: Using a model with alternative transmission ratio and HIV reactivation rate values: Model estimates of the TB incidence in Zimbabwe under six intervention scenarios. The graph shows the observed TB incidence data from Zimbabwe and the average TB incidence produced by the model from 100 runs of the baseline scenario and 100 runs of each intervention scenario when implemented in 2008

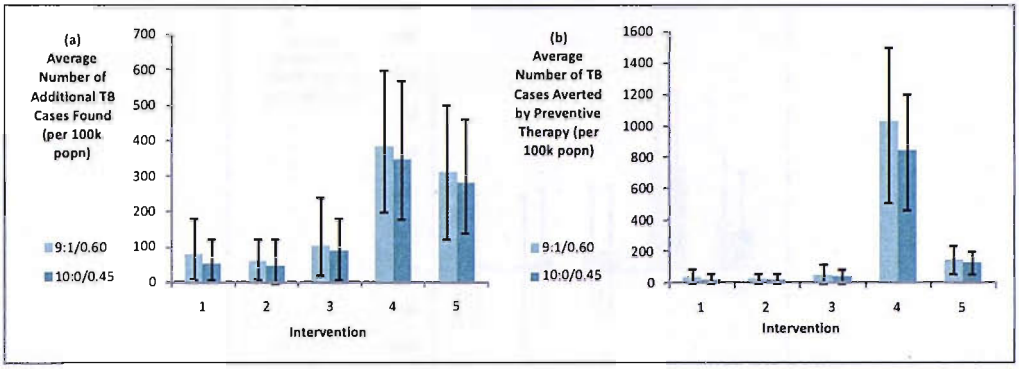


Figure 8.10: (a) The average number of additional TB cases found and (b) the average number of TB cases averted per 100,000 members of the population, by using IPT in each of the household interventions when compared with the base case. Results are shown for the original model when the transmission ratio and HIV reactivation rate are equal to 9:1 and 0.60 respectively, and the model using the alternative values of 10:0 and 0.45 respectively; 90% confidence intervals are included

the values for r and v_{HIV} were to change, targeting HIV-infected households is still the most effective strategy. Scenario 3, visiting households with at least one HIV-positive individual, still performs best when compared to visiting random or TB-diseased households; and scenario 4, visiting households with a member just entering late-stage HIV, means that more cases are found and more TB deaths averted compared to visiting households randomly. However, the results also show that when all TB transmission is assumed to be random, visiting households randomly becomes more effective than contact-tracing TB patients. Finally, if there is no household transmission of TB then doubling the diagnosis rate, whilst not reducing TB as significantly in the short term and therefore averting less TB deaths, would eventually produce a lower TB incidence in the long term (Figure 8.9).

8.4.2 Alternative Long Term Behaviour of HIV Prevalence

The second part of the sensitivity analysis looks at changing the assumptions regarding the expected future behaviour of the HIV epidemic. Currently we fit a double logistic equation to the HIV prevalence data for Zimbabwe to provide estimates of the HIV prevalence for Zimbabwe between 1980 and 2030 (Figure 6.13). The set of HIV prevalence estimates are used to inform the model of the prevalence in a particular year. The estimates obtained for the parameters of the double logis-

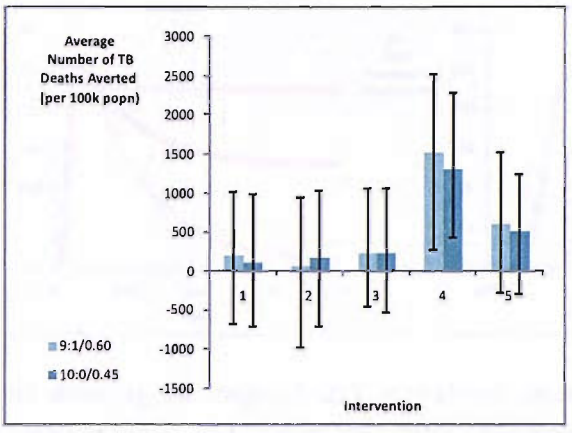


Figure 8.11: The average number of TB deaths averted per 100,000 members of the population, by each of the household interventions when compared with the base case. Results are shown for the original model when the transmission ratio and HIV reactivation rate are equal to 9:1 and 0.60 respectively, and the model using the alternative values of 10:0 and 0.45 respectively; 90% confidence intervals are included

tic equation imply that the HIV epidemic started during 1990, the peak prevalence of the epidemic is 28.2% and that the long-term HIV prevalence will be 13.3% (see Section 6.5).

We look at the effect on the TB epidemic produced by the model and the performance of the interventions, of two different scenarios regarding long-term HIV prevalence. Firstly, we investigate the effect of using a double logistic equation which suggests a higher long-term HIV prevalence, and secondly, a prevalence which is lower, reduces more rapidly and eventually dies out. These two HIV prevalence scenarios can be seen in Figure 8.12 and are compared with the HIV prevalence estimates currently used by the model.

Introducing a higher than predicted long-term HIV prevalence into the model does not have a large impact on the TB epidemic produced, as can be seen in Figure 8.13a. In the scenario where HIV prevalence is higher than predicted, we see that at around 2016, TB incidence becomes higher than expected when the original estimates of HIV prevalence are used. At this same time, approximately 8 years after the interventions are implemented in the model, interventions 1, 2 and 3 show a slight improvement in TB incidence when compared to the baseline scenario (Figure 8.15a). Figure 8.14 compares the number of TB cases found and averted under the different HIV scenarios for the various case-finding strategies. It shows that if long-term HIV prevalence is higher than expected, the performance

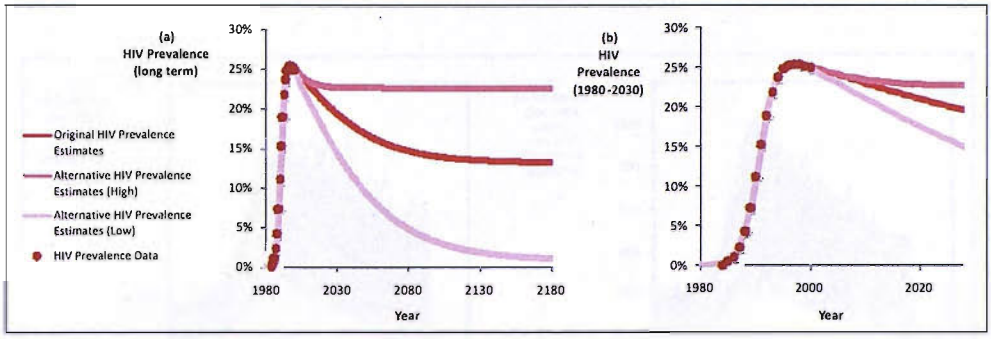


Figure 8.12: Graphs showing the original HIV prevalence estimates used by the model with the alternative lower and higher long-term HIV prevalence scenarios. The graphs show (a) the long-term behaviour of HIV for each scenario and (b) the behaviour during the time being considered by the DES TB model (1980-2028)

of the interventions remains similar with approximately the same number of TB cases being found and averted by the various interventions.

Introducing a lower than predicted long-term HIV prevalence into the model does have an impact on the TB epidemic produced, as can be seen in Figure 8.13b. In the scenario where HIV prevalence is lower than predicted, we see that TB incidence is significantly less than expected when the original estimates of HIV prevalence are used. Figure 8.14 compares the number of TB cases found and averted under the different HIV scenarios for the various case-finding strategies. It shows that if long-term HIV prevalence is lower than expected, the performance of the interventions is less good, with fewer TB cases being found and averted due to the overall lower levels of HIV and TB. The graph also shows that the conclusions regarding the *relative* impact of the interventions on the expected TB epidemic remains the same regardless of which HIV scenario is used. This is supported by Figure 8.15 which shows the model estimates of the TB incidence in Zimbabwe expected from 2008 onwards under the baseline and intervention scenarios when (a) long-term HIV prevalence is high, and (b) long-term HIV prevalence is low. The graphs clearly show the benefit of employing case-finding strategy 4 due to the relative reduction in TB incidence that can be expected regardless of which long-term HIV prevalence is being considered.

Overall, the analysis shows that if the long-term prevalence of HIV is higher than predicted, we can expect a small impact on the TB epidemic and on the performance of the interventions; and if the long-term prevalence of HIV is lower than predicted, the TB epidemic will be smaller and therefore the performance of the interventions less good. The analysis shows, however, that the relative performance

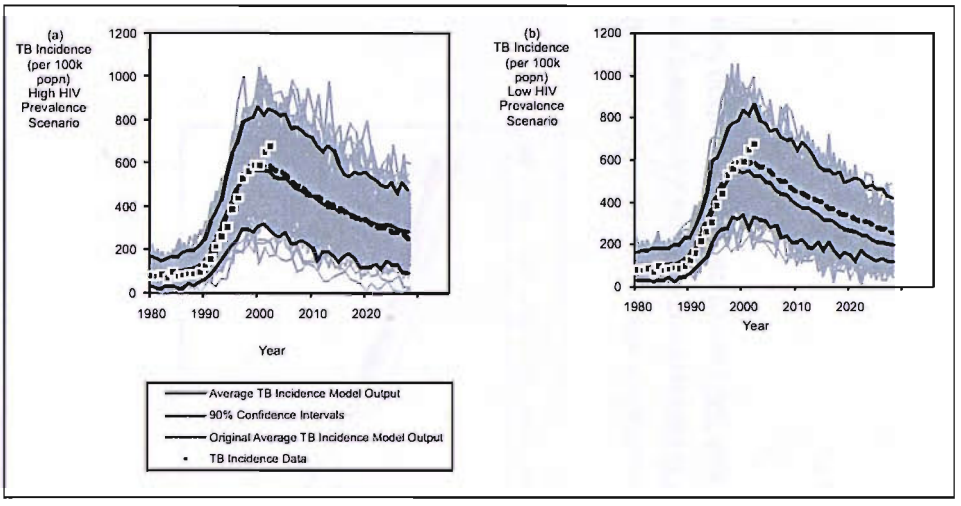


Figure 8.13: Model estimates, when alternative long-term HIV prevalence scenarios are used, of TB incidence in Zimbabwe. The graph shows the observed TB incidence data from Zimbabwe and the output from 100 runs of the model, along with the average result and its 90% confidence intervals when long-term HIV prevalence is (a) higher, and (b) lower than originally predicted. The average result from the original model is also shown for comparison

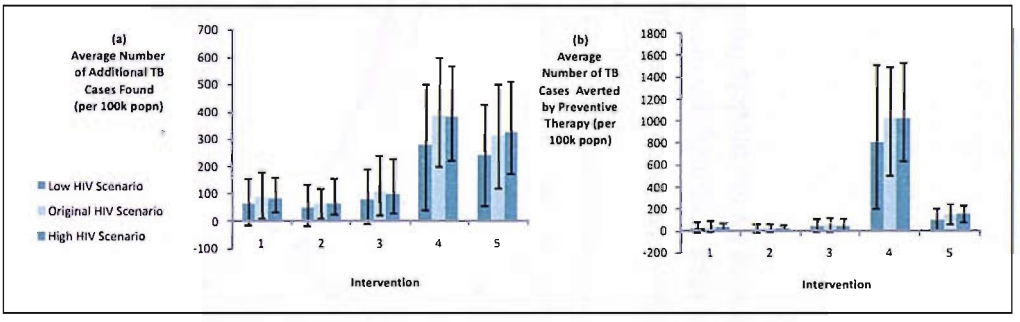


Figure 8.14: (a) The average number of additional TB cases found and (b) the average number of TB cases averted per 100,000 members of the population, by using IPT in each of the household interventions when compared with the base case. Results are shown for when the model uses original HIV prevalence estimates and when its uses estimates suggesting lower or higher long-term HIV prevalence. 90% confidence intervals are included

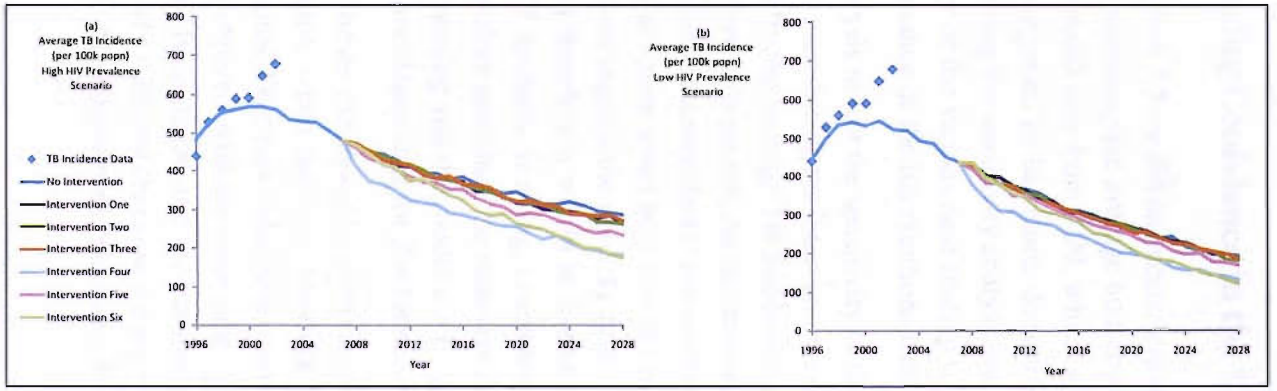


Figure 8.15: Model estimates of the TB incidence in Zimbabwe under six intervention scenarios when long-term HIV prevalence is (a) higher, and (b) lower than originally predicted. The graphs show the observed TB incidence data from Zimbabwe and the average TB incidence produced by the models from 100 runs of the baseline scenario and 100 runs of each intervention scenario when implemented in 2008

of the interventions remains the same regardless of the long-term HIV prevalence scenario being considered, concluding that case-finding strategy 4 would always have the largest impact on reducing TB incidence.

8.4.3 Increasing Confidence in the Conclusions

In Chapter 7, Section 7.5 we did an uncertainty analysis which included investigating the effect of increasing the average household size. It concluded that changing the average household size from 3.99, which was implied by the Harare baseline data, to 5.5, as suggested by literature, does not effect the TB epidemic produced by the model. During the sensitivity analysis we also investigated whether it would affect the efficacy of the various case-finding strategies and found that the conclusions and performance of the interventions were the same.

The final analysis to test the sensitivity of the model's conclusions was to consider the effect of introducing the interventions earlier. The reason for this analysis is two-fold. Firstly, the average TB incidence produced by the model fits ideally to all but the last two data points. As the interventions occur after the last two data points and when TB incidence data is not available, it would be useful to introduce the interventions at a time when we know the model fits accurately. Secondly, the final two data points suggest the unlikely case that TB incidence is still increasing in Zimbabwe and therefore it would be useful to introduce the interventions at a time when the TB epidemic is rising, to see how the interventions perform in this scenario. We therefore introduce the interventions at time 1996, when the TB epidemic is still increasing and the model's average TB incidence output fits well to the available TB incidence data for Zimbabwe.

Figure 8.16 shows the model's estimates of TB incidence under the six intervention scenarios, when they are introduced in 1996. Figure 8.17 shows the number of additional TB cases found and averted by each of the household interventions when compared with the base case. The results show that although the behaviour of the TB epidemic and the number of cases found and averted by the interventions is very different (because we are interfering with the TB epidemic at a different stage), the relative performance of the interventions follows the original conclusions.

Figure 8.17 shows that case-finding strategy 3 is the most effective intervention for finding and averting TB cases when compared to 1 and 2; this suggests a clustering of TB and HIV in households and again suggests that visiting households which contain at least one HIV-positive member is a more effective strategy than

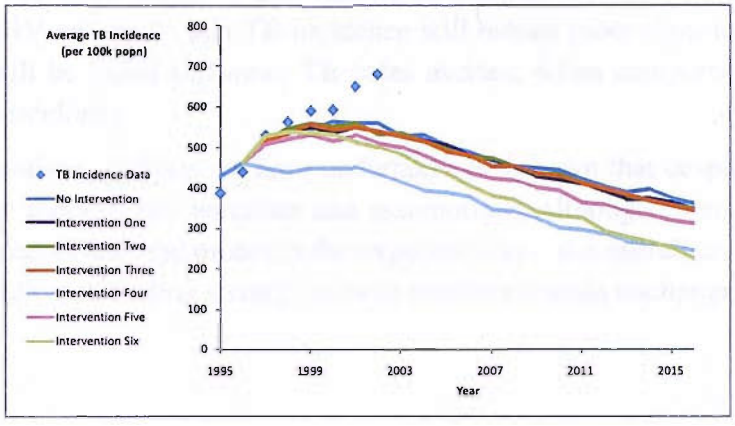


Figure 8.16: Model estimates of the TB incidence in Zimbabwe under six intervention scenarios. The graph shows the observed TB incidence data from Zimbabwe and the average TB incidence produced by the model from 100 runs of the baseline scenario and 100 runs of each intervention scenario when implemented in 1996

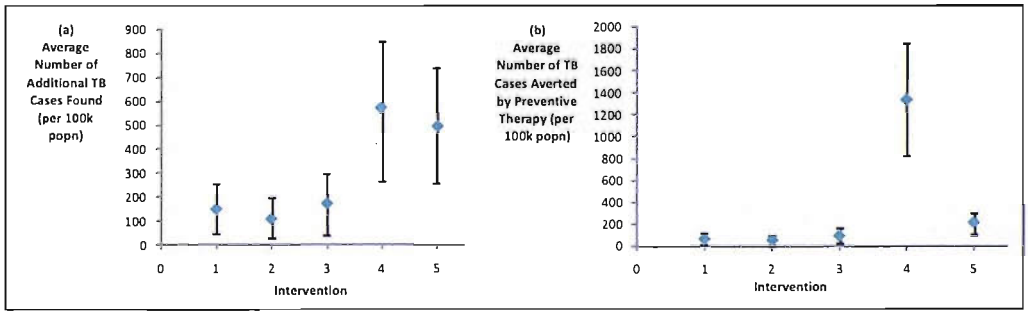


Figure 8.17: (a) The average number of additional TB cases found and (b) the average number of TB cases averted by using IPT (per 100,000 members of the population), in each of the household interventions when implemented in 1996 (when compared with the base case); 90% confidence intervals are included

contact-tracing individuals infected with TB or visiting households randomly. Figures 8.16 and 8.17 clearly show the relative benefit of employing strategy 4 over strategy 5 and again suggests that visiting households with a member entering late-stage HIV will mean that TB incidence will reduce more significantly, more TB cases will be found and more TB cases averted, when compared to visiting households randomly.

The sensitivity analysis we have undertaken has shown that despite changing many of the model's key variables and assumptions, although changing the TB epidemic produced by the model in the expected ways, the overall conclusion regarding which case-finding strategy is most effective remain unchanged.

Chapter 9

Scenario Analysis

We have described the design and development of a discrete event simulation model of TB and HIV which allows us to investigate the significance of household versus community transmission and to analyse interventions targeted at household contacts of TB patients. The results of the modelling and sensitivity analysis suggest that contact-tracing of TB patients is relatively ineffective but that investigating the contacts of persons entering late-stage HIV performs very well and averts more TB deaths and TB cases than any of the other active case-finding strategies.

In this Chapter we explore 8 variations of the original experiment described in Chapter 8. The purpose of the 8 scenarios is to explore the effect of possible changes to the underlying assumptions and design of the case-finding strategies.

9.1 Scenario Design

The original experiment is described in Chapter 8, Section 8.1 and compares the impact of 5 active case-finding strategies, along with a 6th strategy which looks at the effect of doubling the diagnosis rate, on the magnitude of the TB epidemic. The design of these intervention strategies assumes that isoniazid preventive therapy (IPT) is administered to all those household members with a latent infection (diagnosed by a positive tuberculin skin test result) for a duration of 6 months, and that IPT has an efficacy of 50% [52] to account for incomplete protection due to non-compliance or drug resistance.

These assumptions are based on what is widely accepted and used by medical professionals belonging to CREATE (The Consortium to Respond Effectively to the AIDS-TB Epidemic) [167] and involved in intervention trials in Africa. There

is much discussion in the medical community however, over the ideal duration of IPT, to whom IPT should be administered and the level of protection that it offers.

The first set of scenarios that we investigate explore the effect of changing the duration of IPT treatment. Firstly, we look at the effect of administering treatment for a duration of 9 months as recommended by the American Thoracic Society and the United States' Centers for Disease Control [204]; and secondly, a lifetime of IPT is administered on diagnosis of TB infection. This is currently being tried in adults in Botswana [32] and explored in children in Cape Town and Stellenbosch, South Africa [205].

The second set of scenarios that we investigate explore the effect of altering the level of protection gained from IPT. We have modelled IPT as giving 50% protection from TB, to account for the incomplete protection observed in trials, and the combined effects of isoniazid resistant TB strains plus imperfect compliance. As discussed, this assumption regarding the level of protection IPT provides was supplied by medical professionals in CREATE and TB epidemiologists belonging to the WHO [52], it is however useful to look at the sensitivity of the results to changes in this level and we therefore investigate the effect of a 25%, 75% and 100% efficacy.

The final set of scenarios that we investigate involve adjusting the proportion of households visited in case-finding strategy 4 (and therefore 5). The current experiment looks at visiting all of the households of those individuals entering late-stage HIV. This last set of scenarios looks at the effect of visiting only 25%, 50% and 75% of the individuals' households to account for the fact that not all individuals entering late-stage HIV will seek medical attention and/or diagnosis of their HIV infection.

The rest of this Chapter considers the results and implications of the 8 different scenarios. We divide the discussion into three sections to examine the effect of changing the duration of IPT treatment; the effect of altering the level of protection gained from IPT; and the impact of changing the proportion of households visited in interventions 4 and 5.

9.2 Scenario Results: Adjusting IPT Duration

The original experiment assumes the interventions administer IPT to all household members with a latent infection for a duration of 6 months (see Chapter 8, Section 8.1).

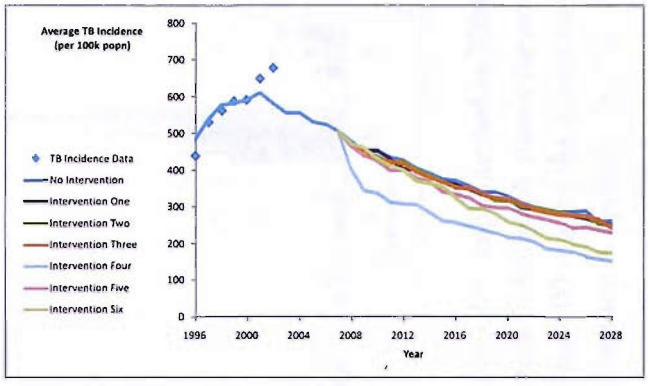


Figure 9.1: Model estimates of the TB incidence in Zimbabwe under six intervention scenarios with IPT being administered for 9 months. The graph shows the observed TB incidence data from Zimbabwe and the average TB incidence produced by the model from 100 runs of the baseline scenario and 100 runs of each intervention scenario when implemented in 2008

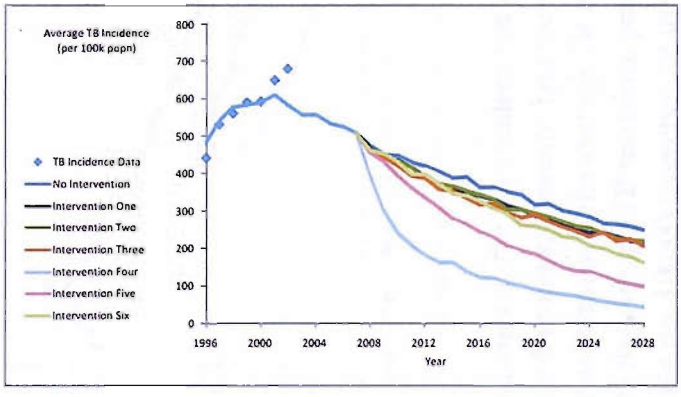


Figure 9.2: Model estimates of the TB incidence in Zimbabwe under six intervention scenarios with IPT being administered for life. The graph shows the observed TB incidence data from Zimbabwe and the average TB incidence produced by the model from 100 runs of the baseline scenario and 100 runs of each intervention scenario when implemented in 2008

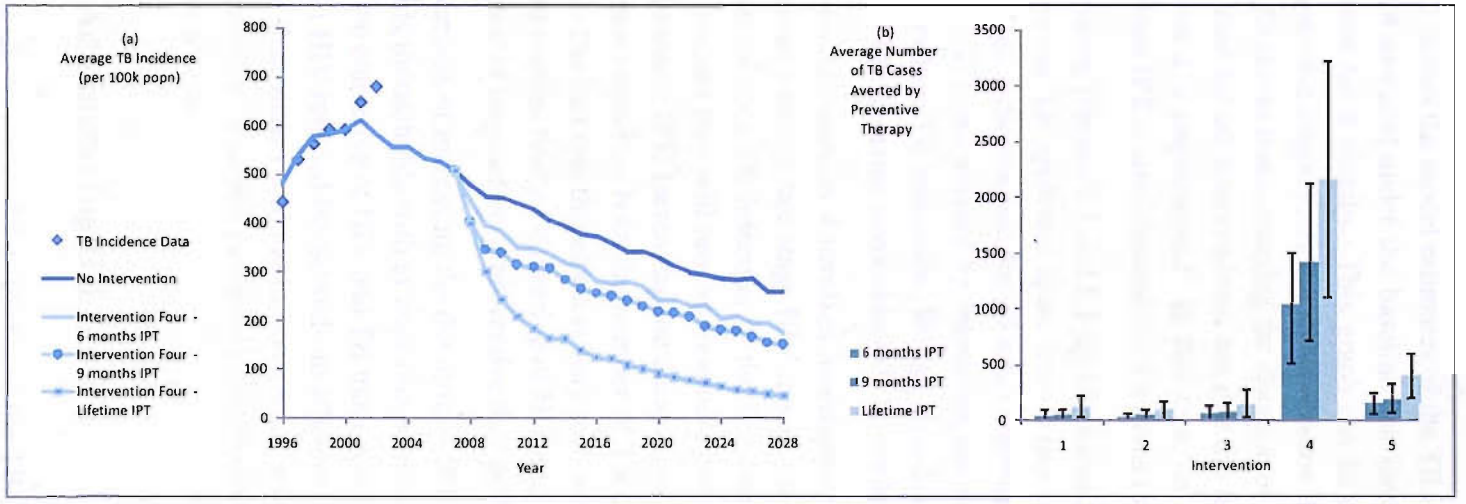


Figure 9.3: (a) Model estimates of the TB incidence in Zimbabwe when intervention scenario 4 is implemented in 2008, with IPT being administered for 6 months, 9 months and the rest of an infected individual's lifetime. The graph shows the observed TB incidence data from Zimbabwe and the average TB incidence produced by the model from 100 runs of the 3 scenarios. The estimates are compared to the average TB incidence produced by the model when no intervention is applied. (b) The average number of TB cases averted per 100,000 members of the population, by using a 6 month, 9 month and lifetime duration of IPT in each of the household interventions when compared with the base case; 90% confidence intervals are included

9.2.1 Administering IPT for 9 months

Figure 9.1 shows the model estimates of the TB incidence in Zimbabwe expected from 2008 onwards under the baseline and intervention scenarios when IPT is administered for 9 months. This graph can be directly compared to Figure 8.1 which shows the original model estimates when IPT is administered for 6 months. Figure 9.3b shows that increasing the duration of IPT increases the number of TB cases averted for *all* interventions, but that the increase is only substantial when intervention 4 is implemented. In this case, an average of 36% more cases are averted when IPT is administered for 9 months rather than 6 months.

Comparing Figures 9.1 and 8.1 shows that increasing the duration of IPT does not impact the TB epidemic apart from when case-finding strategy 4 is implemented. This indicates that for all other interventions, the slight increase in the number of TB cases averted by increasing the duration of IPT, is not significant enough to reduce TB incidence. During case-finding strategy 4 however, the impact on the TB epidemic is considerable as can be seen in Figure 9.3a.

Case-finding strategy 4 involves investigating each member of the household of all persons entering late-stage HIV to see if there are any undetected TB cases or members with a TB infection in the household. If an individual is found to have TB disease they will receive treatment and if found to have a TB infection, they will receive IPT. Increasing the duration of IPT during this intervention has a significant impact on both the number of TB cases averted and the overall TB epidemic. The fact that this effect is only seen when case-finding strategy 4 is implemented implies firstly a clustering of HIV and TB in households, and secondly, the potential of targeted mass preventive therapy (discussed in Section 9.2.2).

The success of increasing the duration of IPT being administered in late-stage households (households with at least one member with late-stage HIV) is further proof of the clustering of HIV and TB that occurs within households which makes looking in HIV-infected households an efficient way of finding TB infection and, further to that, makes looking in households with *late-stage* HIV members even more efficient as it means preventing the progression of TB disease in the most vulnerable group.

9.2.2 Administering IPT for life

Figure 9.2 shows the model estimates of the TB incidence in Zimbabwe expected from 2008 onwards under the baseline and intervention scenarios when IPT is

administered for the rest of an individual's lifetime once their infection has been detected. This graph can be directly compared to Figure 8.1 which shows the original model estimates when IPT is administered for 6 months. Figure 9.3b shows that administering IPT for life significantly increases the number of TB cases averted for all interventions.

Comparing Figures 9.2 and 8.1 shows that administering IPT for the rest of an individual's lifetime impacts the TB epidemic in all interventions. For the first time we even see a reduction in TB incidence in case-finding strategies 1 to 3. This implies that contact-tracing of TB patients will only reduce TB incidence if IPT is administered for life. Strategies 4 and 5 are still by far the best interventions and are now clearly better than doubling the rate at which TB patients present for treatment. The impact of case-finding strategy 4 when a lifetime of IPT is administered on diagnosis of infection, is even more substantial than in previous experiments, with TB incidence expected to halve in just 3 years and return to pre-HIV levels within 13 years.

As previously discussed in Section 9.2.1, the success of increasing the duration of IPT being administered in late-stage households is due to TB being prevented in the groups most vulnerable to disease progression. The success also highlights the potential impact of targeted mass preventive therapy (where IPT is given to a large proportion of a *specific* population). The effect of treating a large proportion of the population on the TB epidemic is apparent, but what is interesting is the obvious benefit of targeting the late-stage households versus the population in general (intervention 4 versus intervention 5 in Figures 9.1 and 9.2); and when late-stage households are targeted, the difference between administering a 6 month, 9 month or lifetime course (Figure 9.3a). This implies that although mass preventive therapy has the potential to rapidly reduce TB incidence in populations, this effect is significantly reduced by a short treatment duration and by not targeting a specific population.

9.3 Scenario Results: Adjusting IPT Efficacy

The original experiment assumes that the IPT administered during the interventions has an efficacy of 50%. The set of scenarios discussed in this Section investigate the sensitivity of the conclusions regarding intervention design to changes in the assumption regarding IPT efficacy. Although we expect to see an improvement in each intervention's performance as the efficacy increases, we are specifically interested in whether the relative performance of each intervention changes if the

efficacy assumption is altered, to ascertain whether the same case-finding strategy would be recommended regardless of actual IPT efficacy.

9.3.1 Administering IPT with an Efficacy of 25%

Figure 9.4a shows the model estimates of the TB incidence in Zimbabwe expected from 2008 onwards under the baseline and intervention scenarios when IPT is administered with an efficacy of 25%. This graph can be directly compared to Figure 9.4b which shows the original model estimates when IPT has an efficacy of 50%.

The graph shows that interventions 1 to 3 still have no impact on the TB epidemic and that if IPT only has an efficacy of 25%, intervention 5 would also have little impact on TB incidence. Intervention 4 still performs the best at preventing TB cases and reducing TB incidence, although the average reduction in TB incidence is less than half that of when the efficacy was 50%. Figure 9.5 shows the average number of TB cases prevented per 100,000 members of the population for each intervention under the different IPT efficacy assumptions, and shows that reducing the efficacy of IPT from 50% to 25% would mean an average decrease of 44% in the number of TB cases prevented by each of the interventions.

9.3.2 Administering IPT with an Efficacy of 75%

Figure 9.4c shows the model estimates of the TB incidence in Zimbabwe expected from 2008 onwards under the baseline and intervention scenarios when IPT is administered with an efficacy of 75%. This graph can be directly compared to Figure 9.4b which shows the original model estimates when IPT has an efficacy of 50%.

The graph shows that interventions 1 to 3 still have very little impact on the TB epidemic and that if IPT efficacy is increased from 50% to 75% there is little improvement in the impact of intervention 5 on the TB epidemic. Intervention 4 still performs the best at reducing TB incidence and preventing TB cases with a 50% increase in efficacy causing just less than a 50% increase in the average reduction of TB incidence. Figure 9.5 shows the average number of TB cases prevented per 100,000 members of the population for each intervention under the different IPT efficacy assumptions, and shows that increasing the efficacy of IPT from 50% to 75% would mean an average increase of 52% in the number of TB cases prevented by each of the interventions.

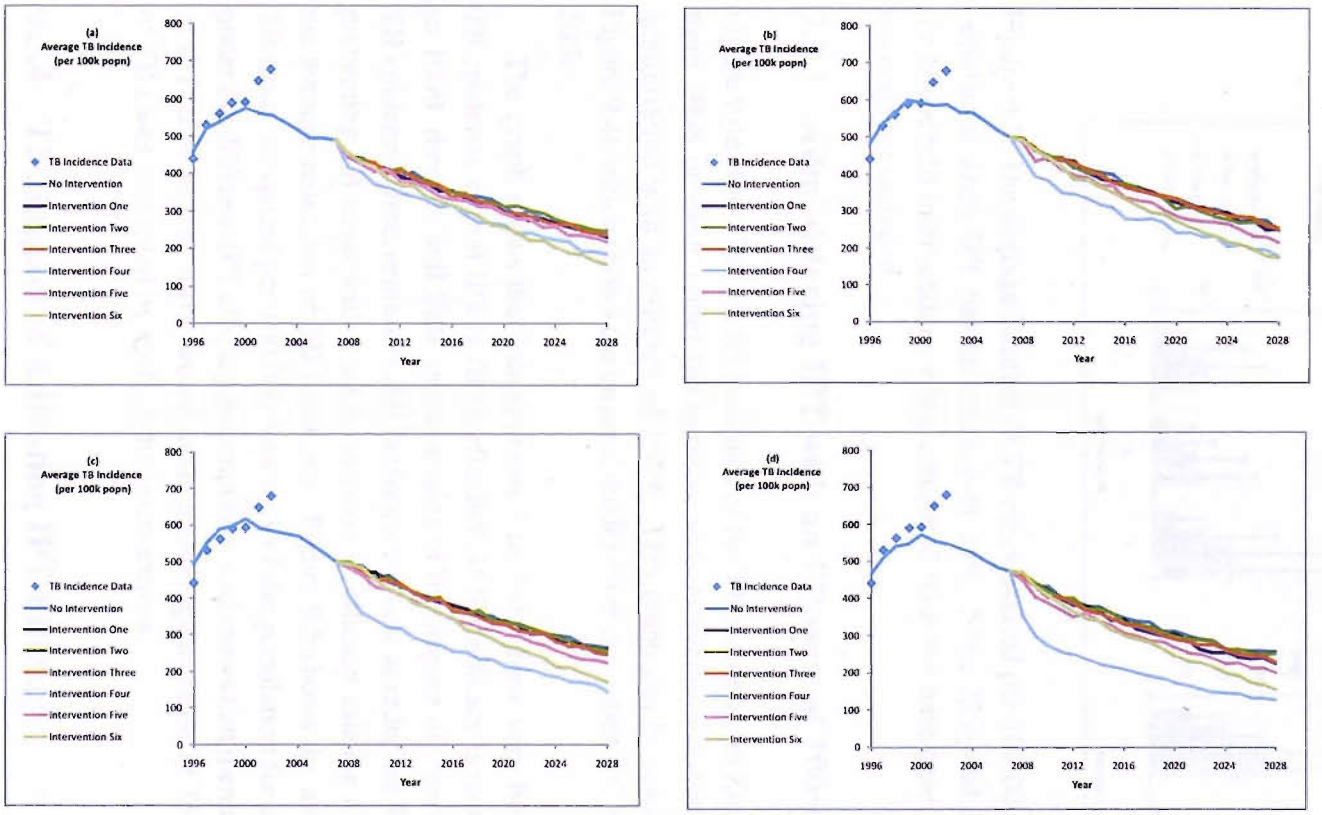


Figure 9.4: Model estimates of the TB incidence in Zimbabwe under six intervention scenarios with IPT having an efficacy of (a) 25%, (b) 50%, (c) 75% and (d) 100%. The graph shows the observed TB incidence data from Zimbabwe and the average TB incidence produced by the model from 100 runs of the baseline scenario and 100 runs of each intervention scenario when implemented in 2008

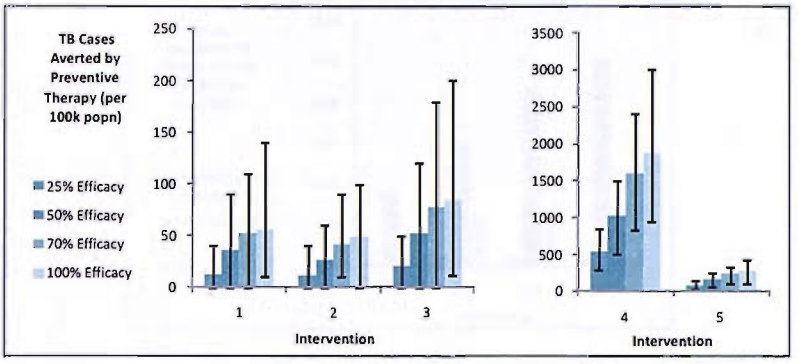


Figure 9.5: The average number of TB cases averted per 100,000 members of the population when IPT has an efficacy of 25%, 50%, 75% and 100% in each of the household interventions when compared with the base case; 90% confidence intervals are included

9.3.3 Administering IPT with an Efficacy of 100%

Figure 9.4d shows the model estimates of the TB incidence in Zimbabwe expected from 2008 onwards under the baseline and intervention scenarios when IPT is administered with an efficacy of 100%. This graph can be directly compared to Figure 9.4b which shows the original model estimates when IPT has an efficacy of 50%.

The graph shows that interventions 1 to 3 still have very little impact on the TB epidemic even if IPT is 100% effective. If IPT efficacy is increased from 50% to 100% there is still little improvement in the impact of intervention 5 on the TB epidemic. Intervention 4 still performs the best at reducing TB incidence and preventing TB cases with a 100% increase in efficacy causing a 72% increase in the average reduction of TB incidence. Figure 9.5 shows the average number of TB cases prevented per 100,000 members of the population for each intervention under the different IPT efficacy assumptions, and shows that increasing the efficacy of IPT from 50% to 100% would mean an average increase of 73% in the number of TB cases prevented by each of the interventions.

9.3.4 The Impact of Adjusting IPT Efficacy

In summary, even increasing the efficacy of IPT to 100% means that interventions 1 to 3 do not impact on the TB epidemic. These interventions rely on TB patients dying or presenting for treatment, causing either a TB, HIV or random household

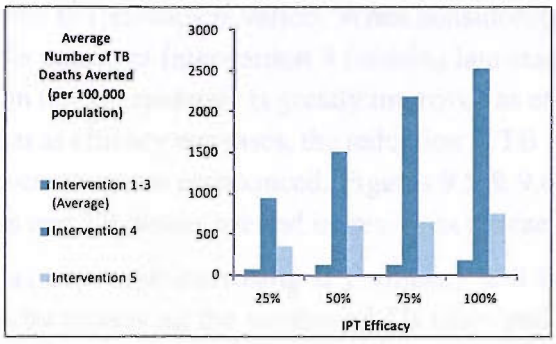


Figure 9.6: The average number of TB deaths averted per 100,000 members of the population when IPT has an efficacy of 25%, 50%, 75% and 100% in each of the household interventions when compared with the base case; 90% confidence intervals are included

to be visited, tested for TB and appropriately treated. Increasing the efficacy of IPT administered to individuals with a TB infection does increase the number of TB cases averted (Figure 9.5) and the number of TB deaths averted (Figure 9.6) however, the number is not significant enough to visibly impact the TB epidemic and cause a reduction in TB incidence.

Regardless of the efficacy of IPT, we find that the original results and conclusions regarding case-finding strategies 1 to 3 (discussed in Chapter 8) hold, suggesting that contact-tracing of TB patients is not the most effective strategy. The fact that regardless of the efficacy of IPT, visiting households with at least one HIV-positive member means more TB cases are found, and more TB deaths and cases are averted, indicates that the clustering of TB in HIV-positive households is so important that even if IPT has a low efficacy, it is still more effective to concentrate interventions in HIV households rather than random households or households with at least one TB diseased member.

When considering the impact of IPT efficacy on the performance of intervention 5 we see that its impact on reducing TB incidence does not change as efficacy increases, but that there is an increase in the average number of TB cases and TB deaths averted (Figures 9.5 & 9.6). Increasing the efficacy from 25% to 50%, for example, means an average of 250 extra TB deaths per 100,000 members of the population would be averted. This indicates that increasing the efficacy of IPT in intervention 5 has a large benefit, but that this benefit will not be seen in the annual TB incidence figures.

Finally, the analysis shows that intervention 4 still out performs the other in-

terventions even when IPT efficacy is varied. When considering the impact of IPT efficacy on the performance of intervention 4 (visiting late-stage households), we see that its impact on the TB epidemic is greatly improved as efficacy is increased. Figure 9.4 shows that as efficacy increases, the reduction in TB incidence produced by intervention 4 becomes more pronounced. Figures 9.5 & 9.6 also show that the number of TB cases and TB deaths averted increases as efficacy improves.

In conclusion, as expected, increasing IPT efficacy will increase the benefit of the interventions by increasing the number of TB cases and TB deaths averted by each of them. The biggest improvement comes from increasing the efficacy from 25% to 50% where doubling the efficacy almost doubles the average benefit in terms of the average number of TB deaths and cases averted per 100,000 members of the population. Improving the efficacy from 25% to 75% or 100% (a 200% and 300% increase) means an average increase in benefit of 165% and 214% making the increases less economic. Although this increase in benefit is seen in all the interventions, the only intervention in which a noticeable change in TB incidence will occur as efficacy is varied, is intervention 4 (visiting late-stage households). Most importantly, the analysis shows that the relative performance of each intervention does not change as the efficacy assumption is altered and therefore the same conclusions are made regarding which case-finding strategy is most efficient, regardless of actual IPT efficacy.

9.4 Scenario Results: Adjusting the Proportion of Late-Stage Households Visited

The original experiment assumes that in case-finding strategy 4, we investigate each member of the household of all persons entering late-stage HIV to see if there are any undetected TB cases or members with a TB infection in the household (we assume that it is around this time that an HIV-positive individual will approach the health services). In reality, not 100% of HIV-positive individuals will seek medical attention when becoming sick and therefore this set of scenarios explores the relative performance of intervention 4 (and therefore 5) when the assumption regarding the proportion of late-stage households that are visited is varied. A late-stage household is a household of an HIV-positive individual who has just entered late-stage HIV. We look at the impact of only 25%, 50% and 75% of these individuals presenting themselves for medical attention, causing their household to be investigated for TB. Graphs showing the average number of additional TB cases

found, TB deaths averted and TB cases averted, per 100,000 members of the population, by each of the household interventions in each of the scenarios can be seen in Appendix O. Appendix N gives details of the number of households visited in each of the scenarios.

9.4.1 Visiting 25% of Late-Stage Households

Figure 9.7a shows the model estimates of the TB incidence in Zimbabwe expected from 2008 onwards under the baseline and intervention scenarios when 25% of late-stage households are investigated. This graph can be directly compared to Figure 9.7d which shows the original model estimates when 100% of late-stage households are visited.

The graph shows that intervention 6 (doubling the rate at which people with TB present for treatment) is the only intervention which causes a noticeable reduction in TB incidence. All of the case-finding strategies have little impact on the TB epidemic and when the other performance measures are examined we see that intervention 4 still performs best at finding the most TB cases and averting the most TB deaths and TB cases (Figure O.1). If only 25% of late-stage individuals present themselves for medical attention, intervention 4 would still be recommended as the most effective case-finding strategy. Intervention 6 would be recommended overall as if a mechanism for encouraging twice as many individuals with TB to present for treatment was established, this would reduce TB incidence and TB deaths more than any active case-finding strategy (Figure O.1).

9.4.2 Visiting 50% of Late-Stage Households

Figure 9.7b shows the model estimates of the TB incidence in Zimbabwe expected from 2008 onwards under the baseline and intervention scenarios when 50% of late-stage households are investigated. This graph can be directly compared to Figure 9.7d which shows the original model estimates when 100% of late-stage households are visited.

The graph shows that interventions 4 (visiting late-stage households) and 6 (doubling the rate at which people with TB present for treatment) are the only interventions which cause a noticeable reduction in TB incidence. All of the other case-finding strategies have little impact on the TB epidemic. When the other performance measures are examined we see that intervention 4 is the case-finding method which finds the most TB cases and averts the most TB deaths and TB

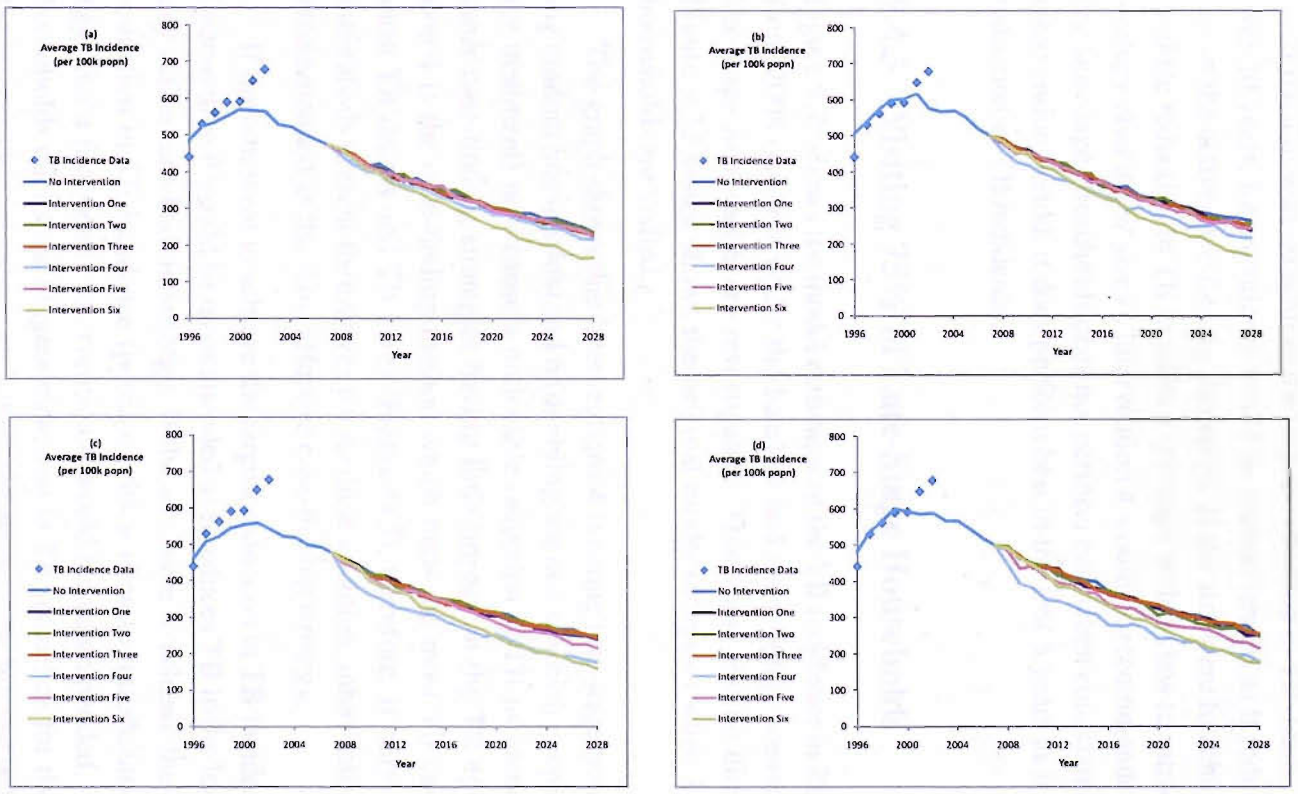


Figure 9.7: Model estimates of the TB incidence in Zimbabwe under six intervention scenarios with (a) 25%, (b) 50%, (c) 75% and (d) 100% of late-stage households being visited. The graph shows the observed TB incidence data from Zimbabwe and the average TB incidence produced by the model from 100 runs of the baseline scenario and 100 runs of each intervention scenario when implemented in 2008

cases (Figure O.2). If only 50% of late-stage individuals present themselves for medical attention, intervention 4 would still be recommended as the most effective case-finding strategy.

If the aim were to achieve the largest reduction in TB deaths and TB incidence over 20 years, intervention 6 would be recommended as it reduces TB more than any of the active case-finding strategies. If the aim were to achieve the largest immediate reduction in TB incidence (perhaps with a view to introducing a different strategy after a few years), intervention 4 would be recommended. Although visiting late-stage households does not perform best when considered over the 20 year intervention period, it does perform best in the first 5 years as it elicits the biggest reduction in TB incidence.

9.4.3 Visiting 75% of Late-Stage Households

Figure 9.7c shows the model estimates of the TB incidence in Zimbabwe expected from 2008 onwards under the baseline and intervention scenarios when 75% of late-stage households are investigated. This graph can be directly compared to Figure 9.7d which shows the original model estimates when 100% of late-stage households are visited.

The graph shows that interventions 4 (visiting late-stage households), 5 (visiting random households) and 6 (doubling the rate at which people with TB present for treatment) now cause a noticeable reduction in TB incidence, with all of the other case-finding strategies having little impact on the TB epidemic. Intervention 4 is the case-finding method which finds the most TB cases and averts the most TB deaths and TB cases (Figure O.3), therefore, if only 75% of late-stage individuals present themselves for medical attention, intervention 4 would still be recommended as the most effective case-finding strategy.

If the aim were to achieve the largest reduction in TB incidence over 20 years, intervention 6 would be recommended as it reduces TB in the long term more than any active case-finding strategy. If the aim were to achieve the largest immediate reduction in TB incidence (perhaps with a view to introducing a different strategy after a few years), intervention 4 would be recommended. Visiting late-stage households elicits the biggest reduction in TB incidence for the first 12 years of the intervention and although it does not produce the biggest reduction in the long term, it does avert the largest number of TB deaths per 100,000 members of the population (Figure O.3).

9.4.4 The Impact of Adjusting the Proportion of Late-Stage Households Visited

In summary, the analysis shows that when the assumption regarding the proportion of late-stage individuals that will seek medical attention (and therefore the proportion of late-stage households that are visited) is varied, the conclusions regarding which intervention should be selected will change.

If only 25% of late-stage individuals present themselves to the medical services causing their household to be visited and examined, establishing a way to double the rate at which individuals with TB present themselves (intervention 6) is the most effective intervention for reducing TB.

If the proportion is 50% or 75%, the recommended strategy would depend on the overall aim of the intervention. If the aim of the intervention were to achieve the biggest possible reduction in TB incidence in the long term (over the 20 years), establishing a way to double the rate at which individuals with TB present themselves (intervention 6) would again be recommended. If the aim were to achieve the biggest possible immediate reduction in TB incidence (in the first 5-10 years of the intervention), visiting late-stage households (intervention 4) would be recommended. If the aim were to reduce the number of TB deaths per 100,000 members of the population, intervention 6 would be suggested if 25% or 50% of late-stage households could be visited, and intervention 4 would be suggested if 75% or 100% of late-stage households could be visited.

In conclusion, assuming the aim of an intervention is to reduce TB incidence as quickly as possible and to reduce the number of TB deaths, at least 50% of late-stage households would need to be visited in order for intervention 4 to be recommended, otherwise intervention 6 would be recommended. If intervention 6 were not possible, in other words, a method for doubling the rate at which individuals with TB present themselves could not be established, a case-finding strategy would have to be considered. Interestingly, intervention 4 always performs the best out of the case-finding strategies and therefore if a case-finding strategy were to be recommended, visiting late-stage households would always be suggested assuming between 25% and 100% of late-stage households could be identified.

9.5 Conclusions

This Chapter has explored 8 scenarios, which are variations of the original experiment described in Chapter 8, to explore the effect of possible changes to the

underlying assumptions and design of the case-finding strategies.

The first set of scenarios investigated the effect of changing the duration of IPT treatment. The analysis highlighted the potential impact of targeted mass preventive therapy. It firstly showed the benefit of administering IPT for the rest of an individual's lifetime once their infection has been detected, and secondly showed the benefit of targeting late-stage households versus the population in general.

The second set of scenarios investigated the effect of altering the level of protection gained from IPT. The analysis showed that increasing IPT efficacy will increase the benefit of the interventions by increasing the number of TB cases and TB deaths averted by each of them. Although this increase in benefit is seen in all the interventions, the only intervention in which a noticeable change in TB incidence will occur is intervention 4 (visiting late-stage households). The analysis also showed that the relative performance of each intervention does not change as the efficacy assumption is altered and therefore the same conclusions are made regarding which case-finding strategy is most efficient, regardless of actual IPT efficacy.

The final set of scenarios investigated the effect of adjusting the proportion of households visited in case-finding strategy 4 (and therefore 5). The analysis showed that which intervention should be selected will depend on the expected proportion of late-stage individuals that will seek medical attention, therefore enabling their household to be visited and examined. It also showed that if less than 100% of late-stage individuals present themselves to the medical services, that the choice of intervention will depend on the aim of the intervention, whether that be to achieve the largest reduction in TB incidence in the short term or the long term.

An interesting result that can be seen throughout all of the scenarios is the relative performance of interventions which target households of people with TB or HIV compared with untargeted or community-wide interventions. We make the assumption that visiting households randomly can represent the effect of an untargeted or community-wide intervention where individuals volunteer themselves to attend a mobile TB clinic in their village, for example. We assume therefore, that interventions 2 and 5, in which random households are visited, can be used to approximate an untargeted or community-wide intervention, as these interventions involve individuals being randomly selected from the population to be tested. When interventions 1 to 3 are compared, we find that if HIV or TB households are targeted, the intervention performs better than if a community-wide intervention were used (represented by intervention 2). More TB cases are found, more TB deaths averted and more TB cases averted in the targeted household inter-

ventions. Similarly, when interventions 4 and 5 are compared, we find that the targeted household intervention (intervention 4) performs significantly better than the community-wide intervention (represented by intervention 5) at reducing TB incidence, finding TB cases, averting TB deaths and averting TB cases. These conclusions are not sensitive to changes in the various assumptions explored in this Chapter meaning we are confident that targeted household interventions may be more effective than untargeted or community-wide interventions.

Chapter 10

Discussion

It is believed that this thesis describes one of the first attempts at an individual-based microsimulation model of TB. The aim of the research is to examine the effect of active case-finding strategies for TB control in high HIV prevalent settings, specifically Harare, Zimbabwe. The active case-finding strategies we investigate involve targeting members of random versus TB or HIV infected households, in order to get a clearer understanding of the role of household versus community transmission of TB.

The research objectives for the work reported in this thesis are provided in Chapter 1, Section 1.1. The next Section summarises the work done under these objectives.

10.1 Summary of Research

1. To develop a mathematical model of TB transmission and disease in Harare, Zimbabwe, in order to enable accurate simulation of possible active case-finding strategies for TB control in HIV prevalent populations.

To answer this research question we began with an examination of the literature to obtain an understanding of TB epidemiology (Chapter 3), we explored the dynamic relationship between TB and HIV using a simple parametric model (Chapter 4), and we considered the structure of previous models to investigate possible representations of the natural history of TB (Chapter 5).

Having looked into the natural history of TB, the interaction between TB and HIV, and previous infectious disease and tuberculosis modelling literature we had

a comprehensive appreciation of which processes would need to be captured and how. This enabled us to develop a DES model of TB transmission and disease (Chapters 6 & 7).

The model describes the movement of individuals through seven possible stages of TB disease. The movements through the pathways of the model are determined by an individual's attributes (age, gender, HIV status), with the distributions used to describe progression currently based on information gathered during the literature review (Chapter 3), development of the simple DCM (Chapter 5), and data from the Harare survey conducted by DETECTB (Chapter 2, Section 2.6.1).

Tuberculosis is usually transmitted by an infectious person coughing; therefore close contact is required for transmission, making transmission within the household particularly important. Hence we modelled transmission within the household separately from transmission within the general community. A further reason for separating household and community transmission was to allow us to evaluate household strategies for active case-finding of persons with active TB disease.

A validation and verification of the model was undertaken (Chapter 7, Section 7.7). This included frequent discussion of the model with DETECTB and expert medical and infectious disease modelling audiences where the assumptions, algorithms and various modelling aspects were discussed. This ensured that any errors in epidemiological understanding were reduced and the necessary epidemiological complexities were identified and satisfactorily represented in the model as it evolved. It also meant that other more general modelling issues such as the way in which the population is initially generated or the way in which births and deaths are managed, could be discussed. Expert suggestion was used throughout the development of the model to improve the accuracy of these and other processes. Areas in which the processes could still be improved upon were highlighted through these discussions and are considered in Sections 10.2 and 10.3 which discuss the limitations of the research and what further work will be done.

Various graphical comparisons of model output to available data were also considered to see whether the outputs of the simulation model have the required accuracy for the purpose of the analysis. This included a comparison of historical TB incidence data with the model output of TB incidence (Chapter 7, Figure 7.11) which suggested that the model is able to reproduce TB incidence rates in Zimbabwe both before and after the onset of the HIV epidemic. The model provides a good description of HIV-negative TB but as the model does not produce as sustained an increase in TB cases as the historical data, it could suggest that the model may be underestimating the effect of HIV on TB disease as the HIV

epidemic matures.

Consideration of this model output and the opinion of various subject matter expects, suggests the simulation model described in this thesis imitates the HIV prevalent population of Harare and the effects of household versus random transmission of TB. The model is regarded as accurate enough to compare the relative effectiveness of household interventions in controlling TB and the outcomes of different active case-finding methods.

This study uses data from a large population-based trial in Harare, Zimbabwe. However only TB incidence and HIV prevalence data for Zimbabwe were available to compare with the model's output. It would have been beneficial to have had longitudinal TB incidence and HIV prevalence data so that the model's output could have been compared with an urban setting rather than country-wide statistics. The model would have also benefited from having data on the sexual behaviour in Zimbabwe, so that the age distribution of new HIV infections in the model could be related to the age distribution of sexual activity in Zimbabwe rather than the UK (Chapter 6, Section 6.5.1). Finally, to improve the distributions used to describe disease progression, it would have been useful to have had data on the time it takes for an individual to become diagnosed with TB, and the time it takes for an individual to default from treatment.

2. To enable comparative projections of the likely impact of possible strategies relative to one another to allow a full assessment of the effectiveness of different contact-tracing and case-finding strategies in HIV prevalent populations.

We considered some active case-finding strategies to test on the model which are described in Chapter 8, Section 8.1. Three were: contact-tracing TB patients, contact-tracing HIV-positive individuals and contact-tracing late-stage HIV individuals (HIV-positive individuals entering late-stage HIV are assumed to seek medical attention at this time). These strategies involve visiting the household members of a person being put on TB treatment or dying of TB; or of a person with a particular HIV status. Those found to have active TB disease are treated promptly with the aim of reducing the time spent with infectious TB and so cutting transmission rates, and those found to have a TB infection are given a course of isoniazid preventive therapy (IPT) with the aim of preventing the infection from progressing to active TB. These three case-finding strategies are compared with visiting the same number of households randomly allowing a direct comparison and full assessment of the effectiveness of the different strategies to be made. An

additional intervention which looked at the effect of doubling the rate at which TB patients present themselves for treatment was also tested.

We ran the simulation model to compare the strategies and output projections of the TB incidence expected under each scenario. We also output the difference between the average number of TB deaths averted, TB cases found and TB cases averted per 100,000 members of the population between the base case and the six strategies. This allowed us to assess the impact of the different interventions relative to each other and to examine their effectiveness as TB control methods in HIV prevalent settings.

The results are explored in Chapter 8, Section 8.2 and showed that contact-tracing of TB patients had a relatively small effect on the TB epidemic due to the small number of households that were visited as part of this intervention. We found that similar TB incidence levels could be expected if we visited the same number of households, but where the households were selected at random or the households had at least one HIV-positive member, and therefore none of the interventions reduced the expected TB incidence levels. When the other performance measures were considered, visiting HIV-infected households performed better than contact-tracing TB patients and contact-tracing TB patients was more effective than visiting households randomly, in terms of the number of TB cases found and averted.

The results showed that visiting households in which a member has recently progressed to late-stage HIV, was clearly the most effective intervention, although this involved visiting six times the number of households and no attempt was made to examine the efficiency or cost-effectiveness of this method.

The scenario analysis carried out in Chapter 9 explored the effect of possible changes to the underlying assumptions and design of the case-finding strategies. It allowed the relative performance of the interventions to be compared as the duration and efficacy of IPT, and the proportion of late-stage households visited was varied. It showed the potential of targeted mass preventive therapy and the benefit of administering IPT for an individual's lifetime once their infection was detected, versus shorter durations. The analysis concluded that the relative performance of the case-finding strategies was not altered by changes to various assumptions.

We have some confidence therefore, that the active case-finding method of contact-tracing TB patients will not significantly impact the TB epidemic; visiting households randomly is the least effective case-finding strategy; and contact-tracing HIV patients always performs better at finding the most TB cases and averting the most TB deaths and TB cases per 100,000 members of the population and

is therefore the most effective case-finding strategy.

3. To predict how variable population attributes are likely to effect the relative success of different interventions.

The results showed that contact-tracing HIV patients would be the most effective active-case finding strategy, compared to contact-tracing TB patients or visiting households randomly.

This is explained by the clustering of HIV and TB infections observed within households. The model was able to examine this effect more closely and found that the large majority of households containing one or more TB cases, contained HIV individuals (Chapter 8, Section 8.2.1). This relationship suggests that if TB is present in a household it is likely that HIV is too.

Analysis also showed that the majority of HIV-associated TB is due to endogenous reactivation (Chapter 8, Section 8.3.1) explaining why administering IPT to households with HIV-positive members means more cases are averted when compared to visiting the same number of random or TB-infected households. These interventions ensure HIV-positive individuals, who are at an increased risk of a latent infection reactivating, receive IPT, and therefore more cases are averted than if IPT were administered to other individuals.

Clearly, visiting households with HIV-infected members is an effective intervention for TB control firstly because this is an efficient way to find TB cases and secondly because preventive treatment is targeted at those susceptible to reactivation of a latent infection; the main cause of the HIV-associated TB epidemic.

The research suggests that the HIV status of members of a household will effect the relative success of a household intervention.

This research does not consider the realities of implementing interventions in which diagnosis of an individual with HIV leads to their household being inspected for TB disease and infection. The HIV status of an individual is a sensitive issue and many individuals with HIV would not want members of their household to know their status. Therefore, the moral and social implications of such an intervention would need to be carefully considered when it was being designed. Treating TB disease and infection at HIV clinics is the obvious substitute however, the analysis (Chapter 8, Section 8.3.2) showed that only a third of the individuals with TB disease would be found compared to if their households were also visited, and that in the majority of cases in which preventive treatment is needed, it is required by members of late-stage individual's household.

4. To determine the relative effectiveness of household interventions in controlling TB in HIV prevalent populations.

We designed the case-finding strategies so that interventions which targeted households of people with TB or HIV could be compared with an untargeted or community-wide intervention, therefore enabling the benefit of household interventions to be evaluated (Chapter 8). We made the assumption that visiting households randomly represented the effect of an untargeted or community-wide intervention where individuals volunteer themselves to attend a mobile TB clinic in their village, for example.

The analysis has shown that targeting households of TB patients (intervention 1) and HIV-positive individuals (intervention 3 & 4) is more effective than an untargeted intervention (intervention 2 & 5). More TB cases were found and more TB cases and TB deaths were averted when a targeted household intervention was used than when a community-wide intervention was explored. Targeted household interventions are therefore more effective than untargeted or community-wide interventions, assuming the same number of participants would be recruited into each.

Two results which support this conclusion can be highlighted. In Chapter 8, Figure 8.1 shows that when targeting households with at least one member with late-stage HIV, there is a significant reduction in TB incidence when compared to the untargeted intervention, despite the same number of individuals being reached. This implies that when individuals are being found at random, a much larger proportion of the population needs to be reached in order to obtain the same benefit as a targeted household intervention. Similarly, in Chapter 9, Section 9.4.1 we investigated targeting less than half the number of individuals through membership to a TB diseased or HIV-infected household, than individuals recruited randomly (Appendix N gives details of the number of households visited as part of each intervention). We find that the targeted household interventions perform the same or better, despite less than half the number of individuals being reached. This indicates that in order to observe the same benefit in a community-wide intervention as a targeted household intervention you would have to recruit more than double the number of participants. These results suggest that interventions based on finding individuals at random in the community are ineffective compared to interventions targeted at TB-diseased or HIV-infected households.

These conclusions are not sensitive to various changes to the model's assumptions and the intervention design (explored in the sensitivity analysis in Chapter 8, Section 8.4 and the scenario analysis in Chapter 9), meaning we have more as-

surance that targeted household interventions are more effective than community-wide interventions. The only exception to this is when the transmission ratio, which determines how the transmission events are allocated between household or casual contacts, was changed so that there was no household transmission of TB. The results (Chapter 8, Section 8.4.1) showed that although the household interventions became less effective if all TB transmission is assumed to be random, targeting HIV-infected households was still more efficient than a community-wide intervention yet a community-wide intervention becomes more effective than contact-tracing TB patients.

This research has relied upon the assumption that the interventions in which random households are visited can be used to approximate a community-wide intervention. This assumption is not ideal, however, to accurately reproduce the effect of a mobile TB clinic visiting villages, for example, geospatial modelling would be required. It was felt that for the purposes of this research, which investigates the relative effectiveness of *household* interventions and was not concerned with community interventions as such, visiting random households is a good approximation of a community intervention for TB control, but there is clearly room for improvement and further modelling will address this issue (discussed in Section 10.3).

We have ascertained that targeted household interventions are more effective than community-wide interventions, however the efficacy of a household intervention depends on recruiting a large number of participants (although less than would be needed if a community-wide intervention were being used). In this respect, contact-tracing TB patients is relatively ineffective because of the small number of individuals reached through this intervention. In an HIV prevalent population, a household intervention such as visiting the households of late-stage HIV members is most effective both because a larger proportion of households will be visited (due to the high prevalence of HIV) and because of the clustering of HIV and TB that occurs within households, as discussed in the evaluation of Objective 3.

In conclusion, the analysis shows that household transmission of TB is important to the design of interventions for TB control, but not as important as the clustering of TB in HIV-infected households. It is clear that designing a household intervention in which contact-tracing of HIV patients is attempted would have the biggest impact on the TB epidemic.

10.2 Limitations of the Research

This is not a cost-effectiveness analysis and we have not attempted to attach financial costs to each of the interventions. We can not draw any conclusions about the real-life applicability of our results since we would need to take into account the relative costs of the different case-finding strategies to evaluate the cost per life year saved.

Data for the validation of the model output was limited to TB incidence per 100,000 members of the population of Zimbabwe. We were not able to compare the model's output with TB incidence for the city of Harare. Data for informing the model of the yearly HIV prevalence was also for the population of Zimbabwe, as data for Harare were unavailable.

Infection with HIV is governed by a static model of HIV, which generates the number of HIV infections to be made each year and assigns the transmission events to particular age groups according to its HIV prevalence and the age distribution of sexual activity. No attempt is made to model HIV transmission so that HIV infection events are assigned to particular households (for instance, infecting the wife of a household with an HIV-positive husband), instead, HIV transmission and infection is randomly distributed throughout the households. However, due the large number of HIV transmissions occurring, household transmission of HIV does happen regularly, and so this may be more of an issue in models in which HIV prevalence and transmission is low, in which case being specific about where the transmissions occur becomes more important.

The model assumes there is no mother-to-child transmission (MTCT) of HIV. In the model, children under 10 years old are unlikely to develop TB (Section 7.3.1.1) and all children are unable to transmit TB which means that whether they are infected with HIV or not will have no impact on TB transmission. The HIV status of children is therefore irrelevant for the purposes of this model, although omitting MTCT may have important implications for the age distribution of the population. MTCT of HIV may mean that children being born infected would die before becoming adults therefore causing the age of the population to shift toward a younger distribution. If the actual population is younger than assumed by the model it would mean the model's predictions of TB incidence are an overestimate.

When modelling the interventions we assume that active case-finding is 100% effective and do not allow for imperfections in screening. We also do not address the issues surrounding possible false-negative results found when skin testing; a complication especially found in HIV-positive individuals.

During the intervention that contact-traces TB patients, households in the model are visited every time one of its members is diagnosed with TB or dies of TB. During the intervention which contact-traces individuals with late-stage HIV, households are visited each time a member advances to late-stage HIV. This means that households can be visited more than once during the intervention period and individuals found to be infected with TB may find themselves on repeated courses of IPT, a practice that does not occur in reality.

Ways in which to implement the various case-finding strategies are not given by the research, therefore we do not attempt to suggest ways in which the rate of TB patients coming forward for treatment could be doubled, nor do we suggest a viable way in which contact-tracing HIV patients could be realised.

In order to make a judgment about whether household interventions are more effective than community-wide interventions, this analysis relies on the assumption that the interventions in which random households are visited can be used to approximate a community-wide intervention. This assumption is not ideal, however, to accurately reproduce the effect of a mobile TB clinic visiting villages, for example, geospatial modelling would be required.

Further to these issues, the model does not consider:

- A contact rate for TB transmission which is dependent on an individual's HIV status.

  Currently we assume that the same contact rate applies to each individual and therefore an individual with late-stage HIV will have the same probability of infecting the same number of people with TB as an early-stage HIV or HIV-negative individual. This may be an oversimplification as you would expect those with advanced HIV and TB disease to be less mobile.

- The declining population in Zimbabwe.

  Zimbabwe has recently seen a downward spiral of the economy. Reports suggest that this is mainly attributable to the mismanagement and corruption of President Robert Mugabe's regime [179] [44] [171]. As a result, the current rate of inflation is thought to be over 66000% [34] and there is a formal unemployment rate of 80% [168]. There have also been widespread reports of starvation and the abuse of various civil and political human rights throughout Zimbabwe, particularly against those opposing the government [35]. The repressive political situation in Zimbabwe has led to a flood of refugees into neighbouring countries.

Currently we assume that the size of the population stays stable, however events that have occurred since the research began, have meant the population is declining. Operation Murambatsvina (translated as "getting rid of the filth") was initiated in May 2005 which saw homes destroyed all over the country and has left thousands homeless. It is currently estimated that 25% of the Zimbabwean population has fled abroad [169]. The impact of this population change on the efficacy of TB control methods has not been considered by this research.

10.3 Further Work

This research fits into a six-year project with funding available for a further 3 years of post-doctoral research.

Future research, being undertaken by the London School of Hygiene and Tropical Medicine, will extend the model to include geographic space, using the baseline data from Harare on geospatial position to define the structure. Transmission between households will be explored using a spatial kernel in which contact between neighbouring households is greater than between distant households. Various forms for this kernel will be explored following analysis of the baseline geospatial distribution of households and will be updated through fieldwork [51].

Geospatial modelling will also allow community-wide interventions to be explored so that strategies such as mobile TB clinics which visit certain communities, sputum collection points at schools, or door-to-door enquiry of households can themselves be evaluated and compared to household contact-tracing interventions.

The research will also incorporate a more comprehensive stratification of an individual's HIV status so that an individual can be susceptible; have a primary infection; be in incubation; be pre-AIDS; or have AIDS. The HIV status of an individual will then not only affect their TB disease progression parameters, but will also impact the number of TB transmissions they are likely to make, so that their contact rate decreases as their HIV infection advances.

The population size will not be stable and instead will incorporate birth and/or fertility rates to obtain more realistic population dynamics.

Finally, costs will be attached to the various interventions in order to carry out a full cost-effectiveness analysis and compare household case-finding strategies in terms of cost per life year saved.

Apart from using the model to investigate household interventions in HIV prevalent populations, the model could be used to address questions regarding:

- Mass Preventive Therapy

  The aim of mass preventive therapy (PT) is to administer preventive treatment to a large proportion of the population in order to rapidly reduce transmission. Questions exist as to the proportion of the population that would need to be reached to reduce transmission to below a certain threshold, and what that threshold would be. It would also be interesting to determine how long it takes for the effect of PT to wane.

- Poverty

  TB has long been a disease of poverty [185], and there is also evidence of an emerging socioeconomic gradient in HIV prevalence in Harare. Cross-sectional data on poverty is being collected through the surveys being conducted by DETECTB in Harare. The hypothesis is that clustering of TB infection and HIV infection in the poorest households may be exacerbating the biological interaction between these two infections, and may reduce the success of TB control interventions, since the poorest households may also tend to have the least access to health care. Incorporating poverty at the household level would enable us to explore what effect poverty has upon the clustering of HIV and TB at the household level, how it affects the interaction of TB and HIV, and how it affects the likely success of interventions.

- Imperfections in tuberculin skin testing (TST)

  Results from Harare show that approximately 25% of infected individuals will be anergic to TST and therefore a large number of false-negatives result from skin testing. The model could investigate the effect of administering preventive therapy to all household members regardless of their TB infection status, in order to assess whether this would be an efficient way to reach those infected individuals in need of IPT but with negative TST results.

- The duration of infectiousness

  The model could be used to examine the differences in the duration of HIV-negative and HIV-associated smear-positive TB disease; what the likely impact of the short duration of infectiousness observed in HIV-associated TB is; and how the interventions will impact the average duration of infectiousness.

- The complicating effects of smoking and alcohol consumption

  There is evidence to suggest that smoking and alcohol consumption are risk factors for TB [106] [90]. Research has shown that smokers are more likely to get infected with TB and are more likely to develop active disease [23] [109]. This may be due to it decreasing the body's immune response or damaging the protective effect of cilia in the airways [13]. The research on the risk of alcohol consumption on developing TB is less clear but there has been some research which suggests it makes individuals more vulnerable to TB, perhaps due to it weakening the body's immune system [82]. These risk factors could be considered by the model to examine whether a strategy such as tobacco control could be considered as an effective preventive intervention.

10.4 Conclusion

This is the first attempt to use individual-based modelling techniques to look at household transmission of TB. It has involved significant modelling effort and it remains unclear from this study whether the benefits of modelling in such detail will outweigh the cost of this effort. TB clinics have gone so far with case detection but can only get rates up to a certain level before they need some assistance, with HIV making it even more important that they do so. This research has shown that under some circumstances, reinforcing existing passive case-finding methods with active case-finding interventions could be an effective way to reduce TB morbidity. The research suggests that targeting households with HIV-infected members is an effective intervention for TB control due to its efficiency at finding TB cases (because of the clustering of TB disease and HIV infection in households), and because preventive treatment is targeted at those susceptible to reactivation of a latent infection; the main cause of the HIV-associated TB epidemic.

Glossary

Active Case Finding	When those with active TB are sought out.
Active Disease	Active TB Disease.
Active TB Disease	Those that have TB disease.
Anergy	A lack of reaction by the body's defence mechanisms to foreign substances.
Antenatal Clinic	A clinic specifically for pregnant woman.
Bacilli	Rod shaped bacteria.
BCG	Bacillus of Calmette and Gurin is a vaccine against tuberculosis.
Casual Contact(s)	The individual(s) that an infected person does not share a household with.
Chemoprophylaxis	The use of a chemical agent to prevent the development of a disease.
Cilia	Small hairlike structures lining the upper respiratory tract
Close Contact(s)	The individual(s) that an infected person shares a household with.

Contact Tracing	The identification and diagnosis of individuals that may have come into contact with an infected person.
Cross Sectional Data	Data collected by observing many subjects without regard for time so that the differences between the subjects can be studied.
Cultured	The growing of microorganisms in a specially prepared nutrient medium.
Deterministic	When there is no uncertainty involved.
Deterministic Compartmental Model	The system is divided into different states according to its attributes. Differential equations are then used to create flows through the various states at specified time steps. The models are deterministic and therefore each run of the model will behave identically as there are no stochastic elements to the model.
Developed Country	A country with a very high Human Development Index (HDI) and high standard of living due to an industrialised economy.
Developing Country	A country with a low to moderate Human Development Index (HDI) and relatively low standard of living due to an underdeveloped economy.
Discrete Event Simulation	A simulation made up of entities and logic statements. Entities are the tangible elements found in the real world, such as people. Logical relationships link the different entities together and these logical relationships are what define the structure of the model.

Drug Resistant TB	When the drugs used to cure TB have a reduced effect.
Endemic	When an infection is prevalent in or restricted to a population.
Endogenous	Originating from within the organism.
Epidemic	An outbreak of a contagious disease that spreads at a rate that substantially exceeds what is expected.
Epidemiological Model	Mathematical Model.
Exogenous	Originating from outside of the organism.
Extrapulmonary TB	When TB is affecting areas other than the lungs, eg. the central nervous system, lymphatic system, bones and joints.
Fast Latent	Those that have a TB infection and will progress quickly to active disease.
Fast Latent - Infectious	Those that have a TB infection and will progress quickly to infectious active disease.
Fast Latent - Non Infectious	Those that have a TB infection and will progress quickly to non infectious active disease.
HAART	Highly Active Antiretroviral Therapy. A treatment consisting of a cocktail of drugs given to suppress the growth of HIV.
HIV	Human Immunodeficiency Virus - A retrovirus that leads to AIDS by infecting helper T cells of the immune system.

HIV-Negative	Those that are HIV-Negative do not have the human immunodeficiency virus HIV.
HIV-Positive	Those that are HIV-Positive have the human immunodeficiency virus HIV.
Household Transmission	An individual with active TB infects someone that they live with.
Immune System	The immune system is the system of organs, tissues, cells, and cell products such as antibodies that protect the body against bacteria and viral infections. If the immune system is weakened the body is more likely to allow viruses to grow as it is less able to defend against pathogens.
Immunocompetent	A person who is able to produce a normal immune response.
Immunocompromised	A person who has an immunodeficiency.
Immunodeficiency	When a person's immune system and ability to fight infectious disease is suppressed or entirely absent.
Incidence	The number of new cases of a disease in a specific time interval.
Induration	The hardening of the skin, because of inflammation.
Infectious Disease	Those that have active TB disease and are able to transmit the infection. Individuals with infectious TB usually suffer from pulmonary TB.
Intervention	A way to interfere with the disease situation so as to reduce morbidity.

Isoniazid Preventive Therapy	Isoniazid is a drug given to patients with a TB infection to prevent it developing into active disease.
Late Stage HIV	The World Health Organisation defines the progression of HIV by four distinct stages. Late stage HIV implies the individual is in stages 3 or above.
Latent TB	Those that have TB infection.
Longitudinal Data	Data collected by observing subjects over time so that a subject's changes can be followed.
Mathematical Model	A static model that is fitted to data to reveal and quantify relationships and processes that would have caused the observations in the data.
Morbidity	The degree or severity of a disease.
Multidrug Resistant TB	When TB has drug resistance to more than one of the drugs used to cure TB.
Natural Cure	See Self Cure.
Non-Infectious Disease	Those that have active TB disease but are not able to transmit the infection, therefore non-infectious TB is not transmissible; for example extrapulmonary TB is usually contained and therefore cannot be expelled or spread to others.
Passive Case Finding	When those with active TB present themselves for diagnosis and treatment.
Prevalence	The ratio of the number of occurrences of a disease in a population to the number of individuals in the population at a specified time.

Preventive therapy	See Isoniazid Preventive Therapy.
Pulmonary TB	Tuberculosis that affects the lungs.
Random Transmission	An individual with active TB infects someone that they do not live with.
Reactivation	When a person has a latent infection and their immune system is weakened so that the defence against the TB infection becomes inadequate and the TB mycobacterium are able to cause damage. The latent infection becomes active TB disease.
Reinfection	When a person with a latent infection is infected again, but they do not produce an effective immune response and therefore get active disease.
Self Cure	When the body naturally cures itself from infectious or non infectious active disease.
Sputum Microscopy	A sputum specimen which has been taken from a patient, cultured, and examined for Mycobacterium tuberculosis organisms.
Sputum Smear Positive	Those that are able to produce TB bacilli that can be cultured.
Sputum Smear Negative	Those that are unable to produce TB bacilli that can be cultured.
Statistical Model	Mathematical Model.
Stochastic	When uncertainty and chance are involved.
Susceptible	Those individuals that are uninfected and therefore susceptible to infection.

TB Disease	The TB mycobacterium are present in the body and they are actively causing damage to body tissues.
TB Infection	The TB mycobacterium are present in the body but they are not actively causing damage to body tissues because the immune system has "walled them off". The infection can lie dormant for years and often only develops into "active" TB when the immune system is weakened.
TB Lesion	The localized change in a bodily organ or tissue due to TB.
Tuberculin Skin Testing	A standard method of determining whether a person is infected with TB. The TST is performed by injecting a small amount of tuberculin purified protein derivative into the forearm. The skin test reaction is measured in millimeters of the induration. DETECTB consider an induration of 5 or more millimeters to indicate a positive result.
Vaccination	Something administered to patients with the intent of conferring immunity against developing a disease.
Virulence	The degree of pathogenicity of a microbe, or the relative ability of a microbe to cause disease.
Virus	A virus is a submicroscopic parasite that often causes disease. It infects cells in biological organisms as it is unable to replicate without a host cell.

Appendix A

DETECTB Data Collection

In Chapter 2, Section 2.6.1 we discuss the household data being collected by DETECTB in Harare, Zimbabwe. The following Figures show a copy of the questionnaire used to conduct the baseline survey.

Figure A.1: DETECTB Dwelling Baseline Form

D01 BAR Dwelling Bar Code — Place Dwelling Bar Code sticker here

D02 ADD Address .....

D03 CLUS Cluster Location- Cluster No. 1= DZIVARASEKWA 2= GLENORAH 3= GLENVIEW 4= HIGHFIED 5= KAMBUZUMA 6= KUWADZANA 7 = MUFAKOSE 8 = RUGARE 9 = WARREN PARK

D04 GPSE Household GPSE East Reading 02

D05 GPSS Household GPS South Reading 80

D06 VRES Visit Result 1= COMPLETED 2= REFUSED 3= DWELLING VACANT

D07 DATE Date of interview D D M M Y Y Y Y

D08 HHDS In total, how many ***households*** reside at this dwelling?
(Pane mhuri ngani dzinogara pano, kusanganisa muridzi wemba nema lodger?)

D09 HTYP What section of the dwelling does each of these ***households*** reside in (i.e. Description)?
(Mhuri imwe neimwe inogara mupanda upi weimba (edza kutsanagudza mhuri nemhuri)?)

HH No.	Resides In Code 01= FULL MAIN HOUSE 02= PART OF MAIN HOUSE 03= COTTAGE/OUT BUILDING 04= PART OF COTTAGE/OUT BUILDING	Household ID	Surname of Head of Household or an Identifying Name (e.g. Baba va Edson)	Total Number of Adults (16 yrs +)	Total Number of Minors (<16 yrs)
1		*Place Household Bar Code sticker here*			
2		*Place Household Bar Code sticker here*			
3		*Place Household Bar Code sticker here*			
4		*Place Household Bar Code sticker here*			
5		*Place Household Bar Code sticker here*			
6		*Place Household Bar Code sticker here*			
7		*Place Household Bar Code sticker here*			
8		*Place Household Bar Code sticker here*			

D10 INF Name of Informant: .....

Figure A.2: DETECTB Household Baseline Form

HOUSEHOLD IDENTIFICATION

A household is defined as a person living alone or a group of persons living together who share meals or other essentials for living; and may be related or unrelated

(Mhuri inosanganisa munhu anogara ega kana kuti vanhu vanogara pamwechete, vachidya pamwechete zvisina basa kuti vane ukama kana kuti havana)

H01 BAR Household ID. — *Place Household Bar Code sticker here*

H02 CLUST Cluster Location-Cluster No — 1= DZIVARASEKWA, 2= GLENORAH, 3= GLENVIEW, 4= HIGHFIED, 5= KAMBUZUMA, 6= KUWADZANA, 7 = MUFAKOSE, 8 = RUGARE, 9 = WARREN PARK — -

H03 ADD Address: _____

H04 GPSE Household GPS East Reading Longitude — **02**

H05 GPSS Household GPS South Reading Latitude — **80**

H06 DATE Date of interview: — - - D D M M Y Y Y Y

The head of the household should supply the following information OR in his/her absence, a representative with sufficient knowledge about the household and permission to answer in place of the head of household.

H07 VRES ***Household*** Visit Result
1 = COMPLETED BY HEAD OF HOUSEHOLD
2 = COMPLETED BY REPRESENTATIVE
3 = CHILD HEADED HOUSEHOLD
4 = HOUSEHOLD REFUSED TO ANSWER
5 = FAILED TO INTERVIEW AFTER 3RD VISIT

H08 INTF Name of interviewee .....

H09 HSURN Surname of Head of Household Or Identifying Name.....

H10 REL Relationship to Head
1 = SELF, 2 = SPOUSE/PARTNER, 3 = CHILD, 4 = PARENT, 5 = SIBLING, 6 = OTHER RELATIVE, 7 = EMPLOYEE, 8 = NON RELATIVE

H11 PART Is the ***household*** willing to be considered for Prevalence Study?
(Mungade kupinda muchirongwa chewongororo yeTB here?)
1= YES, 2= NOT SURE, 3= NO

HOUSEHOLD STANDARD OF LIVING

H12 OWN Does the ***household*** own the dwelling?
(Mhuri yenyu ndiyo muridzi wenzvimbo ino here?)
1= OWN THE DWELLING
2= RENT THE MAIN DWELLING
3= RENT PART OF THE DWELLING/LODGER
4= USE THE DWELLING WITHOUT PAYING RENT

H13 TYP What section of the dwelling does the ***household*** reside in?
(Mhuri yenyu inogara mumupanda upi weimba?)
1= FULL MAIN HOUSE
2= PART OF MAIN HOUSE
3= COTTAGE/OUT BUILDING
4= PART OF COTTAGE/OUT BUILDING

H14 STUC What is the structure the ***household*** dwells in made of?
(Imba yamunogara yakavakwa nei?)
1= BRICK UNDER TILE
2= BRICK UNDER ASBESTOS
3= BRICK UNDER CORRUGATED IRON
4= WOOD CABIN
5= OTHER .....(SPECIFY)

H15 FOOD How often last year did you have problems satisfying the food needs of the ***household***?
(Semhuri makatambudzika kakawanda zvakadii kuti muve nechikafu chinokukwanirai mugore rakapera?)

1= NEVER
2= SELDOM
3= SOMETIMES
4= OFTEN
5= ALWAYS

H16 ROOM How many rooms is the **household** using?
(Mhuri yenyu inoshandisa mipanda mingani paimba ino?)

H17 SLEEP How many rooms does the **household** use for sleeping in?
(Mhuri yenyu inoshandisa mipanda mingani yekurara?)

H18 CROW How many people usually sleep in each of these sleeping rooms?
(Vanhu vangani vanowanzorara mumupanda wega wega?)

Fill in total number of people per room used for sleeping

Room 1	Room 2	Room 3	Room 4	Room 5

H19 HEAL If someone in the household needed medical care, how soon can the ***household*** raise clinic fees?
(Zvinokutorera nguva yakadii kuti muwane mari yekurapisa mumwe anenge arwara mumhuri menyu?

1= IMMEDIATELY
2= WAIT FOR PAYDAY
3 = HAVE TO BORROW
4 = HAVE TO BEG
5 = RELIGIOUS OBJECTION TO CLINIC USE

H20 COPE Some households cope better than others do. Compared to others with a similar income, how would you rate your household's ability to stay in control?
(Dzimwe mhuri dzinokwanisa kupfura dzimwe. Muchitarisawo dzimwe mhuri dzinowana zvakafanana nemi, munofunga kuti murikukwanisa here?)

1 = EXCELLENT
2 = GOOD
3 = FAIR
4 = POOR

TB HISTORY

H21 CTB Is anyone in your ***household*** currently being treated for TB that you can confirm or suspect?
(Pane munhu here wamunogara naye mumba menyu, ari kurapwa chirwere cheTB zvamunonyatsoziva kana kufungidzira?)

1 = YES
2 = NO

H22 CTBT How many household members are currently on TB treatment?
(Vanhu vangani varikurapwa chirwere cheTB mumhuri yenyu?

H23 TBC What have you seen to confirm that these people are on TB treatment?
(Pane chamakaona here chinoratidza kuti vanhu ava varikurapwa chirwere cheTB?)

Circle either Y or N for each person on TB treatment

	Person 1		Person 2		Person 3	
TB Card	Y	N	Y	N	Y	N
TB Tablets	Y	N	Y	N	Y	N
Going for TB Treatment	Y	N	Y	N	Y	N
Saw the Diagnosis	Y	N	Y	N	Y	N
I am the one sick	Y	N	Y	N	Y	N
None of the above	Y	N	Y	N	Y	N
None on TB Treatment	Y	N	Y	N	Y	N

H24 TBH Has anyone else in your ***household*** had TB *within the past two years* (even if deceased)? (***Probe and do not include those in H21***)
(Mumakore maviri apfuura, pane munhu wemhuri ino akamborwara nechirwere cheTB here chero akashaya?)

1 = YES
2 = NO

H27 MEMB **"Information about persons who sleep in the household regularly starting with the Head of the family"**
(Iye zvino ndinoda kubvunza zvinechekuita nevanhu vese vanomborara mumba menyu nguva zhinji)

Relationship Codes

01 = Spouse/Partner
02 = Child (biological)
03 = Child (step/adopted)
04 = Son/Daughter in law
05 = Parent
06 = Father/Mother in law
07 = Sibling
08 = Brother/Sister in law
09 = Grandchild
10 = Grandparent
11 = Uncle/aunt
12 = Nephew/Niece
13 = Cousin
14 = Servants
15 = Other Relative
16 = Non-Relative

Definitions

Usual Member - *Sleeps in household at least once a week*
Regular Visitor - *Spends at least 2 nights in household per month*
Missing -*Away in rural area, on business or at school*
First Degree Relationship - *Spouse, Child, Parent or sibling*

Line	Initials *Prioritise 1st Degree Relationship*	Sex **M =1 F =2**	Age	Usually sleep in room **(Enter in Room 9 if sleeps in another household)** 1	2	3	4	5	9	Membership Status **1** = Usual Member **2** = Regular Visitor **3** = Missing	Relationship to the Head of Household	Other Relationships Spouse of Line ... Child of line...& line... ***(Enter 99 if deceased or does not live there or is not applicable)***		
01 Head											**Self**	**Spouse of**	**Child Of**	**Comments**
02													.....&.....	
03													.....&.....	
04													.....&.....	
05													.....&.....	
06													.....&.....	
07													.....&.....	
08													.....&.....	
09													.....&.....	
10													.....&.....	
11													.....&.....	
12													.....&.....	
13													.....&.....	
14													.....&.....	
15													.....&.....	
16													.....&.....	
17													.....&.....	
18													.....&.....	
19													.....&.....	
20													.....&.....	

Appendix B

Table for Tuberculosis Modelling Literature Review

Chapter 3, Section 3.2 looks at the historical development of TB disease models and provides a concise description of the previous mathematical models which have been built to give a better understanding of tuberculosis epidemiology and effective control measures. The Table in this Appendix is a summary of these models and accompanies the discussion in Chapter 3, Section 3.2.

Key for modelling method used in the following table

Tuberculosis Modelling definition:

A mathematical model describes the behaviour of a system using mathematical language. It is the general characterization of the process, in terms of mathematics, which enables the relatively simple manipulation of variables to be accomplished in order to determine how the process would behave in different situations [178]. There are various different approaches to modelling disease and tuberculosis in particular.

The numbers/letters below are used in the Table to describe the method used in each study:

1. Deterministic Compartmental Model: In a deterministic compartmental model, the population is divided into different epidemiological groups according to their TB (and HIV) disease status. Differential equations are then used to move proportions of the population through the various groups at specified time steps.

2. Mathematical/Epidemiological Model: A mathematical, epidemiological or

statistical model is when the model is static and fitted to particular data to establish various parameter values and epidemiological relationships.

3. Other:

a) **Markov Process:**

A Markov process is when the population is divided into different epidemiological states according to their TB (and HIV) disease status. It is a stochastic process because individuals can belong to one of several states and can pass from one state to another at each time step according to fixed probabilities. A semi-Markov process implies that these probabilities of moving between states are allowed to vary over time.

b) **Decision Tree:**

A decision tree, graphs all of the possible decisions and their possible consequences, including costs, and is used to help make decisions with a specific goal in mind (usually to minimize cost).

c) **Discrete Event Simulation:**

Discrete Event Simulation is a way of building a mathematical model using computer software. The simulation holds all the concepts of the "system" and the entities themselves. The entities are the parts that move through the model, such as individuals. Logical relationships then link the different entities together and define the overall behaviour of the model. Discrete Event Simulation is stochastic as it uses random number generators to move entities through the system according to appropriate random distributions.

EARLY MODELS

Reference	Author (Year)	Population	Method	Model Details (Number of States: State Details)	Objective
[6]	Alling D (1958)	USA	3a	Six: Active TB that will remain active indefinitely, arrested TB that will remain arrested indefinitely, TB that has proved fatal, Arrested TB that is certain sometime to become active, Active TB that is certain sometime to become arrested, Active TB that is certain sometime to be fatal without ever becoming arrested. The patients are divided into two age-dependent sub-groups.	To predict the likely progression of TB disease.
[186]	Waaler H, Geser A, Anderson S (1962)	South India	1	Three: Susceptible, Infected, Infectious	To project different time trends given epidemiological situations and to show the effect of control programs, specifically BCG programs.
[36]	Brogger S (1965)	Thailand	1	Six: Uninfected, Infected, Pulmonary lesions, Cases, Vaccinated, Failures.	Heterogeneity incorporated by classifying patients by age. To model the expected effect of various control programs.

Reference	Author (Year)	Population	Method	Model Details (Number of States: State Details)	Objective
[140]	ReVelle C, Lynn W, Feldmann F (1967)	Developing Nations	1	Nine: Main states include - Susceptible, Vaccinated, Infected, Non-Active Cases, Active, Treated, Natural Recovery	To improve the economic allocation of TB control measures in developing nations.
[71]	Ferebee S (1967)	United States	2	To project the likely course of TB under current control programs and under 2 "what if" scenarios; (1) Vaccinating Uninfected with BCG, (2) Administrating Infected with Isoniazid prophylaxis.	
[182]	Waaler HT (1968a)	Non Specific (eg. Norway)	1	Eight: Non-Infected, Infected $\leq$ 5 years, Infected $>$ 5 years, BCG Protected, Active Non-infectious Cases, Active Infectious Cases, Inactive Previously non-Infectious Case, Inactive Previously Infectious Case	To provide a tool that generates future epidemiological trends in TB given various anti tuberculosis programs.
[183]	Waaler HT (1968b)	United States, Norway	1	Waaler (1968a) [182]	To use a previously developed model (Waaler 1968a) [182] to look at the effect of different BCG vaccination scenarios mainly in low prevalence countries.

Reference	Author (Year)	Population	Method	Model Details (Number of States: State Details)	Objective
[187]	Waaler H, Piot M (1969)	Europe	1	Waaler (1968a) [182]	To develop a clearer understanding of TB dynamics and the factors most likely to affect the efficacy of control methods.
[188]	Waaler HT, Piot MA (1970a)	Non Specific	1	Waaler (1968a) [182] Introduces a utility function with a social time preference parameter.	Part II of Waaler and Piot (1969). Continues to discuss the factors most likely to affect the efficacy of control methods.
[184]	Waaler HT (1970b)	Non Specific	1	Waaler (1968a) [182]	To use the model previously developed to illustrate how to answer typical questions policy makers might ask when designing a control policy.
[141]	ReVelle C, Male J (1970)	Non Specific	3b	Five: Susceptible, Inactive Disease, Pulmonary Lesions, Culture Active, Microscopy Active Cases	To determine the best pattern of testing a population whilst minimising the cost per active case treated. The aim was to maximise the number of active cases treated given the financial limitation.

Reference	Author (Year)	Population	Method	Model Details (Number of States: State Details)	Objective
[43]	Chorba W, Sanders JL (1971)	United States	1	Eight: Susceptible, New Infections, Dormant Infections- low risk, Dormant Infections- high risk, Active Cases, Natural Recoveries, Active Cases Under Treatment, Treated Recoveries. Discrete-state, discrete-time model with a cost-benefit analysis of four different prevalence situations.	To predict the prevalence of TB and predict future trends in new active cases and deaths.
[91]	Horwitz O (1973)	Denmark	1	Four: Non-Infected, Active Cases, Previous Cases, Death	To show the dynamic interplay between the disease parameters and to assess treatment regimens.
[15]	Azuma Y (1975)	Japan	1	Three: Non-Infected, BCG Vaccinated, TB-Infected	To calculate annual trends in TB prevalence and incidence.
[83]	Goh EH, Fam KL (1981)	Singapore	1	Azuma (1975) [15]	To identify the most cost effective control measure to introduce.
[172]	Trefny J, Hejdova E (1982)	Czech Socialist Republic	1	Azuma (1975) [15]	To identify the most effective control measure to introduce.

Reference	Author (Year)	Population	Method	Model Details (Number of States: State Details)	Objective
[152]	Schulzer M, Enarson D, Grzybowski S, Hong Y, Kim S, Lin T (1987)	Taiwan, Korea	3a	Five: New Sensitive, New Resistant, Old Chronic Sensitive, Old Chronic Resistant, Susceptible. Time-homogeneous Markov process, with five-yearly intervals.	To assess the major epidemiological parameters in Taiwan and Korea under a given treatment program.
[94]	Joesoef M, Remington P, Jiptoherijanto P (1989)	Indonesia	1	Seven: Non-Infected, Infected, BCG Vaccinated, Active Radiological Legions, TB positive, Healed, Death by TB	To estimate the cost-effectiveness of three control strategies.

DETERMINISTIC COMPARTMENTAL MODELS FOR EVALUATING CONTROL STRATEGIES

Reference	Author (Year)	Population	Method	Model Details (Number of States: State Details)	Objective
[31]	Blower S, Small P, Hopwell P (1996)	Non Specific	1	Five: Susceptible, Latent (Drug Sensitive/Drug Resistant), Active Disease (Drug Sensitive/Drug Resistant)	To develop a theoretical framework for designing an effective tuberculosis control program.

Reference	Author (Year)	Population	Method	Model Details (Number of States: State Details)	Objective
[37]	Castillo-Chavez C, Feng ZL (1997)	Non Specific	1	Five: Susceptible, Latent (Drug Sensitive/Drug Resistant), Active Disease (Drug Sensitive/Drug Resistant)	To determine the role that a lack of compliance plays on the maintenance of resistant TB strains.
[138]	Porco TC, Blower SM (1998)	Non Specific (Populations where there is no treatment)	1	Five: Susceptible, Latent, Infectious TB, Non Infectious TB, Recovered	To identify which input parameters from their previous models ([30], [31]) significantly affect the severity of a TB epidemic.
[28]	Blower SM, Gerberding JL (1998)	Non Specific	1	Five: Susceptible, Latent (Drug Sensitive/Drug Resistant), Active Disease (Drug Sensitive/Drug Resistant)	To use the theoretical framework developed ([30], [31]) to predict the epidemiological outcome of actual specified approaches to control.
[107]	Lietman T, Blower SM (2000)	Non Specific	1	Preexposure (5 states): Susceptible (Vaccinated/Unvaccinated), Latent (Vaccinated/Unvaccinated), Active TB Disease; Postexposure (5 states): Susceptible, Latent (Vaccinated/Unvaccinated/Waned), Active TB Disease.	To predict the epidemiological effect of both preexposure and postexposure vaccines.

Reference	Author (Year)	Population	Method	Model Details (Number of States: State Details)	Objective
[206]	Ziv E, Daley CL, Blower SM (2001)	Non Specific	1	Four: Susceptible, Latent (Early), Latent (Long Term), Active Disease	To evaluate the effect of targeting therapy specifically to persons with recently acquired latent TB.
[63]	Dye C, Williams BG (2000)	Russia, Dominican Republic, Italy, Korea, Peru	1	Nine: Uninfected, Infected, Fast Latent, Slow Latent, Active Disease (Infectious/Non Infectious), Treatment Failure, Self-Cured, Recovered	To establish a set of epidemiological criteria to eliminate MDR-TB.

DETERMINISTIC COMPARTMENTAL MODELS NOT FOR EVALUATING CONTROL STRATEGIES

Reference	Author (Year)	Population	Method	Model Details (Number of States: State Details)	Objective
[30]	Blower S, Mclean A, Porco T, Small P, Hopwell P, Sanchez M, Moss A (1995)	Non Specific	1	Three: Susceptible, Latent, Active Disease	To gain a better understanding of the intrinsic transmission dynamics of untreated TB.

Reference	Author (Year)	Population	Method	Model Details (Number of States: State Details)	Objective
[181]	Vynnycky E, Fine PEM (1997)	England, Wales	1	Eight: Uninfected, Immune, Latent (¡5 years), Latent, First Primary Episode, Re-infected, Exogenous Disease, Endogenous Disease	To consider the importance of developing TB through initial infection, reactivation or reinfection.
[69]	Feng ZL, Castillo-Chavez C, Capurro AF (2000)	Individuals with high contact rates	1	Four: Susceptible, Latent, Infectious, Recovered	To understand the effect of reinfection on TB dynamics in developing countries and inner cities of developed countries.
[68]	Feng Z, Iannelli M, Milner FA (2002)	Non Specific	1	Four: Susceptible, Infected - Drug Sensitive (Latent/Active), Infected - Drug Resistant	To investigate the effect of variable periods of latency on TB disease dynamics.
[12]	Aparicio JP, Capurro AF, Castillo-Chavez C (2002)	United States	1	Four: Susceptible, Latent (high risk/permanent), Active Disease	To provide evidence that the reductions in active TB incidence are due primarily to slower rates of disease progression.

Reference	Author (Year)	Population	Method	Model Details (Number of States: State Details)	Objective
[77]	Garcia A, Maccario J, Richardson S (1997)	Developed Countries	1	Seven: Susceptible, Protected (by BCG), Latent, Infectious, Non Infectious, Recovered, Failed Treatment	To estimate possible evolutions and trajectories of the disease.
[84]	Gomes M, Margheri A, Rebelo C (2000)	Developed Countries	1	Seven: Unprotected (Susceptible/Latent), Protected (Susceptible/Latent), Infectious, Recovered, Failed Treatment. Improved upon Garcia *et al.* (1997) [77] by including the effect of reinfection.	To estimate possible evolutions and trajectories of the disease.

OTHER TYPES OF MODEL

Reference	Author (Year)	Population	Method	Model Details (Number of States: State Details)	Objective
[164]	Sutherland I, Svandova E, Radhakrishna S (1982)	Netherlands	2		To estimate the risks of developing primary TB, exogenous TB and endogenous TB.

Reference	Author (Year)	Population	Method	Model Details (Number of States: State Details)	Objective
[147]	Salpeter EE, Salpeter SR (1998)	United States	2		To investigate the time delay between initial infection and active disease.
[11]	Aparicio JP, Capurro AF, Castillo-Chavez C (2000)	Non Specific	1	A Cluster Model - Four States: Non Infectious (Susceptible/Latent), Non Infectious but within an epidemiological active cluster (Susceptible/Latent)	To focus on the effect of long and systematic exposure of infectious individuals on susceptible individuals, the effect of clusters (groups of individuals who come into regular and close contacts with an active case).
[155]	Song BJ, Castillo-Chavez C, Aparicio JP (2002)	Non Specific	1	A Cluster Model- Five States: Non Infectious (Susceptible/Latent), Non Infectious but within an epidemiological active cluster (Susceptible/Latent), Infectious	To help understand the role of close and casual contacts in TB transmission.
[122]	Murray M (2002)	Sudan, Algeria, Netherlands	3c	Modelled individuals with various disease characteristics and social and physical space to identify the chain of TB transmission using clusters of cases sharing the same strain.	To explore the impact of different TB transmission dynamics on the population structure of isolates of TB.

Reference	Author (Year)	Population	Method	Model Details (Number of States: State Details)	Objective
[78]	Getoor L, Rhee JT, Koller D, Small P (2004)	San Francisco	2	Extended on Bayesian Networks by using statistical relational models.	To reveal relationships within richly structured data to aid understanding of TB transmission.

MODELS WHICH INCLUDE HIV

Reference	Author (Year)	Population	Method	Model Details (Number of States: State Details)	Objective
[153]	Schulzer M, Enarson D, Grybowski S (1992)	Sub Saharan Africa	2		To quantify the interaction between HIV and TB and predict the rates of TB disease incidence expected, under 4 different risk scenarios.
[25]	Bermejo A, Veeken H, Berra A (1992)	Developing Countries	2		To estimate the impact of HIV on TB incidence in developing countries.

Reference	Author (Year)	Population	Method	Model Details (Number of States: State Details)	Objective
[154]	Schulzer M, Radhamani M, Grybowski S, Mak E, Fitzgerald J (1994)	Sub Saharan Africa, Canada	2	Further developed and generalised the Schulzer *et al.* (1992) [153] model to permit application to other countries.	To study the accelerating impact of HIV infection on the incidence rates of tuberculosis disease.
[113]	Massad E, Burattini MN, Coutinho FAB, Yang HM, Raimundo SM (1993)	Non Specific	1	Sixteen: Main states include - Susceptible, Latent, Active Disease, HIV Positive (Healthy/Positive/Latent/Active Disease), AIDS (AIDS/Latent/Active Disease)	To provide a theoretical framework for the study of the interaction between, and the dynamics of, AIDS and TB co-infection.
[89]	Heymann S (1993)	Africa	3a	Ten: (HIV Negative/HIV Positive)(Susceptible/Recently Infected with TB/Carrying an old asymptomatic TB infection/Active Disease/Dead)	To understand the effect of the interaction of TB and HIV; the impact of expanding TB programs; and the impact of expanding chemoprophylaxis programs on TB incidence and mortality.

Reference	Author (Year)	Population	Method	Model Details (Number of States: State Details)	Objective
[33]	Brewer T, Heymann S, Colditz G, Wilson M, Auerbach K, Kane D, Fineberg H (1996)	United States	3a	3 age groups each with 18 clinical states: (HIV Negative/HIV Positive/AIDS)(No Infection/Low Risk for Active Disease/Intermediate Risk for Active Disease/High Risk for Active Disease/Active Drug-Sensitive Disease/Active Drug-Resistant Disease)	To examine the effects of TB control strategies on projected US cases and deaths.
[139]	Porco TC, Small PM, Blower SM (2001)	United States	3c	Thirty: Main states include - (Susceptible/Infected/Active Disease)(HIV Negative/HIV Stage I/StageII/StageIII/StageIV)	To quantify the potential impact of HIV on TB incidence and the implications on global TB control.

Reference	Author (Year)	Population	Method	Model Details (Number of States: State Details)	Objective
[120]	Murray CJL, Salomon JA (1998a)	Five World Regions (according to patterns of TB epidemiology)	1	HIV Negative/Positive Population, each with 19 States: Main states include- Uninfected, Infected (Slow Breakdown/Fast Breakdown), Superinfected, Active Disease (Untreated/Treated)(Smear Positive Pulmonary/Smear Negative Pulmonary/Extra Pulmonary)(Fast Diagnosis/Slow Diagnosis), Recovered (Fast Relapse/Slow Relapse). Elaborated on Blower *et al.* (1995) and Blower *et al.* (1996) ([30] [31]) to capture other aspects of the TB epidemic like the impact of HIV on TB's development.	To assess the efficacy of using the WHO recommended DOTS strategy alone to control TB.
[121]	Murray CJL, Salomon JA (1998b)	Five World Regions (according to patterns of TB epidemiology)	1	Murray and Salomon (1998a) [120]	To incorporate a cost-benefit analysis into the previously discussed model (Murray and Salomon (1998a) [120]).

Reference	Author (Year)	Population	Method	Model Details (Number of States: State Details)	Objective
[61]	Dye C, Garnett GP, Sleeman A, Williams BG (1998)	Six WHO Regions	1	Similar to Murray and Salomon (1998a) [120] but incorporates age structure.	To assess the potential impact of DOTS on TB control.
[55]	Currie CSM, Williams BG, Cheng RCH, Dye C (2003)	Kenya, Uganda, South Africa	1	Ten: HIV Negative, Stage I and Stage II individuals/HIV Late Stage individuals: Uninfected, Latent, Active Disease, Failed Treatment, Preventive Therapy. The model from Dye *et al.* (1998) [61] was reduced to a single age class but the options for TB control were extended by including three preventative methods as well as case detection and cure.	Due to the observation that even good DOTS programs were failing to check the rapid increase in TB cases in countries with a high HIV prevalence; this model was stimulated by the need to search for new ways to manage TB epidemics.
[56]	Currie CSM, Williams BG, Corbett EL (2005)	Kenya	1	Fourteen: HIV Negative, Stage I and Stage II individuals/HIV Late Stage individuals: Susceptible, Fast Progression to Active Infectious Disease, Latent, Infectious, Failed Treatment, Treatment, Recovered	To investigate the effect on the HIV-associated TB epidemic of rapid progression and high mortality due to the duration of infectiousness for TB of HIV-positives with active TB disease.

Reference	Author (Year)	Population	Method	Model Details (Number of States: State Details)	Objective
[197]	Williams BG, Granich R, Chauhan LS, Dharmshaktu NS, Dye C (2005)	India	1	Forty: (HIV Negative/HIV Stage I/HIV Stage II/HIV Stage III/HIV Stage IV):Susceptible, Latent, Infectious (Infectious/Failed Treatment/Self Cured), Non Infectious (Non Infectious/Failed Treatment/Self Cured)	To determine whether the HIV epidemic will stop India's Revised National TB Control Program of DOTS from achieving its millennium goals.

Table B.5: Summary of tuberculosis models featured in the literature review (Chapter 3)

Appendix C

The Districts of Kenya

In Chapter 4, we describe a simple parametric model which was fitted to district level data from Kenya on TB notification rates and HIV prevalence among women attending ante-natal clinics. Over the 18 years for which TB data are available for this study there have been a number of changes to the district boundaries in Kenya as discussed in Section 4.3.1. This Table outlines those changes and shows the provinces of Kenya (column 1) and the district names between 1985 and 1990. In 1990, for example, Kericho split into Bomet and Kericho. The two districts were again divided in 2000 with part of Bomet and part of Kericho combining to form a new district, Buret. Districts marked in red began reporting on TB separately in 1991, those marked in green in 1993, and those marked in blue in 1995.

	1985–1990	1991–2000	2000+
Nairobi	Nairobi	Nairobi	Nairobi
Central	Kiambu	Kiambu	Kiambu
		Thika	Thika
	Muranga	Muranga	Maragua
			Muranga
	Kirinyaga	Kirinyaga	Kirinyaga
	Nyeri	Nyeri	Nyeri
	Nyandurua	Nyandurua	Nyandurua
Coastal	Kilifi	Kilifi	Malindi
			Kilifi
	Kwale	Kwale	Kwale
	Lamu	Lamu	Lamu
	Mombassa	Mombassa	Mombasa
	Taita Taveta	Taita Taveta	Taita Taveta
	Tana River	Tana River	Tana River
Eastern	Embu	Embu	Mbeere
			Embu
	Isiolo	Isiolo	Isiolo
	Kitui	Mwingi	Mwingi
		Kitui	Kitui
	Machakos	Makueni	Makueni
		Machakos	Machakos
	Marsabit	Marsabit	Marsabit
			Moyale
	Meru	Nyambene	Meru North
		Meru Central	Meru Central
		Tharaka Nithi	Meru South
			Tharaka
Western	Bungoma	Bungoma	Mt. Elgon
			Bungoma
	Busia	Busia	Busia
			Teso
	Kakamega	Vihiga	Vihiga
		Kakamega	Lugari
			Butere
			Mumias
			Kakamega

	1985–1990	1991–2000	2000+
North Eastern	Mandera	Mandera	Mandera
	Garissa	Garissa	Garissa
			Ijara
	Wajir	Wajir	Wajir
Nyanza	Kisumu	Kisumu	Kisumu
			Nyando
	Kisii	Nyamira	Nyamira
		Kisii	Gucha
			Kisii
	Siaya	Siaya	Bondo
			Siaya
	South Nyanza	Homa Bay	Suba
			Rachhuonyo
			Homa Bay
		Migori	Kuria
			Migori
Rift Valley Province	Nakuru	Nakuru	Nakuru
	Kajaido	Kajaido	Kajaido
	Kericho	Bomet	Bomet
			Buret
		Kericho	
			Kericho
	Laikpia	Laikpia	Laikpia
	Samburu	Samburu	Samburu
	Narok	Narok	Transmara
			Narok
	Baringo	Baringo	Koibatek
			Baringo
	Elgeyo Marakwet	Elgeyo Marakwet	Keiyo
			Elgeyo Marakwet
	Nandi	Nandi	Nandi
	Trans Nzoia	Trans Nzoia	Trans Nzoia
	Turkana	Turkana	Turkana
	Uasin Gishu	Uasin Gishu	Uasin Gishu
	West Pokot	West Pokot	West Pokot

Appendix D

Maximum Likelihood Estimation

Throughout the simulation's development, there was a need to fit various simple models to observed data. Mathematical models were proposed, so that they could be used to describe and represent the data in the simulation. To obtain the best estimates of the parameters of these models, we used maximum likelihood estimation.

D.1 The Likelihood Function

The likelihood function is used to generate estimates of the parameter values of a proposed model.

The likelihood function is described well in Kendall's Advanced Theory of Statistics [158].

Let $X = (X_1, ..., X_n)$ be a random vector and $\{f_x(x|\theta) : \theta \in \Theta\}$ a statistical model parameterised by $\theta = (\theta_1, ...\theta_k)$, the parameter vector in the parameter space Θ. The Likelihood function is a map $L : \Theta \rightarrow R$ given by $L(\Theta|x) = f_x(x|\Theta)$. In other words, the likelihood function is functionally the same in form as a probability density function (PDF). However, the emphasis is changed from the x to the θ:

- The PDF is a function of the x's while holding the parameters (θ) constant. For example, the input parameters for the proposed model are known and the observed data are unknown.
- The Likelihood is a function of the parameters (θ), while holding the x's constant. For example, the observed data is known but the input parameters for the proposed model are unknown.

D.2 Maximum Likelihood Estimation

By finding its maximum, the likelihood function $L(\theta)^*$ can be used to estimate the value of the unknown input parameters. Planet Math [136] explain that the parameter vector $\hat{\theta}$, such that $L(\hat{\theta}) = L(\theta)$ for all $\theta \in \Theta$, is called the maximum likelihood estimator (MLE) of θ. Often the density function is the product of independent components. It is then usually easier to compute the MLE of a likelihood function by taking logs first and computing the maximum of the (natural) log of L.

D.3 An Example

As an example, we describe how this method was used to find estimates of the input parameters when fitting a double logistic equation to HIV prevalence data for Zimbabwe (Chapter 6, Section 6.5 and Appendix K). This was done in order to obtain HIV prevalence estimates for the simulation between 1980 and 2030. The expression for a double logistic equation is

$$p(t) = \frac{e^{\alpha(t-\tilde{t})}}{1+e^{\alpha(t-\tilde{t})}}\left(\frac{ae^{-\beta(t-\tilde{t})}}{1+e^{-\beta(t-\tilde{t})}}+b\right) \tag{D.1}$$

The output from this proposed HIV prevalence model can be considered as a function of its parameters $\theta = (a, b, \alpha, \beta, \tilde{t})$. The HIV data observations are treated as fixed and the HIV parameters (θ) are treated as unknown as per the MLE method.

The observed HIV data can be described by the equation

$$P_i = P(t_i|\theta) + \epsilon_i \qquad i = 1, ...n \tag{D.2}$$

where P is the double logistic function, n is the number of HIV observations and ϵ_i are the HIV errors - the difference between the model output and the observed data point at i. It is assumed that the errors are normally distributed with means of 0 and standard deviation σ. Therefore

$$P_i = P(t_i|a, b, \alpha, \beta, \tilde{t}) + \epsilon_i \tag{D.3}$$

where $\epsilon_i \sim N(0, \sigma^2)$ and $\epsilon_i \cong [P_i - P(t_i|a, b, \alpha, \beta, \tilde{t})]$. The likelihood function is the product of the density of the errors.

$$L(\theta) = \prod_{i=1}^{n} \varphi(\epsilon_i) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} exp\left\{\frac{-1}{2\sigma^2}[P_i - P(t_i|a,b,\alpha,\beta,\tilde{t})]^2\right\} \quad \text{(D.4)}$$

In order to find the maximum of this function it is sufficient to find these values at the maximum of the log likelihood function. Taking natural logarithms of equation D.4 gives the log likelihood function.

$$loglik(\theta) = \frac{-n}{2}\log(2\pi) - n\log\sigma - \frac{1}{2\sigma^2}\sum_{i=1}^{n}[P_i - P(t_i|a,b,\alpha,\beta,\tilde{t})]^2 \quad \text{(D.5)}$$

In order to find the values of the unknown parameters that produced the maximum value of the log likelihood function, the Nelder-Mead algorithm was used. The Nelder-Mead algorithm [125] can be described as a simplex based direct search method used to find the minimum (equivalent to the negative of the maximum) of an n-dimensional function. It is based on evaluating a function at the vertices of a simplex, then iteratively modifying the simplex to find better points. This method was selected as it is robust and needs only function evaluations without need for derivatives, and due to its success in previous studies [55]. For more trivial models, such as a Poisson distribution proposed to describe the distribution of household size in Harare (Chapter 6, Section 6.3.1), the maximum of the log likelihood function was found using Solver, the optimisation software found in Microsoft Office Excel.

Appendix E

Bootstrap Resampling Methods

Bootstrap resampling is a simple and effective way of studying the distributional properties of statistical data [41].

The basic process of constructing a statistic is to draw a sample $\boldsymbol{Y} = (Y_1, Y_2, ..., Y_n)$ of size n from a distribution $F(y)$, and then calculate the statistic of interest T from $\boldsymbol{Y}$. The problem is then to find the distribution of T, which bootstrapping allows us to do.

The bootstrap method is based on the idea that as we do not know the distribution $F(y)$ we use the best available estimate, which is the empirical distribution function (EDF) of our sample $\boldsymbol{Y}$. Instead of sampling from $F(y)$, we sample from the EDF of $\boldsymbol{Y}$, i.e. we sample with replacement from $\boldsymbol{Y}$, and carry out this process B times to get B bootstrap samples $\boldsymbol{Y_1^*}, \boldsymbol{Y_2^*}, ..., \boldsymbol{Y_B^*}$. From each of these bootstrap samples we can calculate the statistic of interest $T_k^* = T(\boldsymbol{Y_k^*})$. If we order these statistics it will give us an estimate of the distribution of T.

We generate bootstrap samples by resampling with replacement from the original sample and calculating the bootstrap T for each of the samples. The pseudocode taken from Cheng [41] is as follows:

// $\boldsymbol{y} = (y(1), y(2), ..., y(n))$ is the original sample
// $T = T(\boldsymbol{y})$ is the calculation that produced T from $\boldsymbol{y}$
For $k = 1$ to B
{
  For $i = 1$ to n
  {
    j = Int$(1 + n$ x $Unif())$ //$Unif()$ returns a uniformly distributed
                                   //$U(0,1)$ variate each time it is called
    $y^*(i) = y(j)$
  }
  $T^*(k) = T(\boldsymbol{y}^*)$

}

Now suppose we have fitted a parametric model to data. If this parametric model is correct and accurately describes the form of the data, then the fitted parametric model will be a close representation of the unknown true parametric model. We can therefore generate bootstrap samples not by resampling from the original data, but by sampling from the fitted parametric model [41]. This process is called *the parametric bootstrap.*

E.1 Applying the Bootstrapping Method

The bootstrapping method discussed here which we apply in Chapter 4 to obtain confidence intervals on the fitted HIV prevalence and TB incidence curves, is the parametric bootstrapping process.

Suppose $\eta(t, \boldsymbol{\theta})$ represents the incidence of TB per 100,000 members of the population in year t. We assume that there is no information on $\boldsymbol{\theta}$, but there are observations y_i of the TB incidence at given time points $t_i \quad i = 1, 2, ..., n$. These are subject to errors and thus

$$y_i = \eta(t_i, \boldsymbol{\theta}) + \epsilon_i \qquad i = 1, 2, ..., n. \tag{E.1}$$

We fit $\boldsymbol{\theta}$ to the observations y_i using maximum likelihood estimation and generate bootstrap samples ($\boldsymbol{Y}^*$). The bootstrap samples are obtained by using equation E.1 to construct the sample $\boldsymbol{Y}^*$, only with the unknown $\eta(t, \boldsymbol{\theta})$ replaced by $\eta(t, \hat{\boldsymbol{\theta}})$ and with ϵ sampled from the (fitted) error distribution, $N(0, \hat{\sigma^2})$, assuming normally distributed error.

In our model therefore, instead of the original sample coming from $F(y)$, it comes from a parametric representation of the distribution $F(y, \hat{\boldsymbol{\theta}})$ and the statistic of interest T is the maximum likelihood estimates $\hat{\boldsymbol{\theta}}$.

As each bootstrap sample $\boldsymbol{Y}^*$ is generated, maximum likelihood estimation is used to ascertain $\hat{\boldsymbol{\theta}}^*$. Therefore, generating bootstrap samples from the parametric model enables us to evaluate the distribution of $\hat{\boldsymbol{\theta}}$, and therefore to obtain confidence intervals on each parameter.

Confidence intervals on the TB incidence curve are obtained by using each bootstrap $\hat{\boldsymbol{\theta}}_k^*$ to generate B approximations of the TB incidence curve $\hat{\boldsymbol{Y}}_k$ using

$$\hat{\boldsymbol{Y}}_k = \eta(t, \hat{\boldsymbol{\theta}}_k^*) \qquad k = 1, 2, ..., B. \tag{E.2}$$

These approximations give a distribution of the possible TB incidence at each time point t, and therefore we can obtain confidence intervals for the entire series of time points, providing us with confidence intervals on the TB incidence curve.

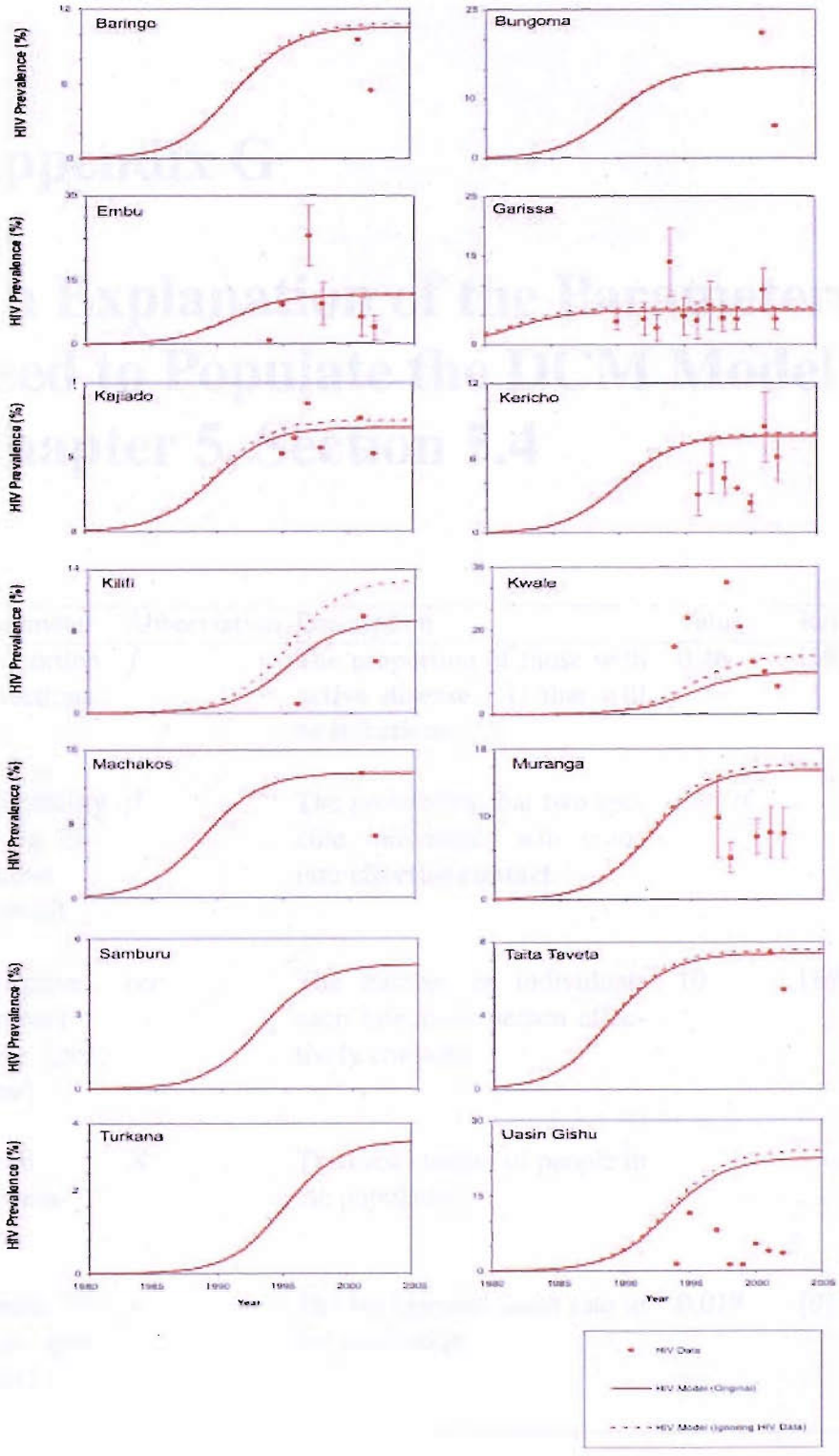
These methods were implemented using Microsoft Visual Basic in Excel and were applied to each district of Kenya in order to obtain confidence intervals for both the TB incidence and HIV prevalence curve. The equation for $\eta(t, \boldsymbol{\theta})$ used to evaluate TB incidence is given by Equation 4.4 in Chapter 4 and the equation for $\eta(t, \boldsymbol{\theta})$ used to evaluate HIV prevalence is given by Equation 4.1 in Chapter 4.

The confidence intervals on the TB incidence and HIV prevalence predictions for each district of Kenya are given in Chapter 4, Figures 4.2, 4.3 & 4.4 and the confidence intervals on each of the parameters are given in Chapter 4, Tables 4.1, 4.2, 4.3 & 4.4).

Appendix F

Comparing Results for all Type II Districts

In Chapter 4, we describe a simple parametric model which was fitted to district level data from Kenya on TB notification rates and HIV prevalence among women attending ante-natal clinics. The model predicts the HIV prevalence in districts using just TB incidence data. Figure F shows the model's predictions of HIV prevalence in the Type II districts of Kenya (Type II districts are those districts where HIV data have only been measured at a few points in time). The Figure compares the model's predictions when the fitting procedure uses just the TB incidence data and then both the TB data and the available HIV data.


Baringo
Bungoma
Embu
Garissa
Kajiado
Kericho
Kilifi
Kwale
Machakos
Muranga
Samburu
Taita Taveta
Turkana
Uasin Gishu
HIV Prevalence (%)
Year
HIV Data
HIV Model (Original)
HIV Model (Ignoring HIV Data)

Appendix G

An Explanation of the Parameters used to Populate the DCM Model in Chapter 5, Section 5.4

Parameter	Abbreviation	Description	Value	References
Proportion Infectious	f	The proportion of those with **active disease** (A) that will be **infectious** (I).	0.46	[197]
Probability of an Effective Contact	β	The probability that two specific individuals will come into **effective contact**.	ecr/N	
Effective Contact Rate (per year)	ecr	The number of individuals each infectious person effectively contacts.	10	[160]
Total Population	N	The total number of people in the population.		
Death Rate (per year)	μ	The background death rate in the population.	0.018	[9]

Parameter	Abbreviation	Description	Value	References
Disease Induced Death Rate (per year)	d	The death rate from active disease.	0.3	[127] [128] [5] [119] [135] [192]
Disease Induced Death Rate - Infectious (per year)	d_I	The death rate from infectious active disease.	0.3	[197]
Disease Induced Death Rate -Non Infectious (per year)	d_{NI}	The death rate from non infectious active disease.	0.2	[197]
Number of Births (per year)	Λ	The total number of births into the population.	Chapter 5, Section 5.4.1	
Proportion Develop Primary Active TB	p	The proportion of those infected that will develop **active disease** (A).	0.14	[163] [72] [76] [164] [159] [102] [103] [180] [181]
Proportion of Reinfections Susceptible	x	The proportion of reinfections that are susceptible to developing **active disease** (A).	0.35	[163] [164] [180] [181]

Parameter	Abbreviation	Description	Value	References
Reactivation Rate (per year)	v	The rate at which individual's **latent** (E) infection becomes active.	0.001	[92] [20] [163] [161] [180] [181]
Diagnosis Rate (per year)	ϕ	The rate at which those with **active disease** (A) will be diagnosed.	0.5	[56]
Diagnosis Rate (per year)	ϕ_I	The rate at which those with **infectious** (I) **active disease** (A) will be diagnosed.	0.5	
Diagnosis Rate (per year)	ϕ_{NI}	The rate at which those with **non infectious** (I) **active disease** (A) will be diagnosed.	0.5	
Cure Rate	cr	The proportion of treated individuals who are successfully treated.	0.7	[56]
Treatment Duration (years)	td	The length of **treatment** (T) for **active disease** (A). How long individuals stay in the **treatment** (T) class.	0.5	[56]
Progression Rate (per year)	r	The rate at which individuals develop **active disease** (A) from the time of infection.	0.62	[181]
Natural Cure Rate - Infectious (per year)	scr_I	The rate at which **infectious** (I) individuals **self cure**.	0.4	[197]

Parameter	Abbreviation	Description	Value	References
Natural Cure Rate - Non Infectious (per year)	scr_{NI}	The rate at which **non infectious** (NI) individuals **self cure**.	0.2	[197]
Relapse Rate (per year)	s	The rate at which individuals that self cured, revert to **active disease** (A).	0.21	[197]
Conversion Rate (per year)	n	The rate at which **non infectious** (NI) individuals convert to being **infectious** (I).	0.015	[197]

Table G.1: An Explanation of the Parameters used to Populate the DCM Model in Chapter 5, Section 5.4

Appendix H

Detailed Simulation Documentation

In Chapter 6, Section 6.2 we discuss the design of the discrete event simulation. This Appendix provides full documentation of the simulation model.

H.1 Class

This Section provides a description of the classes and structures implemented within the DES model.

Class Name	Description
Baby	Makes an instruction (H.3.4) for the simulation on the allocation of a new individual into the appropriate household.
CreateIndividual	Creates a new individual given an instruction from the simulation and ascertains its attributes.
DataFile	Reads in the various input data at the beginning of the simulation. Holds the structures containing the input data and the warmed up populations and schedules (Chapter 6, Section 6.4).
Death	Kills the individual and updates the associated structures in the simulation.

Class Name	Description
EventFunctions	Contains all the functions which execute the events, listed in Section H.2.
Factor	Sets and adjusts all the parameters in the model for use in the sensitivity analysis.
Generate	Generates the population and households at the beginning of the simulation. Generates the household and random transmission events during the simulation.
HIV	Calculates yearly age specific HIV incidence figures; schedules the correct number of HIV transmission events throughout the age groups; and executes the HIV transmission events during the simulation. Records the model's HIV incidence and prevalence figures and holds them in its structures.
Household	Defines a house (Section H.3.3).
HouseholdList	Holds and manages the structures containing the houses and the houses according to size. Adds and removes individuals from households.
Instructions	Defines an instruction (Section H.3.4).
Intervention	Holds all the functions to apply different interventions to the simulation and contains structures to keep track of intervention details and efficacy.
MiscFunctions	Contains different simple functions used throughout the simulation.
Person	Defines an individual (Section H.3.1).

Class Name	Description
Population	Manages and holds structures which contain a list of the population, individuals who are HIV-positive and individuals according to age group.
RandomNumber	Generates random numbers using a combination of the Mersenne Twister and RanRot generators.
ReadInData	Reads in comma separated files of input data and warmed up population and schedule sets for use in the simulation.
rngNonUniform	Implements random number generators from 10 standard probability distributions.
rngUniform	Defines a number of uniform random number generators including the Mersenne Twister and RanRot.
Schedule	Contains, manages and schedules the events on the activity list.
ScheduledEvent	Defines an event (Section H.3.2).
Simulation	Contains the main simulation function used to work through the activity list, execute events and record simulation data. Also contains the simulation clock.
StateDetails	Contains and manages the structure which records the number of people in each TB state.
StatisticalDistributions	Contains the Beta, Poisson, Normal and Gamma distributions.
TB	Contains and manages the structure holding the model's TB incidence and the input TB incidence data.

Class Name	Description
TimeOf	Uses statistical distributions, parameter values and data to calculate the time of various events.
Transmission	Chooses a person to infect with TB or HIV and calls the transmission process.
Treatment	Function to decide whether a person will fail or succeed treatment.
WriteOutData	Writes out simulation information to comma separated files.

Table H.1: Descriptions of classes implemented within the model

H.2 Events

An event is an action upon an individual within the simulation. The following Section lists all the events that are active in the simulation and explains what effect they have.

Event Name	Description	Event(s) it Generates
BecomeLateStage	Progresses an individual from an HIV-positive status to a late stage HIV status. If the individual has latent TB, their time of reactivation is recalculated.	StartActiveDisease
BecomePositiveStage	Progresses an individual from an early-stage HIV status to an HIV-positive status.	

Event Name	Description	Event(s) it Generates
Conversion	Converts a person with non infectious TB disease to infectious TB disease.	StartInfectiousDisease
DeathFromHIV	Kills the individual and creates a new individual in the simulation.	
DiseaseInducedDeath	Kills the individual and creates a new individual in the simulation.	
FinishPT	Changes an individual's TB status from being on treatment to having a latent TB infection.	
HIVTransmission	Randomly chooses an individual from a given age group to infect with HIV.	
HTransmission	Transmits TB to someone within the individual's household (or randomly if no household members can be infected.	RTransmission
NaturalDeath	Kills the individual and creates a new individual in the simulation.	
RTransmission	Transmits TB to someone randomly in the population.	StartLatent
SelfCure	Changes an individual's TB status from an active disease status to self-cured. Determines if and when the individual will relapse to active disease.	StartInfectiousDisease StartNonInfectiousDisease

Event Name	Description	Event(s) it Generates
StartActiveDisease	Decides whether an individual will get infectious or non infectious disease.	StartInfectiousDisease StartNonInfectiousDisease
StartInfectiousDisease	Changes an individual's TB status to an infectious disease status. Calculates when and how the individual will leave the state (receives treatment, self cures or dies) and therefore determines their length of infectiousness. Generates and schedules the individual's household and random transmission events.	StartTreatment SelfCure HTransmission RTransmission DiseaseInducedDeath
StartLatent	Infects or reinfects an individual with TB and sets their TB status to latent. Calculates if and when the individual will progress to active disease.	StartActiveDisease
StartNonInfectious Disease	Changes an individual's TB status to a non infectious disease status. Calculates when and how the individual will leave the state (converts to infectious disease, receives treatment, self cures or dies).	Conversion StartTreatment SelfCure DiseaseInducedDeath

Event Name	Description	Event(s) it Generates
StartPT	Changes an individual's TB status to being on treatment. Administers preventative therapy to the individual and schedules when the person will finish preventative therapy.	FinishPT
StartRecovered	Changes an individual's TB status from being on treatment to being recovered.	
StartTreatment	Changes an individual's TB status to being on treatment. Decides if and when the individual will fail or succeed treatment and schedules the event.	StartRecovered StartInfectiousDisease StartNonInfectiousDisease StartActiveDisease

Table H.2: Events acting on individuals in the simulation

H.3 Objects

The simulation contains four main objects which each have their own member data and functions. This Section details the objects and documents their purpose and members.

H.3.1 An Individual

Individuals make up the model's population and are the entities of the model. The processes associated with the model therefore act upon these individuals. Individual's hold details about both their personal and disease characteristics.

Person Class: Individual
Member Data

Person Class: Individual	
itsID	Holds an individual's unique identification number
itsGender	Holds the gender of the individual
itsAge	Holds the age of the individual
itsHousehold	Holds the unique identification number of the house it lives in
itsTBStatus	Holds the TB infection status of the individual
itsHIVStatus	Holds the HIV infection status of the individual
itsTimeOfDeath	Holds the time that the individual is due to die. When a person is first created, this will be the time of natural death but disease processes may bring this forward and therefore a person can also die from TB or HIV.
itsTimeOfActiveDisease	Holds the time that an individual with latent TB is due to get active disease
itsTimeOfGettingHIV	Holds the time that an individual was infected with HIV
itsAlive	Keeps track of whether the person is alive or has died
Member Functions	
Set/Get ID	Assigns and retrieves an individual's unique identification number
Set/Get Gender	Assigns and retrieves an individual's gender

Person Class: Individual	
Set/Get/Increase Age	Assigns, retrieves and increments an individual's age
Set/Get Household	Assigns and retrieves the ID of the individual's house
Set/Get/Change TBStatus	Assigns, retrieves and updates an individual's TB infection status
Set/Get HIVStatus	Assigns and retrieves an individual's HIV infection status
Set/Get TimeOfDeath	Assigns and retrieves the time an individual is due to die
Set/Get TimeOfActiveDisease	Assigns and retrieves the time a latent individual is due to get active disease
Set/Get TimeOfHIVInfection	Assigns and retrieves the time an individual is infected with HIV
Kill	Kills the individual and removes it from its house. The object itself is not destroyed but recycled
IsAlive	Returns true if the person is alive and false if the person has died

Table H.3: Details of an individual object from the person class

H.3.2 An Event

An event is an action upon an individual in the simulation. A full list of the events active within the model is given in Section H.2. An event holds information about which event function to execute, the name of the event, the ID of the individual the event involves and the time of the event.

ScheduledEvent Class: Event

ScheduledEvent Class: Event	
Member Data	
itsEvent	Holds a pointer to the function which executes the actual event
itsIDofPerson	Holds the identification number of the individual the event acts on
itsTimeofEvent	Holds the time the event is due to be executed
itsNameofEvent	Holds the name of the event
Member Functions	
Set/Get Event	Assigns and retrieves the pointer to the event function
Set/Get IDofPerson	Assigns and retrieves the identification number of the individual that the event acts on
Set/Get TimeofEvent	Assigns and retrieves the time of the event
Set/Get NameofEvent	Assigns and retrieves the name of the event

Table H.4: Details of an event object from the scheduledevent class

H.3.3 A House

A house is an object to which a number of individuals belong. Each individual will belong to a defined house so that transmission of TB can work at both the community and household level. A household contains information about its ID, its size and which individuals live in it.

Household Class: House
Member Data

Household Class: House	
itsID	Holds a household's unique identification number
itsHouseholdSize	Holds the number of occupants within the household
itsOccupants	Holds a list containing the identification numbers of each of the individuals within the household
Member Functions	
Set/Get ID	Assigns a given identification (ID) number to the household and retrieves it
GetHouseholdSize	Retrieves the number of occupants within the household
AddOccupant	Adds an individual to the household: Adds an individual's identification (ID) number to the "itsOccupants" list
DeleteOccupant	Removes an individual from the household: Deletes the individual's ID number from the "itsOccupants" list
GetOccupants	Returns the "itsOccupants" list: Returns a list containing the ID numbers of each of the individuals with the household
GetFellowOccupants	Returns, given the ID number of an occupant, a list containing the ID numbers of the individuals the occupant lives with
ContainsAdult	Returns true if the household contains an adult and false if all the occupants are children

Table H.5: Details of a house object from the household class

H.3.4 An Instruction

An instruction is an object containing information for the simulation on creating a new individual. Each instruction tells the simulation the ID of the new individual, the ID of the new individual's household and whether the individual is to be a baby or an adult. For full details of how the attributes of each instruction are determined, see Chapter 6, Section 6.3.4.

Instructions Class: Instruction	
Member Data	
itsPersonID	Instructs the simulation the ID of the new individual
itsHouseholdID	Instructs the simulation of the household ID the new individual will be assigned to
itsKind	Instructs the simulation whether to create an adult or a child
Member Functions	
Set/Get PersonID	Assigns and retrieves the ID of the individual to be created
Set/Get HouseholdID	Assigns and retrieves the ID of the house the individual is to be assigned to
Set/Get Kind	Assigns and retrieves whether the new individual is to be an adult or a child

Table H.6: Details of an instruction object from the instructions Class

Appendix I

2000 WHO Life Tables for Zimbabwe

In Chapter 6, Section 6.3.3 we describe how the 2000 WHO life tables for Zimbabwe were used to help determine the attributes of an individual when created in the simulation. The life tables are given in this Appendix. The definitions of the life table's column headings are as follows:

Column Heading	Definition
nMx	Age-specific death rate
nqx	The probabilities (or risks) of dying between ages x and $x+n$
lx	The number of people surviving to the beginning of age interval
ndx	The number of people dying during the age interval
nLx	The number of years lived between ages x and $x+n$
Tx	The total number of years lived after exact age x
ex	The average number of years of life remaining at exact age x

Life table: Zimbabwe 2000

Sex	Age range	Actual population	Actual deaths	nMx	nqx	lx	ndx	nLx	Tx	ex
both	'<1'	417800	33257	0.0796	0.0754	100000	7540	94722	3771400	37.7
both	'1-4'	1633560	20702	0.01267	0.0492	92460	4549	358924	3676678	39.8
both	'5-9'	1920810	4218	0.0022	0.01092	87911	960	437157	3317754	37.7
both	'10-14'	1736780	2833	0.00163	0.00812	86951	706	432992	2880597	33.1
both	'15-19'	1509400	9750	0.00646	0.03179	86245	2741	424373	2447605	28.4
both	'20-24'	1243080	23063	0.01855	0.08865	83504	7403	399012	2023232	24.2
both	'25-29'	952270	35383	0.03716	0.16999	76101	12936	348164	1624220	21.3
both	'30-34'	710540	37139	0.05227	0.23114	63165	14600	279323	1276056	20.2
both	'35-39'	570470	28608	0.05015	0.22281	48565	10821	215772	996732	20.5
both	'40-44'	474940	21650	0.04558	0.2046	37744	7723	169414	780960	20.7
both	'45-49'	372090	14746	0.03963	0.18029	30022	5413	136576	611545	20.4
both	'50-54'	255280	8862	0.03471	0.15971	24609	3930	113219	474969	19.3
both	'55-59'	237180	7197	0.03035	0.14103	20679	2916	96103	361750	17.5
both	'60-64'	189230	6449	0.03408	0.15703	17762	2789	81839	265648	15
both	'65-69'	150970	6717	0.04449	0.20019	14973	2997	67372	183809	12.3
both	'70-74'	117070	7561	0.06458	0.27803	11976	3330	51555	116436	9.7
both	'75-79'	74400	7114	0.09562	0.38585	8646	3336	34891	64882	7.5
both	'80-84'	43200	6242	0.14449	0.53073	5310	2818	19505	29991	5.6
both	'85-89'	13870	2949	0.21264	0.69417	2492	1730	8135	10486	4.2
both	'90-94'	3010	914	0.30362	0.79446	762	605	1994	2351	3.1
both	'95-99'	361	152	0.42138	0.85133	157	133	316	357	2.3
both	'100+'	28	16	0.57091	1	23	23	41	41	1.8

Sex	Age range	Actual population	Actual deaths	nMx	nqx	lx	ndx	nLx	Tx	ex
males	'<1'	210160	18144	0.08633	0.08141	100000	8141	94301	3806200	38.1
males	'1-4'	820220	10151	0.01238	0.04808	91859	4416	356835	3711899	40.4
males	'5-9'	962390	2097	0.00218	0.01084	87442	948	434842	3355063	38.4
males	'10-14'	869040	1375	0.00158	0.00788	86495	682	430769	2920221	33.8
males	'15-19'	754370	3373	0.00447	0.02211	85813	1897	424321	2489452	29
males	'20-24'	623730	7456	0.01195	0.05804	83916	4870	407403	2065131	24.6
males	'25-29'	486050	14183	0.02918	0.13598	79045	10749	368355	1657728	21
males	'30-34'	367470	18009	0.04901	0.2183	68297	14909	304211	1289373	18.9
males	'35-39'	291990	15531	0.05319	0.23474	53388	12532	235608	985162	18.5
males	'40-44'	238080	12381	0.05201	0.23011	40856	9401	180775	749554	18.3
males	'45-49'	182320	9022	0.04948	0.22018	31454	6926	139958	568779	18.1
males	'50-54'	122970	5329	0.04333	0.19549	24529	4795	110656	428821	17.5
males	'55-59'	112900	4081	0.03614	0.16574	19734	3271	90492	318164	16.1
males	'60-64'	89050	3624	0.0407	0.1847	16463	3041	74713	227672	13.8
males	'65-69'	69870	3632	0.05198	0.23	13422	3087	59394	152959	11.4
males	'70-74'	53690	3951	0.07359	0.31078	10335	3212	43646	93565	9.1
males	'75-79'	33810	3615	0.10691	0.42181	7123	3005	28105	49919	7
males	'80-84'	19120	3031	0.15854	0.56769	4119	2338	14748	21814	5.3
males	'85-89'	5830	1339	0.22967	0.72948	1781	1299	5655	7066	4
males	'90-94'	1140	369	0.32339	0.82072	482	395	1222	1411	2.9
males	'95-99'	111	49	0.44352	0.86885	86	75	169	188	2.2
males	'100+'	8	5	0.59111	1	11	11	19	19	1.7

Sex	Age range	Actual population	Actual deaths	nMx	nqx	lx	ndx	nLx	Tx	ex
females	'<1'	207640	15113	0.07278	0.06926	100000	6926	95152	3740183	37.4
females	'1-4'	813340	10551	0.01297	0.05032	93074	4684	361057	3645031	39.2
females	'5-9'	958420	2120	0.00221	0.011	88391	972	439522	3283974	37.2
females	'10-14'	867740	1458	0.00168	0.00836	87418	731	435263	2844452	32.5
females	'15-19'	755030	6377	0.00845	0.04136	86687	3585	424472	2409189	27.8
females	'20-24'	619350	15607	0.0252	0.11852	83102	9850	390885	1984716	23.9
females	'25-29'	466220	21200	0.04547	0.20415	73252	14954	328875	1593831	21.8
females	'30-34'	343070	19129	0.05576	0.24469	58298	14265	255827	1264956	21.7
females	'35-39'	278480	13077	0.04696	0.21013	44033	9253	197034	1009129	22.9
females	'40-44'	236860	9268	0.03913	0.17821	34780	6198	158406	812095	23.3
females	'45-49'	189770	5725	0.03017	0.14026	28582	4009	132889	653689	22.9
females	'50-54'	132310	3533	0.0267	0.12516	24573	3076	115178	520800	21.2
females	'55-59'	124280	3117	0.02508	0.11799	21498	2537	101147	405623	18.9
females	'60-64'	100180	2825	0.0282	0.13171	18961	2497	88562	304475	16.1
females	'65-69'	81100	3085	0.03804	0.17369	16464	2860	75170	215913	13.1
females	'70-74'	63380	3610	0.05695	0.24927	13604	3391	59543	140743	10.3
females	'75-79'	40590	3499	0.08621	0.35462	10213	3622	42011	81200	8
females	'80-84'	24080	3211	0.13333	0.49999	6591	3296	24717	39190	5.9
females	'85-89'	8040	1610	0.20029	0.6673	3296	2199	10980	14472	4.4
females	'90-94'	1870	545	0.29156	0.77763	1096	853	2924	3492	3.2
females	'95-99'	250	103	0.41155	0.84319	244	206	500	567	2.3
females	'100+'	20	11	0.56283	1	38	38	68	68	1.8

Figure I.1: Source: World Health Organisation [203]

Appendix J

Heligman-Pollard Method

The Heligman-Pollard (HP) method can be used as a tool for expanding an abridged life table. It is used to graduate a set of age-specific statistics for a standard set of age groups, into a set of single-year statistics. This graduation is accomplished using the following eight-parameter formula known as the Heligman-Pollard model mortality schedule

$$\frac{q_x}{p_x} = A^{(x+B)^C} + Dexp\left[-E(\ln(x/F))^2\right] + GH^x \tag{J.1}$$

where q_x is the probability of an individual at exact age x dying before reaching exact age $x+1$; $p_x = 1 - q_x$; and $\mathbf{C} = A, B, C, D, E, F, G, H$.

Using the same argument and notation as Kostaki [100], if we define the HP equation as $F(x, \mathbf{C})$, then

$$\frac{q_x}{p_x} = F(x, \mathbf{C}) \tag{J.2}$$

which can be rearranged to give

$$q_x = (1 - q_x)F(x, \mathbf{C}) \tag{J.3}$$

$$q_x(1 + F(x, \mathbf{C})) = F(x, \mathbf{C}) \tag{J.4}$$

$$q_x = \frac{F(x, \mathbf{C})}{1 + F(x, \mathbf{C})} \tag{J.5}$$

which we call $G(x, \mathbf{C})$. This is the approximation to q_x, where q_x are the one year probabilities of dying.

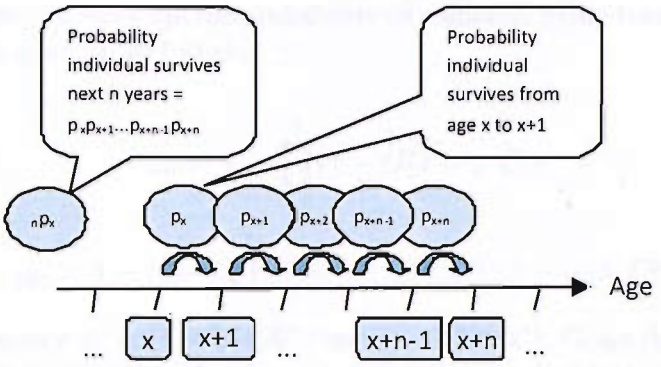


Figure J.1: Illustration of the definition of ${}_np_x$

The probability that an individual aged x will die in the next n years is represented by ${}_nq_x$, which can be defined as ${}_nq_x = 1 - {}_np_x$, where ${}_np_x$ is the probability that an individual will survive an age group. The probability that an individual will survive an age group is the same as multiplying together the incremental probabilities of surviving each year within the age group $(p_x, p_{x+1}, p_{x+2} \ldots)$ as demonstrated by Figure J.1. Therefore ${}_np_x$ can be defined as follows,

$$ {}_np_x = p_x p_{x+1} \cdots p_{x+n-1}, \tag{J.6} $$

which means that

$$ {}_nq_x = 1 - \prod_{i=0}^{n-1} p_{x+i} \tag{J.7} $$

$$ {}_nq_x = 1 - \prod_{i=0}^{n-1} (1 - q_{x+i}). \tag{J.8} $$

We can now substitute our approximation for q_x $(G(x, \mathbf{C}))$ into equation J.8 to give us an approximation of ${}_nq_x$ which we call $nG(x, \mathbf{C})$ and is given by

$$ nG(x, \mathbf{C}) \approx {}_nq_x = 1 - \prod_{i=0}^{n-1} (1 - G(x+i, \mathbf{C})). \tag{J.9} $$

Because the abridged life tables provide us with the ${}_nq_x$ values, we can now approximate $\mathbf{C}$ by minimising the sum of squares

$$ \sum \left(\frac{nG(x, \mathbf{C})}{{}_nq_x} - 1 \right)^2. \tag{J.10} $$

So, for example, to work out the probability of someone dying between the ages 1 and 4, we calculate ${}_3q_1$ as follows:

$$ {}_3q_1 = 1 - \prod_{i=0}^{2}(1 - G(x+i, \mathbf{C})) \tag{J.11} $$

$$ {}_3q_1 = 1 - (1 - G(1, \mathbf{C}))(1 - G(2, \mathbf{C}))(1 - G(3, \mathbf{C})) \tag{J.12} $$

where, for instance $\hat{p_1} = (1 - G(1, \mathbf{C}))$ and $\hat{q_1} = G(1, \mathbf{C})$. Given that we know the value of ${}_3q_1$ from the life tables, we can find estimates for $\mathbf{C}$ which minimise the error between the observed and approximated value.

Given that all of the ${}_nq_x$ values are known, and that we have equations, involving the set of parameters $\mathbf{C}$, to approximate each ${}_nq_x$; we can use Solver, optimisation software in Microsoft Office Excel, to find estimates of $\mathbf{C}$ which minimise the sum of squared errors between the observed and approximated values.

We used this method to obtain a set of q_x values for both males and females, using the ${}_nq_x$ values from the 2000 WHO life tables for Zimbabwe [203]. The starting values for the set of HP equation parameters $\mathbf{C}$ were taken from an available state mortality schedule for Connecticut, USA and are given in Table J.1.

HP Parameter	Starting Value
A	0.00068
B	0.01003
C	0.10752
D	0.00079
E	9.71352
F	21.10103
G	0.000047
H	1.09600

Table J.1: Starting values for the parameters of HP equation.
Source: Connecticut Mortality Schedule [144]

Using Solver to minimise expression J.10 we obtained the following HP parameter values (Table J.2).

HP Parameter	Male: Fitted Value	Female: Fitted Value
A	0.023065687	0.027125667
B	0.00001E-05	0.00001E-05
C	0.253089397	0.282582935
D	0.049452797	0.052293262
E	4.517829352	4.823913966
F	37.9363389	32.18790048
G	0.000103685	6.92359E-05
H	1.094210892	1.097515398

Table J.2: HP parameter values obtained for the male and female population of Zimbabwe, using the 2000 WHO life tables for Zimbabwe [203]

Once the q_x values for both the male and female population were obtained, these were used to construct a complete life table for each gender. The statistics and how they were calculated can be seen in Table J.3 and the resulting distributions of life expectancy and survival are given in Chapter 6, Figures 6.6 and 6.7.

Life Table Statistic	Definition	Method
q_x	The probability of an individual at exact age x dying before reaching exact age $x+1$	Heligman-Pollard
l_x	The number of persons surviving to exact age x	$l_x = l_{x-1}(1-q_{x-1})$

Life Table Statistic	Definition	Method
d_x	Number of deaths at age x	$d_x = l_x q_x$
L_x	The number of person years lived between exact age x and $x+1$	$L_x = l_x - 0.5 d_x$
T_x	The number of person years lived after exact age x	$T_x = L_x + L_{x+1} + L_{x+2} + ...$
e_x	The average number of years of life remaining at exact age x	$e_x = T_x / l_x$

Table J.3: A definition of the statistics used to construct complete life tables for the male and female population in Zimbabwe

Appendix K

Obtaining HIV Prevalence Estimates using a Double Logistic Equation

HIV prevalence data for Zimbabwe is available between 1984 and 2000 [176]. In Chapter 6, Section 6.5 we mention that we fit a double logistic equation to this data to describe its behaviour, and obtain complete estimates for HIV prevalence. This Appendix gives a description of the fitting process.

A double logistic curve was chosen as it is recommended by WHO/UNAIDS [175] and is an established approach to making epidemiological estimates of HIV prevalence in countries with a concentrated epidemic and where there is evidence of a decline in prevalence [111].

Equation K.1 gives the expression for a double logistic. The double logistic equation has been used by previous modellers to estimate HIV prevalence [55] [54]. A double logistic equation is chosen as it allows the initial rate of increase, α, the peak prevalence, a, the final steady-state prevalence, b, the rate of convergence to the steady state, β, and the timing of the epidemic, $\tilde{t}$, to be defined given the HIV prevalence data.

$$p(t) = \frac{e^{\alpha(t-\tilde{t})}}{1+e^{\alpha(t-\tilde{t})}} \left(\frac{ae^{-\beta(t-\tilde{t})}}{1+e^{-\beta(t-\tilde{t})}} + b \right) \tag{K.1}$$

We fit the equation to the HIV data using maximum likelihood estimation (MLE). This method is explained fully in Appendix D. The method defines a likelihood function which is maximised by estimating the value of the unknown parameters, a, b, α, β and $\tilde{t}$.

The result of the fitting process is the double logistic curve shown in Chapter 6, Figure 6.13. The estimates obtained for the parameters of the equation are shown in Table K.1. They imply that the HIV epidemic started during 1990, the peak

prevalence of the epidemic is 28.2% and that long term HIV prevalence will be 13.3%.

Parameter	Estimated Value
a	0.2815
b	0.1327
β	0.0323
$\tilde{t}$	1990.6
α	0.6427

Table K.1: Parameter estimates of the double logistic equation fit to the HIV prevalance data for Zimbabwe using MLE

Appendix L

Uncertainty Analysis

As part of the model validation process described in Chapter 7, we use a technique called *parameter variability-sensitivity analysis*, [150] which involves changing input parameter values to determine the effect on the model's behaviour. This Appendix describes the analysis that was done.

Most of the parameter values for the DES model have been taken from the Harare baseline data, previous modelling literature and expert opinion and therefore we have some confidence in their values. We have however, considered the impact of changing some of the more uncertain parameters such as the length of early- and late-stage HIV, the HIV survival rate and the average household size. Understanding the effect of changing these parameter values gives us insight into how the various aspects influence the model and the TB epidemic it can produce.

L.1 Length of Early- and Late-Stage HIV

Expert opinion suggests that the average duration of early-stage HIV is 3 months. This means that individuals have an increased susceptibility to developing active disease after initial infection with TB for 3 months after infection with HIV. We investigated the effect of changing this duration to see whether reducing or increasing the duration has a significant influence on the average TB epidemic produced by the model. We looked at scenarios where the average duration of early-stage HIV were 0, 1.5, 3, 6 and 9 months, which meant an individual would be early-stage for the first 0%, 1.25%, 2.5%, 5.0% and 7.5% of their HIV infection.

Figure L.1 shows the fit of the model to the TB incidence data for Zimbabwe using the above scenarios. We found that changing the duration of early-stage HIV does not noticeably change the average fit of the model. Using least squares

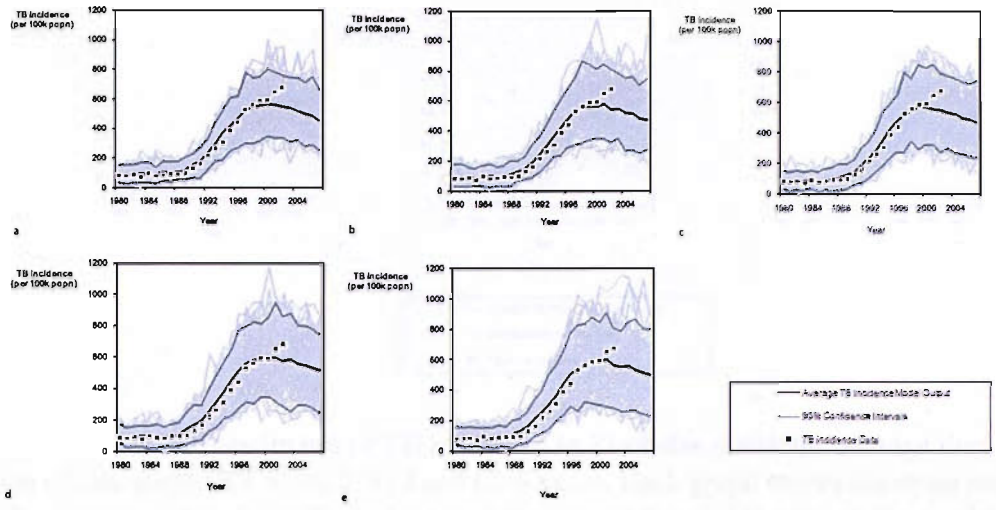


Figure L.1: Model estimates of TB incidence in Zimbabwe with the average duration of early-stage HIV at (a) 0 (b) 1.5 (c) 3 (d) 6 and (e) 9 months. Each graph shows the observed TB incidence data from Zimbabwe and the output from 100 runs of the model, along with the average result and 90% confidence intervals for each scenario

analysis[1] to measure the fit of the average TB incidence curves to the TB data suggests that using a duration of 3 months is optimal as it produces the lowest sum of squares value.

Changing an individual's average duration for early-stage HIV does not significantly effect the TB epidemic produced by the model. This implies that the duration is too short for the temporary increase in susceptibility to TB disease progression to have a considerable impact on the TB epidemic. There is some argument to suggest, therefore, that modelling early-stage HIV is unnecessary.

Expert opinion suggests that the average duration of late-stage HIV is 4 years, occurring on average 6 years after initial infection with HIV. This means that individuals have an increased susceptibility to developing active disease after initial infection with TB and that they are at an increased risk of endogenous reactivation after 6 years of being infected with HIV. We investigated the effect of changing this duration to see whether reducing or increasing the duration has a significant influence on the average TB epidemic produced by the model. We looked at scenarios where the average duration of late-stage HIV were 2, 4 or 6 years, which meant an individual would be late-stage for the last 20%, 40% or 60% of their HIV

[1]This method simply calculates the difference between the model's average predicted values and the values observed in the data. The sum of the residuals squared gives our 'sum of squares' value, which allows the fit of different models to be compared.

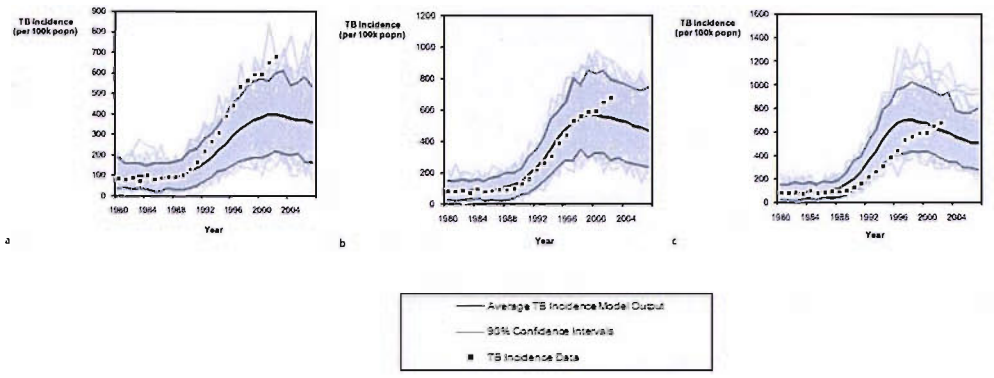


Figure L.2: Model estimates of TB incidence in Zimbabwe with the average duration of late-stage HIV at (a) 2 (b) 4 and (c) 6 years. Each graph shows the observed TB incidence data from Zimbabwe and the output from 100 runs of the model, along with the average result and 90% confidence intervals for each scenario

infection.

Figure L.2 shows the fit of the model to the TB incidence data for Zimbabwe using the above scenarios. We found that changing the duration of late-stage HIV does cause a very noticeable change in the average fit of the model and that the duration of late-stage seems to effect the timing of the epidemic and the amplifying effect of HIV. Using an average duration of 4 years gives the best model fit, with 2 years causing a TB epidemic which is too small and occurs too late, and 6 years causing a TB epidemic which is too large and occurs too early.

L.2 HIV Survival Rate

In our model we assume that the survival time of a newly infected HIV individual can be modelled using a Weibull distribution with a mean of 10 years as per Salomon and Murray [146]. Although this is acceptable and has been used frequently there is some uncertainty surrounding the time from infection to death. We investigated the effect of changing this survival function to see whether reducing or increasing the length of survival has a significant influence on the average TB epidemic produced by the model. We looked at an individual's survival time being on average 8 years, 10 years and 12 years after infection by adjusting the α parameter value of the Weibull distribution but maintaining the same shape parameter.

Figure L.3 shows the fit of the model to the TB incidence data for Zimbabwe using the various survival distributions. We found that changing the average survival time of an HIV individual effects the timing of the epidemic. The average

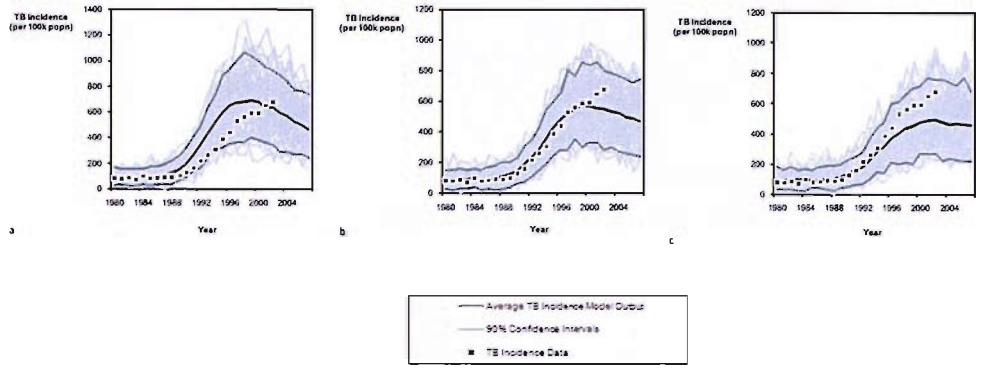


Figure L.3: Model estimates of TB incidence in Zimbabwe with the average duration of survival after infection with HIV at (a) 8 (b) 10 and (c) 12 years. Each graph shows the observed TB incidence data from Zimbabwe and the output from 100 runs of the model, along with the average result and 90% confidence intervals for each scenario

length of early- and late-stage HIV were kept the same. This meant that individuals became late-stage earlier if they were alive for 8 years, than if they were alive 12 years. This is reflected in the resulting model fits that show that using a survival time of 8 years causes the TB epidemic to occur too early and too late when 12 years is used.

L.3 Size of Household

To describe the distribution of household size in Harare, we use a Poisson distribution with mean 3.99, which was obtained by using observed data from Harare on the number of individuals within each household. We are confident that this gives us an accurate representation of the distribution of household size in our study area, but literature suggests that for Zimbabwe as a whole, the average household size would be higher [1] [65] [157] [190] and therefore it would be interesting to see what effect it would have if we introduced a higher household size into the model. We looked at using a Poisson distribution with mean household size 5.5 which is the average of the estimates found in the literature.

Figure L.4 shows the fit of the model to the TB incidence data for Zimbabwe using an average household size of 5.5. We found that changing the distribution of household size does not notably change the average fit of the model. This suggests that changing the average size of households from 3.99 to 5.5 does not effect the TB epidemic produced.

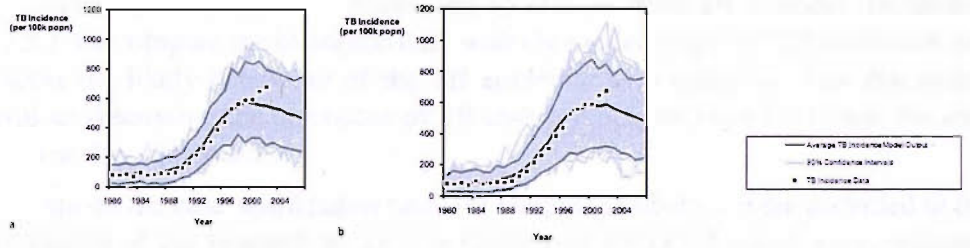


Figure L.4: Model estimates of TB incidence in Zimbabwe with the average household size at (a) 3.99 and (b) 5.5. Each graph shows the observed TB incidence data from Zimbabwe and the output from 100 runs of the model, along with the average result and 90% confidence intervals for each scenario

Appendix M

TB Case-Notification Rate for Harare, Zimbabwe

In Chapter 7, Section 7.7 we discuss the validation of the DES model. In Section 7.7.5.1 we compare the historical data with the model output of TB incidence and discuss the likely behaviour of the TB epidemic as it matures. This discussion involves discussing the behaviour of TB case-notification rates in Harare; the data given in this Appendix.

Data on the case-notification rates for Harare, Zimbabwe were provided in the last month of the research by Dr Liz Corbett of DETECTB and were extracted from reports by the Harare City Health Department, Zimbabwe.

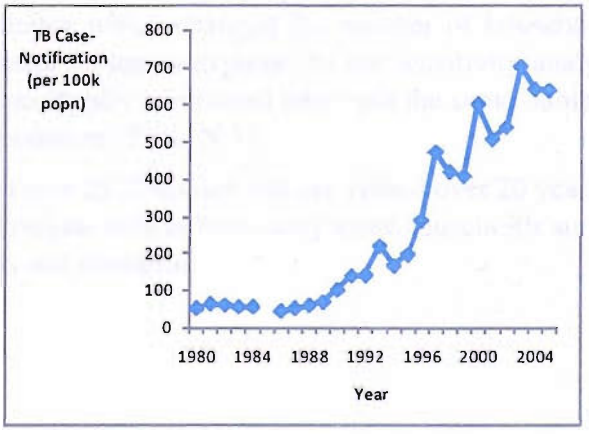


Figure M.1: TB case-notification rates (per 100,000 members of the population) for Harare, Zimbabwe

Appendix N

Number of Households Visited in Each of the Active Case-Finding Interventions

Table N.1 shows the number of households visited per 10,000 households as part of each 20 year active case-finding intervention defined in Chapter 8. It shows that 2972 households are visited over 20 years in interventions 1-3 but that over 6.5 more households are visited through interventions 4 and 5.

Table N.2 shows the number of households visited per 10,000 households as part of each 20 year active case-finding intervention when other scenarios were investigated during the sensitivity analysis and scenario analysis in Chapters 8 and 9. Only scenarios which changed the number of households being visited are shown. The other scenarios explored in the sensitivity analysis and scenario analysis and not specifically mentioned here visit the same number of households as the original experiment (Table N.1).

Given that we know 2972 households are visited over 20 years in interventions 1-3, the figure in brackets tells us how many more households are visited under the other interventions and scenarios.

Intervention	Number of Households (per 10,000 households)
1-3	2972
4-5	19300 (x 6.5)

Table N.1: Table showing the number of households visited per 10,000 households as part of each 20 year active case-finding intervention

Intervention	**Number of Households (per 10,000 households)**	Introducing Interventions in 1996 (Section 8.4.3)	Visiting 25% of Late-Stage Individual's Households (Section 9.4.1)	Visiting 50% of Late-Stage Individual's Households (Section 9.4.2)	Visiting 75% of Late-Stage Individual's Households (Section 9.4.3)
1-3		3849 (x 1.3)	2972 (x 1)	2972 (x 1)	2972 (x 1)
4-5		21844 (x 7.3)	4822 (x 1.6)	9666 (x 3.3)	14320 (x 4.8)

Table N.2: Table showing the number of households visited per 10,000 households as part of each 20 year active case-finding intervention

Appendix O

Scenario Analysis Results

In Chapter 8 we define the original experiment which assumes that in case-finding strategy 4, we investigate each member of the household of all persons entering late-stage HIV to see if there are any undetected TB cases or members with a TB infection in the household (we assume that it is around this time that an HIV-positive individual will approach the health services). In reality, not 100% of HIV-positive individuals will seek medical attention when becoming sick and therefore Chapter 9, Section 9.4 explores the relative performance of intervention 4 when the assumption regarding the proportion of late-stage households that are visited is varied. A late-stage household is a household of an HIV-positive individual who has just entered late-stage HIV. We look at the impact of only 25%, 50% and 75% of these individuals presenting themselves for medical attention, causing their household to be investigated for TB. This Appendix shows graphs of the average number of additional TB cases found, TB deaths averted and TB cases averted, per 100,000 members of the population, by each of the household interventions in each of the scenarios.

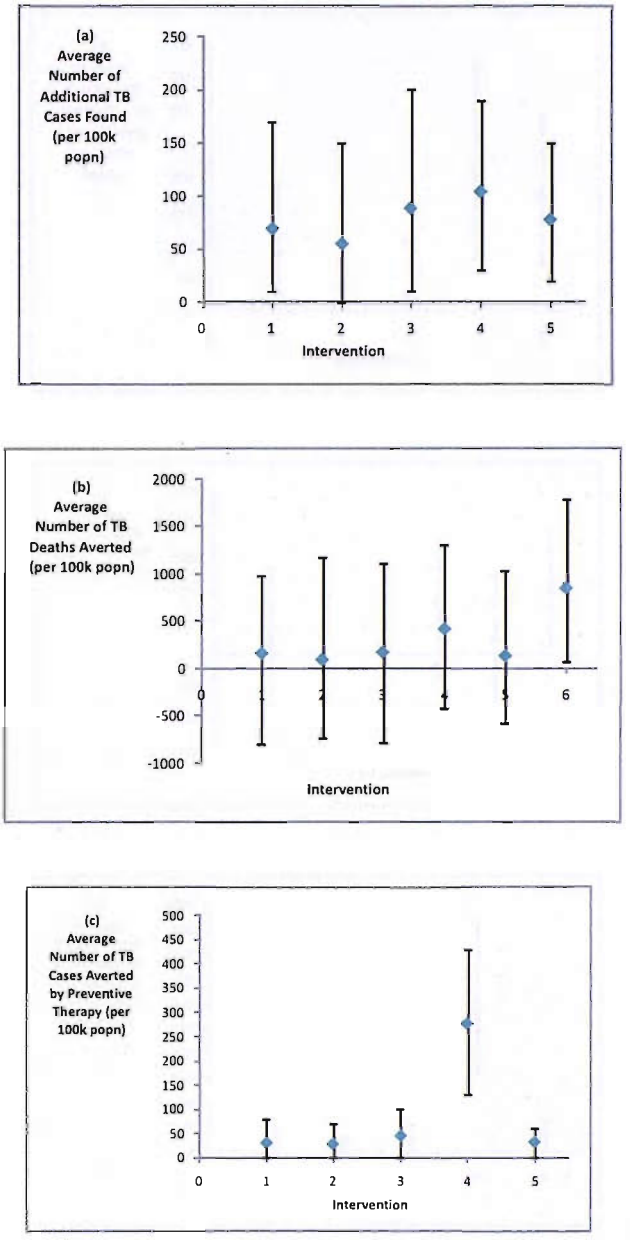
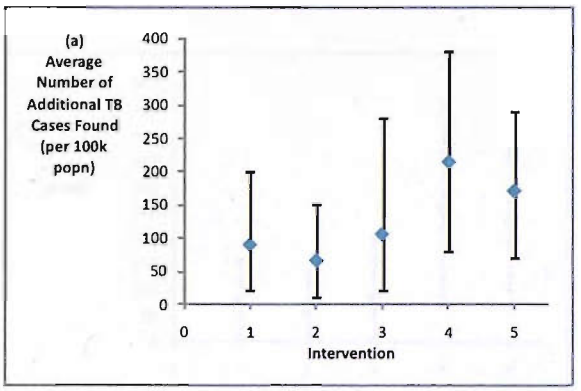
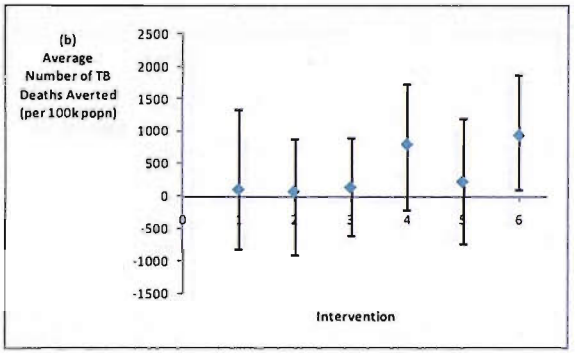


Figure O.1: **Visiting 25% of Late-Stage Households:** The average number of additional (a) TB cases found (b) TB deaths averted and (c) TB cases averted, per 100,000 population, by each of the household interventions when compared with the base case; 90% confidence intervals are included





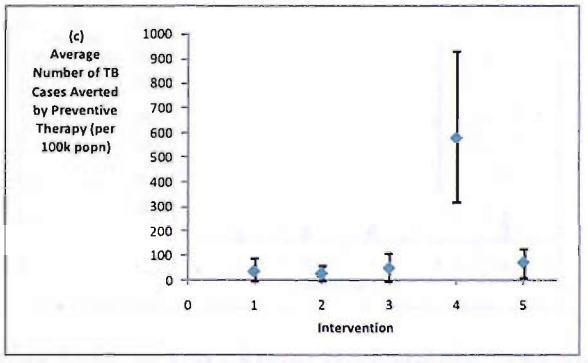


Figure O.2: **Visiting 50% of Late-Stage Households:** The average number of additional (a) TB cases found (b) TB deaths averted and (c) TB cases averted, per 100,000 population, by each of the household interventions when compared with the base case; 90% confidence intervals are included

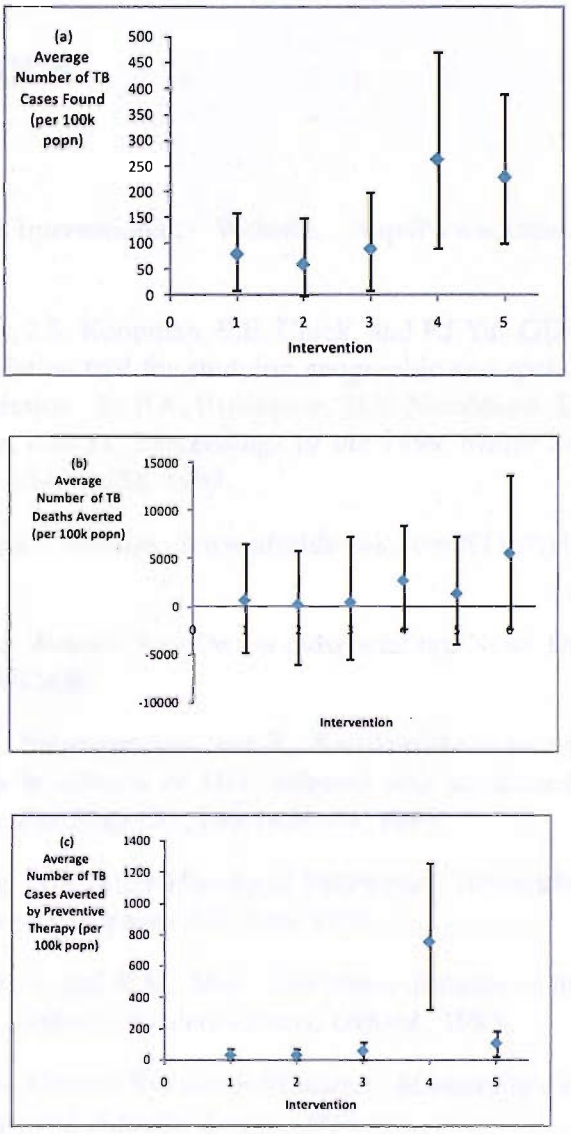


Figure O.3: **Visiting 75% of Late-Stage Households:** The average number of additional (a) TB cases found (b) TB deaths averted and (c) TB cases averted, per 100,000 population, by each of the household interventions when compared with the base case; 90% confidence intervals are included

References

[1] Action Aid International. Website. http://www.kubatana.net, accessed 24/05/06.

[2] A.L. Adams, J.S. Koopman, S.E. Chick, and P.J Yu. GERMS: an epidemiologic simulation tool for studying geographic and social effects on infection transmission. In P.A. Farrington, H.B Nembhard, D.T. Sturrock, and G.W. Evans, editors, *Proceedings of the 1999 Winter Simulation Conference*, pages 1549–1556, 1999.

[3] Afraid to Ask. Website. www.afraidtoask.com/STD/frHIV.html, accessed 24/05/06.

[4] AIDS Portal. Website. http://www.aidsportal.org/News_Details.aspx?ID=6238, accessed 19/02/08.

[5] S. Allen, J. Batungwanyo, and K. Kerlikowske. Two-year incidence of tuberculosis in cohorts of HIV infected and uninfected urban Rwandan woman. *Am Rev Resp Dis*, 146:1439–44, 1992.

[6] A.W. Alling. The After-History of Pulmonary Tuberculosis: A Stochastic Model. *Biometrics*, pages 527–546, 1958.

[7] R.M. Anderson and R.M. May. Infectious diseases of humans: dynamics and control. *Oxford University Press, Oxford.*, 1993.

[8] Anonymous. Central Bureau of Statistics. *Ministry of Economic Planning and Development, Nairobi, Kenya*, 2000.

[9] Anonymous. The World Health Report. *World Health Organization*, 2000.

[10] Anonymous. National HIV Prevalence in Kenya. *The National AIDS and STDs Control Programme*, 2003.

[11] J. P. Aparicio, A. F. Capurro, and C. Castillo-Chavez. Transmission and dynamics of tuberculosis on generalized households. *Journal of Theoretical Biology*, 206(3):327–341, 2000.

[12] J. P. Aparicio, A. F. Capurro, and C. Castillo-Chavez. Markers of Disease Evolution: the case of TB. *J Theor Biol*, 215(2):227–37, 2002.

[13] Archives News Releases. Website. http://pubs.ama-assn.org/media/2007a/0226.dtl#2, accessed 27/09/07.

[14] C. Aschwanden. Spatial simulation model for infectious viral diseases with focus on sars and the common flu. *Proceedings of the 37th Hawaii International Conference on System Sciences*, pages 1–5, 2004.

[15] Y. Azuma. A simple simulation model of tuberculosis epidemiology for use without large-scale computers. *Bull World Health Organ*, 52(3):313–22, 1975.

[16] N.T.J. Bailey. A simple stochastic epidemic. *Biometrika*, 37:193–2002, 1950.

[17] N.T.J. Bailey. The total size of a general stochastic epidemic. *Biometrika*, 40:177–185, 1953.

[18] N.T.J. Bailey. The mathematical theory of epidemics. *Charles Griffin and Company*, 1957.

[19] F.G. Ball and O.D. Lyne. *Epidemics among a population of households.* Mathematical approaches for emerging and reemerging infectious diseases. Springer-Verlag, New York, 2002.

[20] G. Barnett, S. Grzybowski, and K. Styblo. Present risk of developing active tuberculosis in Saskatchewan according to previous tuberculin and X-ray status. *Bull IUATLD*, 45:51–74, 1971.

[21] M.S. Bartlett. Some evolutionary stochastic processes. *Journal of the Royal Statistical Society*, 11(2):211–229, 1949.

[22] M.S. Bartlett. Measles periodicity and community size. *Journal of the Royal Statistical Association*, 120(1):48–70, 1957.

[23] M.N. Bates, A. Khalakdina, M. Pai, L. Chang, F. Lessa, and K.R. Smith. Risk of tuberculosis from exposure to tobacco smoke: A systematic review and meta-analysis. *Arch Intern Med.*, 167:335–342, 2007.

[24] Berkeley Madonna. Website. www.berkeleymadonna.com, accessed 24/05/06.

[25] A. Bermejo, H. Veeken, and A. Berra. TB Incidence in Developing Countries with High HIV Prevalence of HIV Infection. *AIDS*, 6:1203–6, 1992.

[26] D. Bernoulli. Essai d'une nouvelle analyse de la mortalit cause par la petite vrole et des advantages de l'inoculation pour la prvenir. *Mm.Math.Phys.Acad.Roy.Sci.*, pages 1–45, 1760.

[27] B.R. Bloom and J.D. McKinney. The death and resurrection of tuberculosis. *Nature Medicine*, 5:872–4, 1999.

[28] S. M. Blower and J. L. Gerberding. Understanding, predicting and controlling the emergence of drug-resistant tuberculosis: a theoretical framework. *Journal of Molecular Medicine-Jmm*, 76(9):624–636, 1998.

[29] S. M. Blower, K.V. Koelle, and T. Lietman. Antibiotic resistance: to treat...(or not to treat). *Nature Medicine*, 5:358, 1999.

[30] SM. Blower, AR. Mclean, TC. Porco, PM. Small, PC. Hopwell, MA. Sanchez, and AR. Moss. The intrinsic transmission dynamics of tuberculosis epidemics. *Nature Medicine*, 1:815–821, 1995.

[31] SM. Blower, PM. Small, and PC. Hopwell. Control strategies for tuberculosis epidemics: new models for old problems. *Science*, 273:497–500, 1996.

[32] Botswana Press Agency. Website. http://www.gov.bw/cgi-bin/news.cgi?d=20040513&i=BOTUSA_to_undertake_study_in_two_cities, accessed 04/09/07.

[33] TF. Brewer, SJ. Heymann, GA. Colditz, ME. Wilson, K. Auerbach, D. Kane, and HV. Fineberg. Evaluation of tuberculosis control policies using computer simulation. *Journal of the American Medical Association*, 276(23):1989–1903, 1996.

[34] British Broadcasting Company. Website. http://news.bbc.co.uk/1/hi/business/7244769.stm, accessed 19/02/08.

[35] British Broadcasting Company. Website. http://news.bbc.co.uk/1/hi/world/africa/6449691.stm, accessed 19/02/08.

[36] S. Brogger. Systems analysis in Tuberculosis control: a model. *American Review of Respiratory Disease*, 95(3):419–34, 1965.

[37] C. Castillo-Chavez and Z. L. Feng. To treat or not to treat: The case of Tuberculosis. *Journal of Mathematical Biology*, 35(6):629–656, 1997.

[38] S. Cauchemez, F. Carrat, C. Viboud, A.J. Valleron, and P.Y. Boelle. A bayesian mcmc approach to study transmission of influenza: application to household longitudinal data. *Statistics in Medicine*, 23:3469–3487, 2004.

[39] CDC. Website. http://www.cdc.gov/hiv/pubs/facts/hivtb.htm, accessed 24/05/06.

[40] Chaotic Random Number Generators with Random Cycle Lengths. Website. www.agner.org/random/theory, accessed 24/05/06.

[41] R.C.H. Cheng. *Bootstrap Resampling Methods.* University of Southampton, GTP Course, 2005.

[42] M.R. Chernick. *Bootstrap Methods, A Practitioner's Guide.* Wiley, New York, 1999.

[43] W Chorba and J.L. Sanders. Planning Models for Tuberculosis Control Programs. *Health Services Research*, 6(2):144–64, 1971.

[44] CNN. Website. http://archives.cnn.com/2000/WORLD/africa/04/18/ zimbabwe.land.03/, accessed 24/05/06.

[45] Tanzania Tuberculin Survey Collaboration. Tuberculosis control in the era of the HIV epidemic: risk of Tuberculosis infection in Tanzania, 1983-1998. *International Journal of Tuberculosis and Lung Disease*, 5(2):103–112, 2001.

[46] E. L. Corbett, S. Charalambous, K. Fielding, T. Clayton, R. J. Hayes, K. M. De Cock, and G. J. Churchyard. Stable incidence rates of tuberculosis (TB) among human immunodeficiency virus (HIV)-negative South African gold miners during a decade of epidemic HIV-associated TB. *J Infect Dis*, 188(8):1156–63, 2003.

[47] E. L. Corbett, C. J. Watt, N. Walker, D. Maher, B. G. Williams, M. C. Raviglione, and C. Dye. The growing burden of Tuberculosis: global trends and interactions with the HIV epidemic. *Archives of Internal Medicine*, 163(9):1009–21, 2003.

[48] E.L. Corbett. The impact of HIV on TB prevalence in Southern Africa. *Int J Tuberc Lung Dis*, page S8:S14, 2004.

[49] E.L. Corbett, S. Charalambou, V.M. Moloi, K. Fielding, A.D. Grant, C. Dye, K.M. De Cock, R.J. Hayes, BG. Williams, and GJ. Churchyard. Human Immunodeficiency Virus and the Prevalence of Undiagnosed Tuberculosis in African Gold Miners. *Am J Respir Crit Care Med*, 170:673–679, 2004.

[50] Corbett, E. L. Behaviour of TB incidence and HIV prevalence in Zimbabwe. Personal Email, October 2007.

[51] Corbett, E. L. Bill and Melinda Gates Foundation Funding Proposal. Report, August 2007.

[52] Corbett, E. L. and Williams, B. G. Strategies for reducing the burden of TB infection and disease in South African goldminers: the role of preventive therapy. Report, September 2001.

[53] F. A. Coutinho, L. F. Lopez, M. N. Burattini, and E. Massad. Modelling the natural history of HIV infection in individuals and its epidemiological implications. *Bulletin of Mathematical Biology*, 63(6):1041–62, 2001.

[54] C. S. M. Currie, K. Floyd, B. G. Williams, and C. Dye. Cost, affordability and cost-effectiveness of strategies to control tuberculosis in countries with high HIV prevalence. *BMC Public Health*, 5(130), 2005.

[55] C. S. M. Currie, B. G. Williams, R. C. H. Cheng, and C. Dye. Tuberculosis epidemics driven by HIV: is prevention better than cure? *AIDS*, 17(17):2501–2508, 2003.

[56] C. S. M. Currie, B. G. Williams, and E. L. Corbett. Assessing the impact of HIV on the annual risk of TB infection using a mathematical model. *Proc. TSRU*, 2005.

[57] A.C. Davison and D.V. Hinkley. *Bootstrap Methods and their Application.* Cambridge University Press, Cambridge, 1997.

[58] K.M. De Cock, B. Soro, I.M. Coulibaly, and S.B. Lucas. Tuberculosis and HIV infection in sub-Saharan Africa. *JAMA*, 268:1581–1587, 1992.

[59] S. M. Debanne, R. A. Bielefeld, G. M. Cauthen, T. M. Daniel, and D. Y. Rowland. Multivariate Markovian modeling of Tuberculosis: Forecast for the United States. *Emerging Infectious Diseases*, 6(2):148–157, 2000.

[60] J.M. Dreesman. An array based simulation approach for predicting the impact of different measles vaccination strategies in lower saxony. *Supplement to the APL Berlin 2000 proceedings*, pages 5–12, 2000.

[61] C. Dye, G. P. Garnett, A. Sleeman, and B. G. Williams. Prospects for worldwide Tuberculosis control under the WHO DOTS strategy. *Lancet*, 352(9144):1886–1891, 1998.

[62] C. Dye, S. Scheele, P. Dolin, V. Pathania, and MC. Raviglione. Global burden of Tuberculosis: estimated incidence, prevalence, and mortality by country. *Journal of the American Medical Association*, 282:677–686, 1999.

[63] C. Dye and B. G. Williams. Criteria for the control of drug-resistant Tuberculosis. *Proc. Natl. Acad. Sci. USA*, 97(14):8180–8185, 2000.

[64] Massad E. A homogeneity mixing model for the AIDS epidemic. *Mathematical and Computer Modelling*, 12(1):89–96, 1989.

[65] Earthscape. Website. http://www.earthscape.org/r1/ES16562, accessed 08/08/07.

[66] B. Efron and R.J. Tibshirani. *An Introduction to the Bootstrap*. Monographs on Statistics and Applied Probability 57. Chapman and Hall, London, 2003.

[67] L.R. Elveback, J.P. Fox, E. Ackerman, A. Langworthy, M. Boyd, and L. Gatewood. An influenza simulation model for immunization studies. *American Journal of Epidemiology*, 103(2):152–165, 1976.

[68] Z. Feng, M. Iannelli, and F. A. Milner. A two-strain Tuberculosis model with age of infection. *Siam Journal on Applied Mathematics*, 62(5):1634–1656, 2002.

[69] Z. L. Feng, C. Castillo-Chavez, and A. F. Capurro. A model for Tuberculosis with exogenous reinfection. *Theoretical Population Biology*, 57(3):235–247, 2000.

[70] Zhilan Feng, Wenzhang Huang, and Carlos Castillo-Chavez. On the Role of Variable Latent Periods in Mathematical Models for Tuberculosis. *Journal of Dynamics and Differential Equations*, 13(2):425–452, 2001.

[71] SH Ferebee. An Epidemiological Model of Tuberculosis in the United States. *NTA Bulletin*, pages 4–7, 1967.

[72] SH Ferebee. Controlled chemoprophylaxis trials in Tuberculosis: a general review. *Adv Tuberc Res*, 17:28–106, 1970.

[73] Forum for Science, Industry and Business. Website. www.innovations-report.com/html/reports/medicine_health/report-23774.html, accessed 24/05/06.

[74] J Frimodt-Moller. A community-wide Tuberculosis study in a South Indian rural population, 1950-1955. *Bull World Health Organisation*, 22:61–170, 1960.

[75] W.H. Frost. How much control of Tuberculosis? *American Journal of Public Health*, 27:759–766, 1937.

[76] Comstock G. Epidemiology of Tuberculosis. *Am Rev Resp Dis*, 125:8–16, 1982.

[77] A Garcia, J Maccario, and S Richardson. Modelling the annual risk of Tuberculosis infection. *Int. J. Epidemiol.*, 26(1):190–203, 1997.

[78] L. Getoor, J. T. Rhee, D. Koller, and P. Small. Understanding Tuberculosis epidemiology using structured statistical models. *Artificial Intelligence in Medicine*, 30(3):233–256, 2004.

[79] W. Githui, P. Nunn, E. Juma, F. Karimi, R. Brindle, R. Kamunyi, S. Gathua, C. Gicheha, J. Morris, and M. Omwega. Cohort study of HIV-positive and HIV-negative tuberculosis, Nairobi, Kenya: comparison of bacteriological results. *Tubercle and Lung Disease*, 73:203–209, 1992.

[80] Global Health Council. Website. http://globalhealth.org/view_top.php3?id =227, accessed 24/05/06.

[81] J. R. Glynn, A. C. Crampin, B. M. Ngwira, F. D. Mwaungulu, D. T. Mwafulirwa, S. Floyd, J. M. Ponnighaus, D. K. Warndorff, and P. E. Fine. Trends in Tuberculosis and the influence of HIV infection in northern Malawi, 1988-2001. *Aids*, 18(10):1459–63, 2004.

[82] P. Godoy, A. Dominguez, J. Alvarez, N. Camps, JM. Jansa, J. Alcaide, and S. Minguell. Risk factors associated with cases of pulmonary tuberculosis coinfected by HIV with a smear positive for Mycobacterium tuberculosis. *Int Conf AIDS*, 14:7–12, 2002.

[83] EH Goh and KL Fam. A Dynamic Model of Tuberculosis Epidemiology for Singapore. *Annals Academy of Medicine*, 10(1):40–49, 1981.

[84] MC. Gomes, A. Margheri, and C. Rebelo. Stability and persistence in a compartment model of pulmonary TB. *Nonlinear Analysis: Theory, methods and applications*, 48(4):617–636, 2000.

[85] E. Halloran, I. Longini, D.M. Cowart, and A. Nizam. Community interventions and the epidemic prevention potential. *Vaccine*, 20:3254–3262, 2002.

[86] W. H. Hamer. On epidemic disease in england - the evidence of variability and of persistency of type. *The Lancet*, 167:733–739, 1906.

[87] L. Heligman, J.H. Pollard, and M.A. Pollard. The age pattern of mortality. *Journal of the Institute of Actuaries*, 107:49–80, 1980.

[88] G. Hertzberg. The infectiousness of human tuberculosis; an epidemiological investigation. *Acta Tuberc Scand Suppl*, 38:1–146, 1957.

[89] SJ. Heymann. Modeling the efficacy of prophylactic and curative therapies for preventing the spread of Tuberculosis in Africa. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 87:406–411, 1993.

[90] PC. Hill, D. Jackson-Sillah, SA. Donkor, J. Otu, RA. Adegbola, and C. Lienhardt. Risk factors for pulmonary tuberculosis: a clinic-based case control study in The Gambia. *BMC Public Health*, 6:156, 2006.

[91] O Horwitz. Disease, Cure and Death: Epidemiologic and Clinical Parameters for Chronic Diseases Illustrated by a Model - Tuberculosis. *American Journal of Epidemiology*, 97(3):148–159, 1973.

[92] O. Horwitz, E. Wilbek, and PA. Erickson. Epidemiological basis of Tuberculosis eradication. 10 longitudinal studies on the risk of Tuberculosis in the general population of a low-prevalence area. *Bull World Health Org*, 41:95–113, 1969.

[93] International Union Against Tuberculosis Committee on Prophylaxis. Efficacy of various durations of isoniazid preventive therapy: five years of follow-up in the IUAT trial. *Bulletin of the World Health Organisation*, 60(4):555–564, 1982.

[94] MR Joesoef, PL Remington, and PT Jiptoherijanto. Epidemiological model and cost-effectiveness analysis of Tuberculosis treatment programmes in Indonesia. *Int. J. Epidemiol.*, 18(1):174–179, 1989.

[95] AM Johnson, AH Mercer, B Erens, AJ Copas, S McManus, K Wellings, KA Fenton, C Korovessis, W Macdowall, K Nanchahal, S Purdon, and J Field. Sexual behaviour in Britain: partnerships, practices, and HIV risk behaviours. *Lancet*, 358:1835–42, 2001.

[96] T. Kenyon, S. Valway, W. Ihle, I. Onorato, and K. Castro. Transmission of Multidrug-Resistant Mycobacterium tuberculosis during a Long Airplane Flight. *NEJM*, 334:933–938, 1996.

[97] W. O. Kermack and A. G. McKendrick. A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London*, 115(772):700–721, 1927.

[98] W. O. Kermack and A. G. McKendrick. Contributions to the mathematical theory of epidemics. ii. the problem of endemicity. *Proceedings of the Royal Society of London*, 138(834):55–83, 1932.

[99] W. O. Kermack and A. G. McKendrick. Contributions to the mathematical theory of epidemics. iii. further studies of the problem of endemicity. *Proceedings of the Royal Society of London*, 141(843):94–122, 1933.

[100] A Kostaki. The heligman-pollard formula as a tool for expanding an abridged life table. *Journal of Official Statistics*, 7(3):311–323, 1991.

[101] M. Kretzschmar, Y. Van Duynhoven, and A. Severijnen. Modeling prevention strategies for gonorrhea and chlamydia using stochastic network simulations. *American Journal of Epidemiology*, 144(3):306–317, 1996.

[102] V. Krishnamurthy, S. Nair, and G. Gothi. Incidence of Tuberculosis among newly infected populations and in relation to the duration of infected status. *Indian J Tuberc*, 33:1–3, 1976.

[103] Chaudhuri K Krishnamurthy V. Risk of pulmonary Tuberculosis associated with exogenous reinfection and endogenous reactivation in a south Indian rural population - A mathematical estimate. *Indian J Tuberc*, 37:63–67, 1990.

[104] Averill M. Law and Michael G. McComas. How to build valid and credible simulation models. In B. A. Peters, J. S. Smith, D. J. Medeiros, and M. W. Rohrer, editors, *Proceedings of the 2001 Winter Simulation Conference*, pages 22–29, 2001.

[105] S.A. Levin. New directions in the mathematics of infectious diseases. In C. Castillo-Chavez et al, editor, *Mathematical approaches for emerging and reemerging infectious diseases*, pages 1–5. Springer Verlag, New York, 2002.

[106] C. Lienhardt, K. Fielding, JS. Sillah, B. Bah, P. Gustafson, D. Warndorff, M. Palayew, I. Lisse, S. Donkor, S and. Diallo, K. Manneh, R. Adegbola, P. Aaby, O. Bah-Sow, S. Bennett, and K. McAdam. Investigation of the risk factors for tuberculosis: a case-control study in three countries in West Africa. *International Journal of Epidemiology*, 34(4):914–923, 2005.

[107] T. Lietman and S. M. Blower. Potential impact of Tuberculosis vaccines as epidemic control agents. *Clinical Infectious Diseases*, 30:S316–S322, 2000.

[108] T. Lietman, TC. Porco, and S. M. Blower. Leprosy and Tuberculosis: the epidemiological consequences of cross-immunity. *American Journal of Public Health*, 87:1923–7, 1997.

[109] H.H. Lin, M. Ezzati, and M. Murray. Indoor Air Pollution and Tuberculosis: A Systematic Review and Meta-Analysis. *PLoS Medicine*, 4(1):e20 doi:10.1371/journal.pmed.0040020, 2007.

[110] I. Longini Jr and J.S. Koopman. Household and community transmission parameters from final distributions of infections in households. *Biometrics*, 38(1):115–126, 1982.

[111] R. Lyerla, E. Gouws, J.M. Garca-Calleja, and E. Zaniewski. The 2005 Workbook: an improved tool for estimating HIV prevalence in countries with low level and concentrated epidemics. *Sex. Transm. Inf*, 82:41–44, 2006.

[112] P. Manfredi and J.R. Williams. Realistic population dynamics in epidemiological models: the impact of population decline on the dynamics of childhood infectious diseases. Measles in Italy as an example. *Mathematical Biosciences*, page 153175, 2004.

[113] E. Massad, M. N. Burattini, F. A. B. Coutinho, H. M. Yang, and S. M. Raimundo. Modelling the Interaction between Aids and Tuberculosis. *Mathematical and Computer Modelling*, 17(9):7–21, 1993.

[114] M. Matsumoto and T. Nishimura. Sex Differences in the Epidemiology of Tuberculosis in San Fransisco. *ACM Transactions on Modeling and Computer Simulation*, 8(1):3–30, 1998.

[115] Medic8 Family Health Guide. Website. www.medic8.com/healthguide, accessed 24/05/06.

[116] G.R. Mellor, J. Kangangi, J. Mansoer, L. Marum, R.C.H. Cheng, C. Dye, and B.G. Williams. The impact of HIV on TB at a district level in Kenya. *Unpublished*, 2007.

[117] G. B. Migliori, A. Borghesi, A. Spanevello, P. Eriki, M. Raviglione, G. Maciocco, A. Morandi, L. Ballardini, and M. Neri. Risk of infection and estimated incidence of Tuberculosis in northern Uganda. *Eur Respir J*, 7(5):946–53, 1994.

[118] D. Morgan, C. Mahe, B. Mayanja, and JAG. Whitworth. Progression to symptomatic disease in people infected with HIV-1 in rural Uganda: prospective cohort study. *British Medical Journal*, 324:193–197, 2002.

[119] D. Mulder, A. Nunn, and A. Kamali. Two-year HIV-1 associated mortality in a Ugandan rural population cohort. *Lancet*, 343:1021–1023, 1994.

[120] C. J. L. Murray and J. A. Salomon. Modeling the impact of global Tuberculosis control strategies. *Proceedings of the National Academy of Sciences of the United States of America*, 95(23):13881–13886, 1998a.

[121] C. J. L. Murray and J. A. Salomon. Expanding the WHO Tuberculosis control strategy: rethinking the role of active case-finding. *Int. J. Tuberc Lung Dis*, 2(9):S9–S15, 1998b.

[122] M Murray. Determinants of cluster distribution in the molecular epidemiology of Tuberculosis. *Proceedings of the National Academy of Sciences of the United States of America*, 99(3):1538–1543, 2002.

[123] N.J.D. Nagelkerke, Prabhat Jha, S. J. de Vlas, L. KorenrompEline, Stephen Moses, J. F. Blanchard, and F. A. Plummer. Modelling HIV/AIDS epidemics in Botswana and India: Impact of interventions to prevent transmission. *Bulletin of the World Health Organization*, 80(2), 2002.

[124] R. Narain, S.S. Nair, G.R. Rao, and P. Chandrasekhar. Distribution of tuberculosis infection and disease among households in a rural community. *Bulletin of the World Health Organization*, 34:639–654, 1966.

[125] J.A. Nelder and R. Mead. A simplex method for function minimization. *Computer Journal*, 7:308–313, 1965.

[126] WHO Stop TB News. Tb-hiv: Fuelling each other. *Global Alliance for TB drug development*, 2001.

[127] P. Nunn, R. Brindle, and L. Carpenter. Cohort Study of human immunodeficiency virus infection in patients with TB in Nairobi, Kenya: analysis of early (6 month) mortality. *Am Rev Resp Dis*, 146:849–54, 1992.

[128] P. Nunn and M. Felten. Surveillance of resistance to anti Tuberculosis drugs in developing countries. *Tuberc Lung Dis*, 75:163–7, 1994.

[129] Object Services and Consulting, Inc. Website. http://www.objs.com/x3h7/cplus.htm, accessed 10/07/07.

[130] World Health Organization. Acquired immunodeficiency syndrome (AIDS). Interim proposal for a WHO staging system for HIV infection and disease. *Weekly Epidemiological Record*, 65:221–224, 1990.

[131] World Health Organization. Global tuberculosis control: surveillance, planning, financing. *WHO/TB/2003.316*, 2003.

[132] World Health Organization. Global tuberculosis control: surveillance, planning, financing. *WHO/TB/2005*, 2005.

[133] World Health Organization. Global tuberculosis control: surveillance, planning, financing. *WHO/TB/2006*, 2006.

[134] World Health Organization. Global tuberculosis control: surveillance, planning, financing. *WHO/HTM/TB/2007.376*, 2007.

[135] J.H. Periens, M.E. St Louis, and Y.B. Mukadi. Pulmonary Tuberculosis in HIV-infected patients in Zaire: a controlled trial of treatment for either 6 or 12 months. *N Engl J Med*, 332:779–84, 1995.

[136] Planet Math. Website. http://planetmath.org/encyclopedia/Likelihood Function.html, accessed 18/07/07.

[137] Catharina P. B. Van Der Ploeg, Carina Van Vliet, Sake J. De Vlas, Jeckoniah O. Ndinya-Achola, Lieve Fransen, Gerrit J. Van Oortmarssen, and J. Dik F. Habbema. Stdsim: a microsimulation model for decision support in std control. *Interfaces*, 28(3):84–100, 1998.

[138] T. C. Porco and S. M. Blower. Quantifying the intrinsic transmission dynamics of Tuberculosis. *Theoretical Population Biology*, 54(2):117–132, 1998.

[139] T. C. Porco, P. M. Small, and S. M. Blower. Amplification dynamics: Predicting the effect of HIV on Tuberculosis outbreaks. *Journal of Acquired Immune Deficiency Syndromes*, 28(5):437–444, 2001.

[140] CS. ReVelle, WR. Lynn, and F. Feldmann. Mathematical models for the economic allocation of TB control activities in developing nations. *American Review of Respiratory Disease*, 96(5):893–909, 1967.

[141] CS. ReVelle and J. Male. A mathematical model for determining case-finding and treatment activities in Tuberculosis control programs. *American Review of Respiratory Disease*, 102:403–11, 1970.

[142] S. B. Richards, M. E. St Louis, P. Nieburg, I. M. Coulibaly, D. Coulibaly, L. Abouya, H. D. Gayle, and K. M. De Cock. Impact of the HIV epidemic on trends in Tuberculosis in Abidjan, Cote d'Ivoire. *Tuber Lung Dis*, 76(1):11–6, 1995.

[143] H. L. Rieder. Contacts of Tuberculosis patients in high-incidence countries. *Int J Tuberc Lung Dis*, 7(12):S333–S336, 2003.

[144] A. Rogers and K. Gard. Applications of the Heligman-Pollard model mortality schedule. *Population Bulletin of the United Nations*, 30:79–105, 1991.

[145] R. Ross. The prevention of malaria. *2nd edn, Murray, London*, 1911.

[146] J.A. Salomon and C.L. Murray. Modeling HIV/AIDS epidemics in Sub-Saharan Africa using seroprevalence data from antenatal clinics. *Bulletin of the World Health Organization*, 79:596–607, 2001.

[147] EE Salpeter and SR Salpeter. Mathematical Model for the Epidemiology of Tuberculosis, with estimates of the Reproductive Number and Infection-Delay Function. *American Journal of Epidemiology*, 142(4):398–406, 1998.

[148] MA Sanchez and SM Blower. Uncertainty and sensitivity analysis of the basic reproductive rate. Tuberculosis as an example. *Am. J. Epidemiol.*, 145(12):1127–1137, 1997.

[149] RG Sargent. Simulation model verification and validation. In B.L. Nelson, W.D. Kelton, and G.M. Clark, editors, *Proceedings of the 1991 Winter Simulation Conference*, pages 37–47, 1991.

[150] RG Sargent. Verification and validation of simulation models. In D.J. Medeiros, E.F. Watson, J.S. Carson, M.S. Manivannan, and N.J. Piscataway, editors, *Proceedings of the 1998 Winter Simulation Conference*, 1998.

[151] L. Sattenspiel. Epidemics in nonrandomly mixing populations: A simulation. *American Journal of Physical Anthropology*, 73(2):251–265, 1987.

[152] M Schulzer, DA Enarson, S Grzybowski, YP Hong, SJ Kim, and TP Lin. An analysis of pulmonary Tuberculosis data in Taiwan and Korea. *Int. J. Epidemiol.*, 16(4):584–589, 1987.

[153] MJ. Schulzer, DA. Enarson, and S. Grybowski. An estimate of the future size of the Tuberculosis problem in sub-Saharan Africa resulting from HIV infection. *Tubercule and Lung Disease*, 73:52–58, 1992.

[154] MJ. Schulzer, MP. Radhamani, S. Grybowski, E. Mak, and JM. Fitzgerald. A mathematical model for the prediction of the impact of HIV infection on Tuberculosis. *International Journal of Epidemiology*, 23:400–407, 1994.

[155] B. J. Song, C. Castillo-Chavez, and J. P. Aparicio. Tuberculosis models with fast and slow dynamics: the role of close and casual contacts. *Mathematical Biosciences*, 180:187–205, 2002.

[156] H.E. Soper. The interpretation of periodicity in disease prevalence. *Journal of the Royal Statistical Society*, 92(1):34–73, 1929.

[157] Southern African Regional Poverty Network. Website. http://www.sarpn.org.za/documents/d0000022/page5.php, accessed 08/08/07.

[158] A. Stuart, J.K. Ord, and A.O. Hagan. *Kendall's Advanced Theory of Statistics*. Oxford University Press, London, 1994.

[159] K Styblo. Tuberculosis control and surveillance. *Advances in Respiratory Medicine*, pages 77–108, 1976.

[160] K Styblo. The relationship between the risk of tuberculous infection and the risk of developing infectious tuberculosis. *Bull Int Union Tuberc Lung Dis*, 60:117–9, 1985.

[161] K Styblo. Epidemiology of Tuberculosis. *The Hague: Royal Netherlands Tuberculosis Association (KNCV)*, 1991.

[162] I Sutherland. An estimation of the scope for BCG vaccination in preventing Tuberculosis among those aged 15-19 years in England and Wales at the present time. *Tubercule and Lung Disease*, 40:413, 1959.

[163] I Sutherland. The ten-year incidence of clinical Tuberculosis following 'conversion' in 2,550 individuals aged 14-19 years. *Tuberculosis Surveillance and Research Unit Progress Report. The Hague, Royal Netherlands Tuberculosis Association (KNCV)*, 1968.

[164] I. Sutherland, E. Svandova, and S. Radhakrishna. The development of clinical Tuberculosis following infection with tubercle bacilli. *Tubercule*, 62:255–68, 1982.

[165] The American Social Health Association. Website. www.ashastd.org/stdfaqs/asha_hiv.html#HIV, accessed 24/05/06.

[166] The Bloomsbury Wellcome Trust Centre. Website. http://www.lshtm.ac.uk/wbc/, accessed 04/09/07.

[167] The Consortium to Respond Effectively to the AIDS-TB Epidemic, Clinical Research Unit, London School of Hygiene and Tropical Medicine. Website. http://www.lshtm.ac.uk/cru/research/cru_research_detail.php?id=122, accessed 04/09/07.

[168] The Economist. Website. http://www.economist.com/world/africa/displaystory.cfm?story_id=9475943, accessed 19/02/08.

[169] The Guardian. Website. http://www.guardian.co.uk/world/2007/jul/01/zimbabwe.southafrica, accessed 19/02/08.

[170] The National AIDS and STDs Control Program. National hiv prevlence in kenya, 2003.

[171] Time Magazine. Website. http://www.time.com/time/printout/0,8816,203620,00.html, accessed 19/02/08.

[172] J Trefny and E Hejdova. A Model of the epidemiology of Tuberculosis in the Czech Socialist Republic. *Bull Int Union Tuberc*, 57(3-4):206–211, 1982.

[173] UK Coalition of People Living with HIV and AIDS. Website. www.ukcoalition.org/tb/history.html, accessed 24/05/06.

[174] UNAIDS. AIDS epidemic update - Sub-Saharan Africa, 2006.

[175] World Health Organisation & UNAIDS. Website. http://data.unaids.org/pub/Presentation/2007/workbook_2007_pres_en.pdf, accessed 02/10/07.

[176] UNAIDS/WHO. UNAIDS/WHO Global HIV/AIDS Online Database, 2006.

[177] J.S Urban Hjorth. *Computer Intensive Statistical Methods*. Chapman and Hall, London, 1994.

[178] U.S. Department of Labor Office of Administrative Law Judges. Website. www.oalj.dol.gov/public/dot/refrnc/glossary.htm, accessed 24/05/06.

[179] USA Today. Website. http://www.usatoday.com/news/world/2002/06/25/zimbabwe.htm, accessed 19/02/08.

[180] E. Vynnycky. An investigation of the transmission dynamics of M. Tuberculosis. *PhD Thesis, University of London*, 1996.

[181] E. Vynnycky and P. E. M. Fine. The natural history of Tuberculosis: the implications of age- dependent risks of disease and the role of reinfection. *Epidemiology and Infection*, 119(2):183–201, 1997.

[182] HT. Waaler. A dynamic model for the epidemiology of Tuberculosis. *American Review of Respiratory Disease*, 98:591–600, 1968a.

[183] HT. Waaler. Cost-Benefit Analysis of BCG-Vaccination under Various Epidemiological Situations. *Bull Int Union Tuberc*, 41:42–52, 1968b.

[184] HT. Waaler. Model Simulation and Decision-Making in Tuberculosis Programmes. *Bull Int Union Tuberc*, 8:337–44, 1970b.

[185] HT. Waaler. Tuberculosis and Poverty. *International Journal of Tuberculosis and Lung Disease*, 6(9):745–746, 2004.

[186] HT. Waaler, A. Geser, and S Anderson. The use of mathematical models in the study of the epidemiology of Tuberculosis. *American Journal of Public Health*, 52(6):1002–1013, 1962.

[187] HT. Waaler and MA. Piot. The use of an epidemiological model for estimating the effectiveness of Tuberculosis control measures. Sensitivity of the effectiveness of Tuberculosis control measures to the coverage of the population. *Bull World Health Organ*, 41(1):75–93, 1969.

[188] HT. Waaler and MA. Piot. Use of an Epidemiological Model for Estimating the Effectiveness of Tuberculosis control measures. Sensitivity of the effectiveness of Tuberculosis control measures to the Social Time Preference. *Bulletin of the World Health Organization*, 43:1–16, 1970a.

[189] DF. Wares, M. Akhtar, S. Singh, and H. Luitel. Is TB contact screening relevant in a developing country setting? Experiences from eastern Nepal, 1996-1998. *International Journal of Tuberculosis and Lung Disease*, 4:920–924, 2000.

[190] Zimbabwe Emergency Water and Sanitation Project. Website. http://pdf.dec.org/pdf_docs/PDACG298.pdf, accessed 08/08/07.

[191] D. Weycker, J. Edelsberg, E. Halloran, I. Longini Jr, A. Nizam, V. Ciuryla, and G. Oster. Population-wide benefits of routine vaccination of children against influenza. *Vaccine*, 23(10):1284–1293, 2005.

[192] C. Whalen, C. Horsburgh, and D. Hom. Accelerated course of human immunodeficiency virus infection after Tuberculosis. *Am J Resp Crit Care Med*, 151:129–35, 1995.

[193] P. Whittle. The outcome of a stochastic epidemic - a note on bailey's paper. *Biometrika*, 42:116–122, 1955.

[194] WHO Stop TB. Website. www.stoptb.org, accessed 09/08/06.

[195] B. G. Williams and C. Dye. Maximum-Likelihood for Parasitologists. *Parasitology Today*, 10(12):489–493, 1994.

[196] B. G. Williams and C. Dye. Antiretroviral drugs for Tuberculosis control in the era of HIV/AIDS. *Science*, 301:1535–1537, 2003.

[197] B. G. Williams, R. Granich, L.S. Chauhan, N.S. Dharmshaktu, and C. Dye. The impact of HIV/AIDS on the control of Tuberculosis in India. *Proc. Natl. Acad. Sci. USA*, 102(27):9619–9624, 2005.

[198] B.G. Williams, E Gouws, M Colvin, F Sitas, G Ramjee, and SS Abdool Karim. Patterns of infection: using age prevalence data to understand the epidemic of HIV in South Africa. *South African Journal of Science*, 96:1–9, 2000.

[199] B.G. Williams, A. Kochi, and C. Dye. The impact of antiretroviral therapy on the life expectancy of people living with HIV/AIDS. *Submitted for publication*, 2003.

[200] S.D. Wilson-Clark, S.L. Deeks, E. Gournis, K. Hay, S. Bondy, E. Kennedy, I. Johnson, E. Rea, T. Kuschak, D. Green, Z. Abbas, and B. Guarda. Household transmission of SARS, 2003. *Canadian Medical Association Journal*, 175(10):1219–23, 2006.

[201] WordIQ. Website. www.wordiq.com/definition/tuberculosis, accessed 24/05/06.

[202] World Health Organisation. Website. www.who.int/mediacentre/factsheets/fs104/en/, accessed 24/05/06.

[203] World Health Organisation Life Tables. Website. http://www.who.int/whosis/database/life_tables/life_tables_process.cfm?path=whosis,life_tables&language=english, accessed 02/07/07.

[204] World Health Organisation Regional Office for South-East Asia. Website. http://www.searo.who.int/en/Section10/Section18/Section356/Section415_1714.htm, accessed 04/09/07.

[205] H.J. Zar, M.F. Cotton, S. Strauss, J. Karpakis, G. Hussey, H.S. Schaaf, H. Rabie, and C.J. Lombard. Effect of isoniazid prophylaxis on mortality and incidence of tuberculosis in children with HIV: randomised controlled trial. *BMJ*, 334(7585):136, 2007.

[206] E. Ziv, C. L. Daley, and S. M. Blower. Early therapy for latent Tuberculosis infection. *American Journal of Epidemiology*, 153(4):381–385, 2001.