EPIDEMIOLOGICAL MODELS FOR TUBERCULOSIS CASES PROGRESSION

HILLARY KIPCHUMBA KIPRUTO

DOCTOR OF PHILOSOPHY

(Applied Statistics)

JOMO KENYATTA UNIVERSITY OF AGRICULTURE AND TECHNOLOGY

2016

Epidemiological Models for Tuberculosis Cases Progression

Hillary Kipchumba Kipruto

A thesis submitted in fulfilment for the Degree of Doctor of Philosophy in Applied Statistics in the Jomo Kenyatta University of Agriculture and Technology

2016

DECLARATION

This thesis is my original work and has not been presented for a degree in any other University.

Signature: . . . . . Date: . . . . . .

Hillary Kipchumba Kipruto

This thesis has been submitted for examination with our approval as University Supervisors.

Signature: . . . . . Date: . . . . . .

Dr. Joseph K. Mung'atu

JKUAT, Kenya

Signature: . . . . . Date: . . . . . .

Dr. Samuel M. Mwalili

JKUAT, Kenya

Signature: . . . . . Date: . . . . . .

Dr. Kenneth Ogila

JKUAT, Kenya

DEDICATION

To God Who alone knows the desires of our hearts and what beholds the future of every mortal.

ACKNOWLEDGEMENT

First and foremost I thank the Almighty God for giving good health and providence during my research period. I also would like to express my sincere appreciation to my supervisors: Dr.Joseph Kyalo Mung'atu, Dr.Keneth Ogila and Dr.S.M. Mwalili for their guidance, advice, patience, moral and professional support during the entire learning period. Besides their tight schedules, they still had time with me over the duration. Additionally, I thank all the colleagues from the Ministry of Health and Kenya Medical Research Institute whom we worked and continue to work closely in the fight of tuberculosis, specifically Dr.Jane Ongango, Dr.Joseph Sitienei, Dr.Enos Masini, Richard Kiplimo , Gilbert Sang and Eliud Kibuchi your comments and insights enabled me to get this far. I will forever be grateful to you, To my friends Dr.Bernard Langat, Aiban Ronoh, Dr.Humphrey Karamagi and Wesley Tomno thank you for all the encouragement through the years. In a special way, I am exceedingly grateful to all the members of the Statistics and Actuarial Science (STACS) department, JKUAT, for the guidance and support I received at the different stages in the development of my work, Special appreciation goes to the Chair of the Department Dr.George Orwa for the constant encouragement. Lastly and most graciously to my loving family thank you for being there for me through this period. I will forever cherish the sacrifices you made for me. To get this far I can only say it is through the mighty hand of God and forever be praises unto him the most high.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

ABBREVIATIONS AND ACRONYMS

AIDS	Acquired Immunodeficiency Syndrome
ARV	Anti-retroviral
BMI	Body Mass Index
CPT	Cotrimoxazole Preventive Therapy
CDR	Case Detection Rate
CNR	Case Notification Rate
DCM	Deterministic Compartmental Model
DLTLD	Division of Leprosy, tuberculosis and Lung Disease
DOTS	Directly-Observed Treatment, Short-course
EPTB	Extra Pulmonary Tuberculosis
GLM	Generalized Linear Model
PDF	Probability density fuction
df	Degrees of freedom
DTLC	District tuberculosis and Leprosy Coordinator
GoK	Government of Kenya
H1N1	Influenza A virus containing Haemagglutinin and Neauraminidase glycoproteins
HIV	Human Immunodeficiency Virus
PTB	Pulmonary tuberculosis
ICM	Individual Contact Model
PTP SS-	Pulmonary TB, sputum smear negative

PTP SS+	Pulmonary TB, sputum smear positive
TB	tuberculosis
SAE	Spatial Area Estimation
SIR	Susceptible Infected Recovered
SI	susceptible Infected
SIS	susceptible Infected Susceptible
INLA	Integrated Nested Laplace Approximation
EVD	Ebola Virus Disease
PQL	Penalized Quasi-likelihood
BHGLMMS	Bayesian Hierarchical. Generalized Linear Mixed Models
STAR	Structured additive regression
ICD-10	International Classification of Diseases version 10
WHO	World Health Organization
EPTB	Extra-Pulmonary tuberculosis
MDR-TB	Multi Drug Resistant TB
XDR-TB	Extremely Drug Resistant TB
MCMC	Markov Chain Monte Carlo methods

NOMENCLATURE

$S(t)$	Number susceptible at time t
$\bar{S}$	Number susceptible in a steady state
$s(t)$	Fraction of the susceptible relative to the population
$I(t)$	Number of infectives at time t
$\bar{I}$	The number infected in a steady state
$i(t)$	Fraction of the infective relative to the population
$R(t)$	Number recovered at time t
$\bar{R}$	The number recovered in a steady state
μ	Per capita death rate
$r(t)$	Fraction of the recovered relative to the population
r	Fraction of those who recovered
λ	Force of infection
β	Average number of infected contacts
γ	Average number of infected contacts who recover
ν	Intervention modification rate
η	Mass action coeficients
θ	Latent period
N	Total population
R_0	Reproduction number
ρ	Relative removal rate
S	Number susceptible
I	Number infected
R	Number recovered
E	Number exposed

ABSTRACT

This research developed epidemiological models for the susceptible infected recovered for tuberculosis epidemic in Kenya using stochastic and deterministic models as well as spatial temporal models. Tuberculosis disease is transmitted within and between communities when infected and susceptible individuals interact. Interest in the epidemiology of TB was triggered by the re-emergence of tuberculosis in the early 1990's with the advent of HIV and falling economic status of many people which subjected them to poverty. In this study, we focused on the period 2000-2013 and all the notified data in Kenya was included. Data on estimates of TB incidence, prevalence and mortality was extracted from the WHO global tuberculosis database. The study was guided by the following objectives: to develop epidemic models for TB progression; to estimate the expected number of individuals with the disease at any given time t; to formulate small area estimation models for TB progression; and to develop spatial-temporal models for the TB progression. The results showed that there was an average decline of 5% over the last 8 years with the highest decline being reported in the year 2012/13. TB continues to disproportionately affect the male gender with 58% being male and 42% being female. Kenya has made significant efforts to address the burden of HIV among TB patients with cotrimoxazole preventive therapy (CPT) uptake reaching 98% with and ART at 74% by the end of 2013. The gains in the decline of TB could be attributed in part to in the outcomes of integrating TB and HIV services and these gains should be sustained. What is equally notable is the clear epidemiologic shift in age indicating reduced transmission in the younger age groups. The spatial reference regions considered were the 47 Kenyan counties. The covariates considered were gender,

HIV positive proportion, directly observed Treatment (DOTs), average weight, average Body Mass Index (BMI) and average age. From the results of all notified TB cases, only average BMI was excluded from the spatial temporal model since it was not statistically significant (p-value > 0.05). The estimated risk of case notification rates per 100,000 were found to be highest in the following counties Marsabit, Isiolo, Nairobi, Lamu, Mombasa, Machakos, Kajiado, Makueni, Kisumu, Siaya and Homabay. The study recommends that efforts must be made in addressing the risk factors for TB which is geographically differentiated.

Chapter 1

INTRODUCTION

1.1 Background Information

In this work we begin by exploring the reasons why mathematical modelling in the study of disease epidemics and epidemiological process has a huge significance and later review the important field of stochastic modeling. A discussion on its implications when applied in the modelling of infectious diseases will also be looked at. The continued challenge of infectious diseases in human beings and animals has brought to the fore the need for rigorous study of mathematical epidemiology which entails developing models for predicting the nature of dynamics of the spread of an infection or a disease. It is widely held in discourse and literature that mathematical models development, their analysis and implementation play a natural and important role in obtaining such understanding. It has been noted that model formulation processes when carried out systematically; clarifies assumptions, variables, and parameters. Further, models provide conceptual results such as thresholds, basic reproduction numbers, contact numbers, and replacement numbers, Hethcote (2000).

The understanding of the transmission characteristics of infectious diseases is paramount particularly in communities, regions, and countries can lead to better design of interventions geared towards decreasing the transmission rates of diseases pathogens. Mathematical models play an important role in comparing, planning, implementing, evaluating, and optimizing various disease detection, prevention, treatment, and control programs, Hethcote (2000).

Mathematical models of the spread of a disease in given populations play a central role in the following key components; understanding the transmission, predicting the future direction the epidemic is going to take when there is an outbreak, its extinc-

tion time and evaluating the efficiency and effectiveness of disease control measures. The validity, richness and reliability of the resulting intervention depends on the reliability and accuracy of the model, Jacob (2010).

An epidemic model can be considered as a simplified means of describing the transmission patterns of communicable disease either through individuals from contact to contact or agents. The term "stochastic" has been used in this thesis to mean being or having a random variable characteristic. A stochastic model thus is a tool for estimating probability distributions of potential outcomes by allowing for random variation in one or more inputs over time. Stochastic models depend on the chance variations in risk of exposure, disease and other illness dynamics. They are mostly used when it is considered that random changes are important, Trottier and Philippe (2001).

The main reason why stochastic models were considered in this study is because of the inherent challenges that are evident and clear in deterministic models which assume that populations are randomly mixing, that is,. every individual contact interacts with every other individual with the same rate, or in a multi-type process, contacts between individuals of certain types all happen at the same rate but these assumptions would not be true of tuberculosis because duration of exposure is an important factor. In the case of tuberculosis an individual my be exposed to the bacteria and could enter into a latent stage or proceed on to get the disease. Some of the major drawbacks of deterministic models have been highlighted by Murray et al. (1986) and Murray (1989) in which it was proposed that a deterministic model be used to predict the dynamics of the prevalence of rabies among foxes in England. In this study, the dynamics predicted by the model produced results that the number of infected foxes will rapidly increase until the number of available susceptible foxes

is too low and then the disease seems to disappear, but after a period of about 2 years there is a sudden reappearing of rabies in foxes predicted. Murray et al. (1986) used a continuous approximation of the number of infected foxes and during the years that the infection seemed to have disappeared (in the predictions based on the model), there still was a minimum of around one infected atto-fox (10^{-18} of a fox) per square kilometre and this fox-part eventually caused the new wave of infection as clearly pointed out by Molison (1991) who tried to explain the phenomenon that Murray et al. (1986) had encountered previously. This shows some of the inherent dangers of continuous approximations of the number of individuals being used in the modelling framework.

From the foregoing challenges of deterministic models, its clear that they are not the most relevant for modelling the start and the end of an epidemic. This is particularly when the number of infectious individuals is small and for modelling the spread of infections on complex interactions of populations when the number of available susceptible individuals that an infective individual can infect, is small. In this thesis, the main focus is on stochastic branching process models and develop stochastic models for tuberculosis diseases which can be generalized to other infectious diseases that are important from epidemiological, mathematical and modelling point of view. Specifically, three problems are considered. Firstly, the problem of evaluating the conditional probability that an epidemic has died out given that no symptomatic cases have been observed, although infected cases with latent form of TB may still be present. Secondly, the effect of predisposing factors for the development of the infectious disease (HIV status, Nutritional status, socio economic and delay in diagnosis) in the presence of spatio-temporal distribution and geographical differences. Finally, a model is developed for tuberculosis in order to assess the trends and de-

duce the future trends of the disease given the current trends and interventions.
In our model, we introduced branching process in a changing (but not random) situation. The study proceeds to estimate the probability of disease dying out and the expected number of infected individuals at a given time t and as a consequence the expected number of TB cases can be estimated. In addition, this branching process is used to calculate the generating function of the number of infected individuals at any given moment in time. The model and methods are designed using tuberculosis data reported in Kenya for the last 14 years.

1.1.1 Basics of Tuberculosis

Tuberculosis (TB) is one of the infectious diseases of public health concern globally. It is caused by *bacillus* bacteria and the most common causative organism is the *Mycobacterium tuberculosis*. Other causative agents that are occasionally implicated are *Mycobacterium bovis* which is transmitted through contaminated milk and *Mycobacterium africanum*, Cadmus et al. (2010). The transmission of the bacteria is through infectious aerosolized droplet nuclei generated by coughing, laughing, talking, sneezing and singing. The ability to generate infectious aerosolized droplet nuclei is dependent on the infectivity of the patient where a sputum smear positive patient is considered most infectious, Dooley et al. (1990). Infection with the *Mycobacterium* does not always lead to development of disease as the immune system is able to contain the infection and the *bacilli* remain dormant. The risk of infection is dependent on the extent to which exposure happens; longer durations of exposure to infected persons who are not on treatment increase the chance of infection, Dooley et al. (1990). Those with the highest risk to develop TB include children under five, the elderly and those who are immunosuppressed.

Globally, the World Health Organization (WHO) estimates that 2 billion people, or 1/3 of the world's population, are infected with (*Mycobacterium tuberculosis*) the *bacillus* that causes tuberculosis (TB). In 2012, there were 8.7 million incident TB cases with 1.4 million people dying of TB, making it the leading infectious cause of death worldwide, WHO (2012).

For individuals who have untreated latent TB infection and immunity is not compromised, the estimated risk of developing tuberculosis disease is 5% to 10% over a lifetime, with about 50 – 80% of that risk occurring during the first 2 years following infection. Individuals with immunocompromised conditions, in particular by HIV co-infection, the risk of developing the disease increases to 5% to 10% per year MACET (2009). Healthy immune systems can contain TB *bacilli* in the vast majority of cases. Individuals at higher risk of developing TB are those with lowered immunity – notably people who are HIV positive, those living in poverty as well as those with other chronic conditions. While HIV attacks the body's immune system directly, characteristics of poverty such as poor nutrition, inadequate housing, limited access to clean water and bad sanitation reduces resistance to TB.

Those who develop TB will develop one of two types: First, the Pulmonary TB (PTB), which accounts for about 80% of TB cases WHO (2012). This is highly infectious and attacks the lungs. There are two sub-categorizations of PTB: one that shows up in a sputum sample (PTB SS+) and that which requires further tests, such as a chest x-ray, to diagnose (PTB SS-). The second type, Extra-pulmonary TB (EPTB) affects organs other than the lungs, such as the spine, lymph nodes or abdomen. EPTB is not infectious, unless it is accompanied by PTB. Figure 1.1.1 shows the flow of the treatment process of TB.

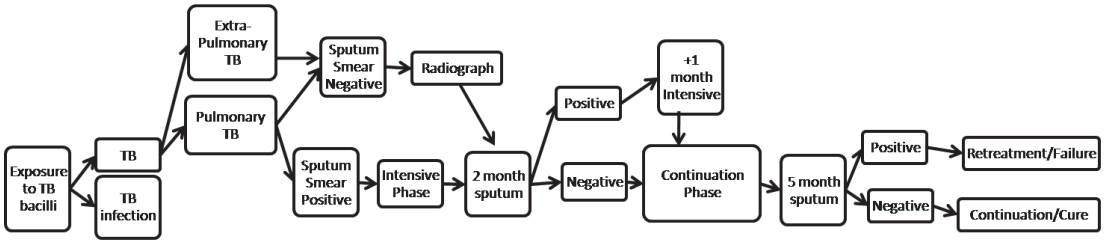


Figure 1.1.1: Tuberculosis treatment process

To combat all forms of TB, a global strategy for diagnosis and treatment for TB was adopted by the WHO, for implementation by national TB programmes, WHO (2006). The strategy has six elements with major strategy being Directly-Observed Treatment Short course popularly known as DOTS, which is simple since it is based on identifying TB cases in the community and treating these cases by directly observing that patients take the correct drug treatment for six (6) to eight (8) months (depending on the type of treatment). The objectives of this strategy are to ensure that the patient is cured thus minimizing the chances that he/she will relapse, and prevent the development of drug-resistant TB. Under DOTS, TB patients require a guardian to ensure and document that they have taken their treatment – often this is a family member, community health worker or a neighbour. Their role is particularly crucial during the first two months of treatment, known as the intensive phase, when the strongest drugs are used. After two months, patients undergo a sputum examination to determine the concentrations of TB *bacilli* in their lungs. Depending on the results, the drug therapy is changed and TB patients enter the continuation phase of treatment for four (4) or six (6) more months. In the intensive phase, most patients or their guardians are required to pick up their drugs from the designated health facility once per week. During the continuation phase, this changes to twice per month.

The length of treatment, frequency of drug collection, and human resources required by DOTS are all onerous aspects of TB treatment, involving direct financial and indirect opportunity costs. As it is widely held in the discourse that poor socioeconomic standing and TB are closely linked, the susceptibility of those who are already poor to a disease that leads to further impoverishment due to illness and treatment demands constitutes a double burden of the most economically vulnerable. Furthermore, where TB is widespread, an undiagnosed TB case can lead to between ten (10) and fifteen (15) new infections per year, WHO (2007). One of the most widely-associated symptoms of TB is haemoptysis (coughing up blood), but in fact, this is a late-developing symptom - WHO guidelines recommend that patients seek medical attention after two weeks of a chronic cough, WHO (2003). As such, the pre-diagnosis phase is a crucial time period for the individual as well as for TB control.

Kenya is among the twenty two (22) TB high burden countries in the world which contribute 80% of the global TB burden WHO (2012). The absolute number of TB cases notified has increased more than ten-fold since 1990 while the TB incidence has increased from below 50 per 100,000 in 1990 to 329 per 100,000 population in 2008, DLTLD (2013). The HIV epidemic is the single most contributing factor for this massive increase in the burden of TB in Kenya. From the Kenya AIDS Indicator Survey 2013 the prevalence of HIV in Kenya currently stands at 5.6%, NASCOP (2013), while the TB HIV co-infection rate was at 39% in 2012, DLTLD (2013). In 2013, diagnosis and treatment of TB was carried out by general health care workers in 1,900 TB diagnostic centres and over 4,300 TB treatment centres in Kenya. This translated to one diagnostic centre per 37,634 populations and one treatment centre

per 18,411 populations.

Diagnosis of TB is through sputum smear microscopy for new smear positive cases. Smear negative cases are diagnosed via a diagnostic algorithm as per the national TB guidelines. The diagnosis of extra pulmonary TB is based on clinical suspicion and the collection of appropriate specimens for TB bacteriology where this is feasible. Kenya developed a medium term TB control strategic plan covering the period 2011-2015 which was modelled along the WHO Stop TB Strategy. The global targets of TB control are to achieve at least 70% TB case detection rate and 85% treatment success rate, WHO (2006).

Kenya has made good progress in the fight against TB; TB Case detection rate (CDR) reached the global target of 80% in 2007, the treatment success rate was 85% in 2007. By the end of 2008, 83% of TB patients were tested for HIV against a national target of 80%, DLTLD (2009).

Despite this progress, major problems still remain: low access to anti-retroviral treatment for HIV infected TB patients, significant delays in TB diagnosis which facilitates TB transmission and is associated with a higher frequency of the poor sequel of TB and the emerging problem of drug resistant tuberculosis. There are concerns that inadequate infection control measures in health care settings may be facilitating the transmission of TB to health care workers, patients and their visitors which has significant consequences for vulnerable groups such as HIV infected persons.

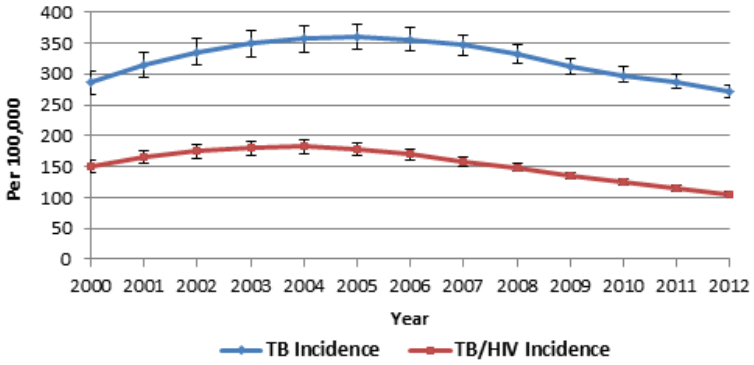
The emergence of multidrug-resistant, *Mycobacterium tuberculosis* (MDR-TB) worldwide poses a serious problem to the treatment of tuberculosis. These MDR strains are at least resistant to two primary chemotherapeutic agents rifampicin and isoniazid, WHO (2003) which require treatment with more costly and more toxic second-line drugs.

1.1.2 Tuberculosis Epidemiology

To understand TB epidemiology, changes in the trends of the burden of TB disease over a period of time are considered. The main indicators for the burden of TB disease include; incidence, prevalence and mortality. These estimates have been derived from WHO databases on TB. Data on TB prevalence and mortality are hardly available due to low coverage of the civil registration systems in the country. Kenya has not conducted a national TB prevalence survey in the recent past (the last TB prevalence survey was conducted in 1956), but it is currently carrying out a prevalence survey (2015/2016). There is currently no national level vital registration system with standard ICD-10 coding in place in Kenya. Less than half of deaths are recorded, and approximately 10% of deaths receive any ICD code. Results from a prevalence survey and vital registration systems can be used to assess the current levels of TB disease and mortality and could also provide important evidence about the effectiveness of current TB programmatic efforts and actions needed to improve TB care and control.

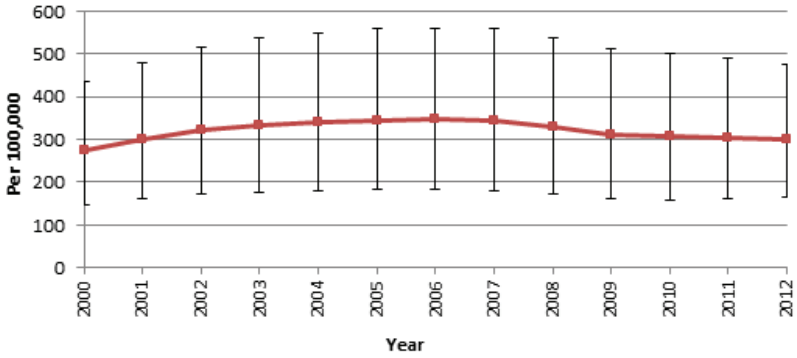
1.1.2.1 Burden of TB Disease

The burden of a disease is looked at from three contexts; incidence, prevalence and mortality. Figure 1.1.2 suggests a consistent decline in new TB cases over time, with the decline in TB cases starting in 2005 following the decline in TB/HIV cases which started in 2004. After peaking in 2006, there has been a slowing phase in the rate of decline in estimates of TB prevalence, Figure 1.1.3. TB mortality estimates suggest an increase in TB deaths in 2011-2012, Figure1.1.4. However, the wide confidence intervals indicate considerable uncertainty in the estimates, suggesting the need for other more direct methods to measure prevalence and mortality.



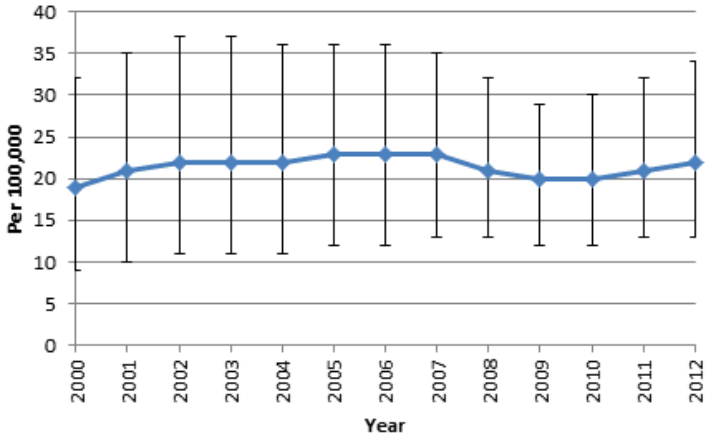
Source: WHO global TB database

Figure 1.1.2: Trend of tuberculosis incidence in Kenya, 2000-2012



Source: WHO global TB database

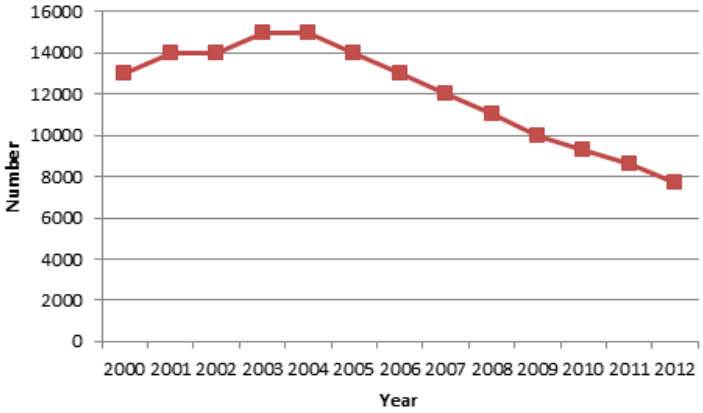
Figure 1.1.3: Trend of tuberculosis prevalence in Kenya, 2000-2012



Source: WHO global TB database

Figure 1.1.4: Trend of tuberculosis patients' mortality in Kenya, 2000-2012

With regards to deaths ocurring to HIV positive TB patients, there has been a consistent decline in the estimated TB-related deaths among people living with HIV, Figure 1.1.5. This could be attributed to intervetions in both HIV and TB programs beging to bear fruits and could need to be further sustained.



Source: WHO global TB database

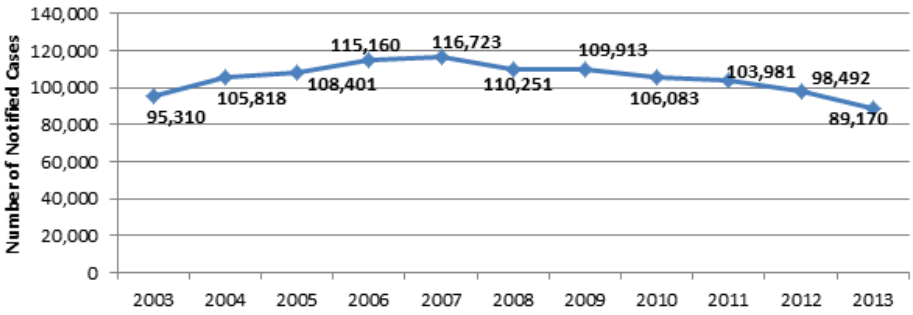
Figure 1.1.5: Trend of deaths due to TB among HIV patients in Kenya, 2000-2012

1.1.2.2 TB Case Notifications

In Kenya the number of notified TB cases (all forms) increased from approximately 95,000 cases in 2003 to a peak of over 116,000 cases in 2007, Figure 1.1.6. After 2007, the number of notified TB cases has steadily declined until 2013, when the number of notified TB cases was approximately 89,000 – the lowest it has been in over a decade. Approximately 10,000 fewer cases were reported in 2013 compared to 2012. This observed trend is consistent with the trend of incidence and prevalence as estimated by WHO.

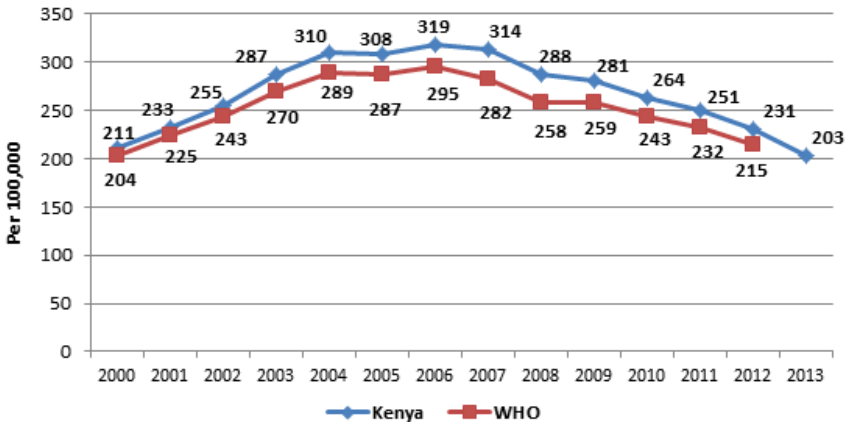
The case notification rate for all TB cases (new and retreatment) shows three distinct phases, Figure 1.1.7. From 2000 to 2004, the TB case notification rate increased; from 2004 to 2006, rates remained constant; and from 2006 to 2013, rates steadily declined, with rates in 2013 lower than those recorded in 2000. Based on national data, the case notification rates fell rapidly between 2011 and 2013, decreasing approximately 8% and 12% each year respectively as shown in Table 1.1. For all years, notification rates based on Kenyan data are slightly higher than WHO estimates (Figure 1.1.7) this may be explained by use of different population estimates.

Case notification rates were calculated using population estimates from Kenya's 2009 population and housing census for the years 2009-2013 and updated population estimates from the 1999 population and housing census for 2003-2008. Changes in population estimates between 2008 and 2009 did not appear to have a large effect on national level case notification rates.



Source: Ministry of Health TB database

Figure 1.1.6: TB notifications in Kenya, 2003-2013



Source: Ministry of Health TB database

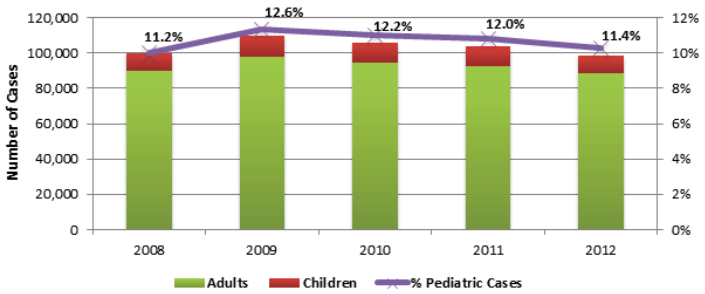
Figure 1.1.7: TB case notification rates in Kenya, 2000-2012

Table 1.1: Percentage change in all forms of notified TB case rate in Kenya, 2003-2012

Year	National TB Case Notification Rate (per 100,000)	% Change from Previous Year
2000	210.8	.
2001	233.2	10.60%
2002	254.8	9.30%
2003	287.3	12.70%
2004	309.6	7.80%
2005	308.5	-0.40%
2006	318.7	3.30%
2007	313.9	-1.50%
2008	288.0	-8.20%
2009	281.2	-2.40%
2010	263.6	-6.20%
2011	251.0	-4.80%
2012	230.9	-8.00%
2013	203.1	-12.10%

Source: Ministry of Health TB database

As shown in Figure 1.1.8, the percentage of new childhood and adult TB cases has remained consistent between 2008 – 2012. The ratios were fairly consistent with the rates that have been proposed by WHO in their standards and benchmarks tool which suggests that the data collected at the national level have a high level of external consistency if the percentage of children diagnosed with TB ranges between 5-15% in low - and middle-income countries.

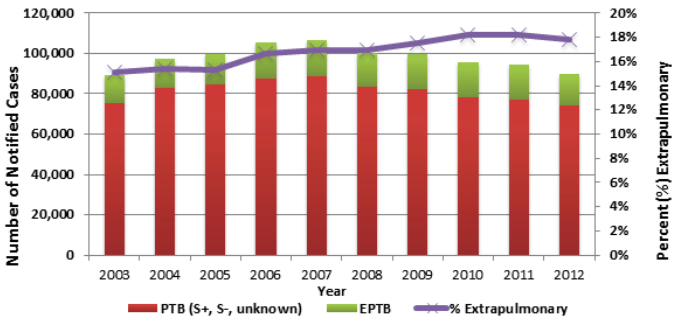


Source: Ministry of Health TB database

Figure 1.1.8: Comparison of TB cases in adults and children in Kenya, 2008-2012

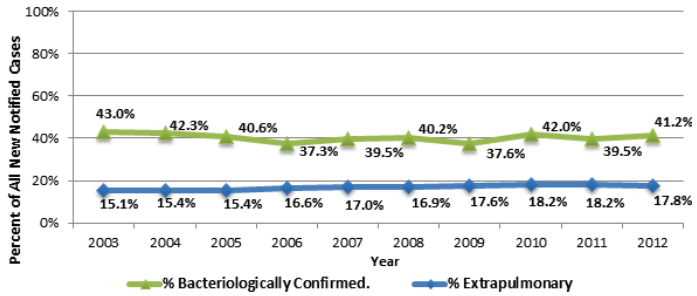
1.1.2.3 TB Cases by Disease Type

The relative numbers of new bacteriologically confirmed (smear positive) cases and extrapulmonary TB cases have remained fairly consistent over time, Figure 1.1.10. The percentage of new cases that are bacteriologically confirmed ranged from 37.3 to 43.0% between 2003 to 2012, fluctuating slightly from year-to-year. The percentage of new cases that are extrapulmonary, however, increased gradually since 2003 but maintained a narrower range: 15.1% to 18.2%, Figure 1.1.9. This slight increase could result in a reduction in TB transmission because there were fewer pulmonary TB cases in the community.



Source: Ministry of Health TB database

Figure 1.1.9: Comparison of new pulmonary and extrapulmonary TB in Kenya, 2003-2012

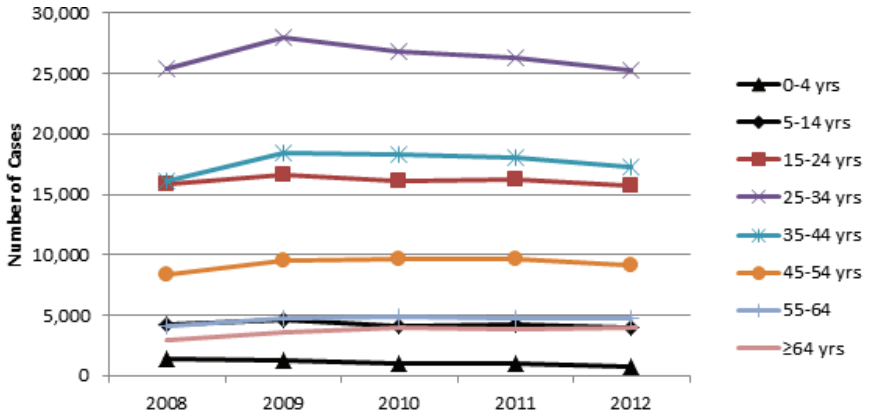


Source: Ministry of Health TB database

Figure 1.1.10: Proportions of new smear positive pulmonary and extrapulmonary TB in Kenya, 2003-2012

1.1.2.4 TB Cases by Age

The largest number of TB cases occurred among young adults as shown in Figure 1.1.11, with the most cases reported for adults aged 25-34 years, followed by adults aged 35-44 and those aged 15-24 years. The fewest cases were reported in children aged 0-4 years.



Source: Ministry of Health TB database

Figure 1.1.11: All forms of TB cases by age group in Kenya, 2008 - 2012

1.2 Statement of the Problem

In the 1950's and 60's as the health systems began to be stregthened and with effectiveness of improved sanitation systems, introduction of potent antibiotics, and initiation of mass vaccination programs created confidence that the war against infectious diseases would soon be won and most of the infectious diseases would be eliminated. As the burden of infectious diseases began to decline in the West more attention has turned to addressing cardiovascular disease and cancer. But infectious diseases have continued to be the major causes of suffering and mortality in developing countries. Moreover, infectious disease agents adapt and evolve, so that new infectious diseases have emerged and some existing disease have reemerged. This

coupled with lack of mathematical models interventions to guide targeted disease control measures in developing countries, the problem has been compounded. In Kenya, TB continues to remains a major public health concern with over 100,000 new caseas each year and the disease being the leading cause of mortality particularly among the HIV+ patients (DLTLD (2013)).

Kenya is a high TB burden country ranked $13th$ amongst the 22 countries contributing 80% of the Global TB case load, WHO (2012). Currently there is no model that can be used to predict the direction the epidemic will take given the current interventions continue being implemented. In addition, there is no prediction currently on the direction the epidemic will take given presence of other disease predisposing factors or determinants for tuberculosis disease. The aim of this study was to develop an epidemilogical model for tuberculosis disease and apply them in Kenya based on the surveillance data reported in the last 24 years.

1.3 Justification

The mathematical models developed will provide an in-depth examination of the inherent dynamics of TB epidemiology and its spread over time given the existing predisposing factors for the development of tuberculosis disease. Through the development of a stochastic branching process model, prediction of the TB epidemic will become apparent. The epidemiological modeling and understanding of the TB epidemic particularly in high HIV settings will act as a tool for advising TB control policies in Kenya and other developing countries. Although this model has been used in other infectious diseases, the proposed models have not been used for TB progression. It is anticipated that the model can be used to study the effects of the variation of interventions and assist in designing proper interventions since there

has been several interventions implemented by the government which have not been evaluated appropriately to the TB disease dynamics. The study also provides a considerable contribution to the tuberculosis epidemic modelling and prediction in Kenya.

1.4 Objectives

1.4.1 The General Objective

The general objective was to model tuberculosis epidemic progression.

1.4.2 The Specific Objectives

1. To develop epidemic models for TB progression;
2. To estimate the expected number of individuals with the disease at any given time t;
3. To formulate small area estimation models for TB progression;
4. To develop spatial-temporal models for the TB progression.

1.5 The Scope

This study covered the reported cases of TB in Kenya in the period from 1987 to 2014. All TB cases, that is, smear positive, smear negative and extra pulmonary TB. The case notification data were obtained from the national TB database while data on disease burden (incidence, prevalence and mortality) were obtained from WHO global database. The methodologies used include: the deterministic models, stochastic models, branching processes, small area estimation for mapping.

Chapter 2

LITERATURE REVIEW

2.1 Introduction

In modeling the nature of infectious diseases at the population level two types of models are useful: these are stochastic and deterministic models. Stochastic models rely on chance variation in risks of exposure, disease, and other factors. They provide much more insight into an individual-level modeling, taking into consideration small population size where every individual plays an important role in the model i.e. the individuals in a way contribute to the progression of the infections in a random manner. They are more useful in scenarios where infection is as a result of contacts with another individual. Hence, they are used when known heterogeneities are important as in small or isolated populations. Stochastic models have several advantages. More specifically, they allow close watching of each individual in the population on a chance basis. They, however, can be laborious to set up as they need many simulations to yield useful predictions. These models can become mathematically very complex and do not contribute to an explanation of the dynamics. Deterministic models, also known as compartmental models, attempt to describe and explain what happens on the average at the population scale. They fit well to large populations. These models categorize individuals into different subgroups (compartments). There is a large body of literature which has been developed in the field of epidemic modelling in an attempt to address the behaviour of different disease spread.

2.2 Predictive Model

A fine predictive model should be built as far as possible in a rigorous mechanistic way starting from the mechanism of exposure/infection of each individual and taking into account their variability (Jacob (2010)). Bernoulli (1766), was one of the most foremost mathematicians to attempt to develop a model on the effects of the disease in the population. The work developed a deterministic model to show that inoculation with a mild form of the small pox virus would reduce the death rate of the population of France, Bernoulli's work was reviewed extensively by Klaus and Heesterbeek (2002) in their study, where they revisited Bernoulli's work who has been credited as having developed the first compartmental epidemic model. The main objective of this work was to calculate the gain in life expectancy at birth if small pox was to be eliminated as a cause of death. The main drive of this work was on prolongation of life expectancy at any age since the focus then was on annuities and any change of life expectancy would have had immediate financial impact in the insurance business. Bernoulli's work has been further credited for developing a method for dealing with competing risks (D'Alembert (1761)). In Bernoulli's model the population is divided into two compartments: Susceptible that is, those who have not been infected and immunes, being those who have been immunized for the rest of their lives after one infection. In this model certain parameters are considered namely; death rate due to all causes except due to infection and was denoted by $\mu(a)$ and the force of infection, $\lambda(a)$, which is the rate according to which susceptible are infected. In the model, there is a proportion $s(a)$ who survive to become immune with the rest $c(a) = 1 - s(a)$ dying due to the infection. If $u(a)$ denotes the probability for a new born individual to be alive and susceptible at age a, this probability satifies the differential equation.

$$\frac{du}{da} = -\left[\lambda(a) + \mu(a)\right] u \tag{2.2.1}$$

with initial condictions $u(0) = 1$.

The probability $w(a)$ to be imunne and alive is given by

$$\frac{dw}{da} = \left[1 - c(a)\right] \lambda(a)\mu(a) - \mu(a)w \tag{2.2.2}$$

with initial condictions $w(0) = 0$.

The solutions to these equations were obtained as:

$$u(a) = \exp\{-[\Lambda(a) + \mathrm{M}(a)\} \tag{2.2.3}$$

$$w(a) = e^{-m(a)} \int_0^a [1 - c(\tau)]\lambda(\tau)e^{-\Lambda(\tau)}d\tau \tag{2.2.4}$$

where

$$\Lambda(a) = \int_0^a \lambda(\tau)d\tau \tag{2.2.5}$$

and

$$\mathrm{M}(a) = \int_0^a \mu(\tau)d\tau \tag{2.2.6}$$

If $l(a)$ denotes the probability to survive age to a then, $l(a) = u(a) + w(a)$. This is because the two states susceptible and immune are complementatry to each other. The survival function then in the population without small pox would then be

$$l_0(a) = e^{-M(a)} \tag{2.2.7}$$

while the survival function in the presence of small pox was given as

$$l(a) = l_0(a)[e^{-\Lambda(a)} + \int_0^a [1 - c(\tau)]\lambda(\tau)e^{\Lambda(\tau)}d\tau]. \tag{2.2.8}$$

With the presentation of Bernoulli's work, Klaus and Heesterbeek (2002) gave an account of a rejoinder to Bernoulli's work by D'Alembert (1761) who gave a critique of Bernoulli's compartmental model by developing a generalized model which is not restricted to an immunizing disease. This model starts by letting $\mu_d(a)$ denote the force of mortality due to some disease d. The force of mortality due to other causes was denoted by $\mu(a)$. Letting $\phi_d(a)$ denote the rate at which deaths due to particular cause for individuals who die at age a, then

$$\phi_d(a) = \mu_d(a)l(a). \tag{2.2.9}$$

Thus, if survival function is known, force of mortality for a particular disease can be calculated by dividing $\phi_d(a)$ by $l(a)$. The survival function without particular cause of death is thus as shown in Equation 2.2.10.

$$l_0(a) = e^{-M(a)} = l(a) \exp\left(\int_0^a \mu_d(\tau)d(\tau)\right) \tag{2.2.10}$$

D'Alembert (1761)method is viewed as being more desirable if the task at hand is to calculate the survival function after eliminating a particular cause of death. While Bernoulli's method provides more insight for the interpretation of infectious disease data, an early reference to the non-linearity of epidemic models is made in a study

by Hamer (1906). Hamer (1906) presupposed that the probability of an infection in the next period of time (in a discrete time model) was proportional to the number of infectious individuals multiplied by the number of susceptible individuals which is an assumption of the mass action law. One of the first epidemic models to incorporate the randomness observed in real life epidemics was given by McKendrick (1926). This model is a stochastic continuous time version of the Deterministic General Epidemic Model. With a deterministic models the values of the parameters and variables in the model are known with certainity, as there are no random fluctuations in value. The system can be completely defined at any time using the initial conditions specified. Another early discrete-time model is the chain Binomial model of Reed and Frost, Bailey (1975), in which the number of infectious cases to appear in the next time unit follows a Binomial distribution with the probability of infection dependent on the number of infectious cases in the current time unit. *Bartlett. (1949)* studied McKendrick's model, that stochastic models in continuous time were examined more extensively. Since then, research has been directed towards the study of a wide variety of models, and their statistical analysis.

Several epidemic models have been used to model recent disease outbreaks Siettos et al. (2015). The study developed an agent based model using small-world network constructed using Watts and Strogatz (S&W) algorithm with a variable edge density. The edge density being defined as the number of links divided by the total possible links attributed to the agent. Further work on epidemic modelling was illustrated by Keeling and Danon (2009) who showed that Mathematical models allow us to extrapolate from current information about the state and progress of an outbreak, to predict the future and, most importantly, to quantify the uncertainty in the predictions. It also illustrated the mathematical epidemic principles by utilizing

the data form the Influenza A virus containing Haemagglutinin and Neauraminidase glycoproteins (H1N1) epidemic. Given the wealth of information mathematical models provided in the last two years alone due to disease outbreaks in particular Ebola virus disease (EVD) there is a wealth of literature utilizing epidemic models methodology as described in the work by (Fasina et al. (2014), Muyembe-Tamfum et al. (2012), Shuaib et al. (2014), Tomori (2015)) among others. In the preceeding work, a review of some of the models used is given.

2.2.1 The Susceptible Infective Epidemic Model

There has been a long history of mathematical epidemiology modeling dating back to the 18th century, Bernoulli (1766). It was until the early 20th century that there was popular dynamic systems approaches being implemented, Anderson and May (1991a), Hethcote (2000). Over the last decades, epidemiologists have used epidemic models to obtain a purview of disease epidemic processes and dynamics. One such epidemic model is Susceptible-Infective model I, this model was first formulated by Bernoulli (1766). In this model only two compartments exist i.e. the susceptible and infective compartment. A number of authors have reviewed and utilized the I models. Watson and Galton (1968) considers an SI model with constant population size and two kinds of susceptibles having very different infection rates. Then he gives the exact solution in an implicit form and also derives an approximation to the solution which permits simple estimates of the infection rates. There have also been efforts to develop numerical methods to approximate the solution of McKendrick (1926) models. Lewis and Glass (1991) uses discrete models to qualitatively capture all the dynamics of these models. He develops two discrete models based on the Lewis-Glass hypercube projection method Lewis and Glass (1991) and the Laubenbacher-Stigler

polynomial interpolation method Laubenbacher and Stigler (2004) to find discrete approximations of the dynamics of the associated McKendrick (1926) model which are qualitatively similar to the dynamics of the continuous model. This model is the basic epidemic model and assumes that all the susceptible will enventually become infected and at that point the epidemic comes to an end.

2.2.2 The Susceptible Infective Recovery Epidemic Model

The standard SIR (Susceptible Infective Recovered) epidemic model is a well studied model for the spread of an infectious disease through a fixed population. This section sets out how the model is constructed, and then generalise it to have multiple types of communities with different rates of predisposing factors using the structured model developed by Anderson and Tom (2000).

Glass et al. (2007) applied Markov Chain Monte Carlo methods to a discrete time branching process model with two types of infectious individual: diagnosed and hidden. The study obtained estimates of the distribution of the number of infectious individuals in the $(t+1)th$ generation of the infection, given that the number of diagnosed individuals has been observed for the first t generations. Since it is necessary to know to which generation an infectious form belongs, the method by Glass et al. (2007) can only be used for data on the very early generations of infection. However, if the reproduction number is close to one, the highly stochastic first phase of an epidemic may last for many more generations than this. Eichner and Dietz (1996) simulated a continuous time Markov process model for the spread of polio with vaccination, in order to find the probability that the disease has died out given a case free period of length t. They calculated this probability by simulating the epidemic and recording the proportion of case free periods with length greater than t in which

the disease had died out at the end. All of the simulations begin from a state of endemic polio in a completely unvaccinated population of 200,000 individuals. They found that a case free period of 3 years gives a 95% probability of eradication. O'Neill and Roberts (1999) described Markov Chain Monte Carlo methods for estimating the parameters for the Markov SIR epidemic model in which the removal times are observed until time t. Branching processes play a fundamental role in epidemic theory, underpinning our understanding of the threshold behaviour of epidemics and the calculation of the critical vaccination threshold as well as providing a simple model for the early stages of epidemic spread. Thus, much recent work on complex epidemic models making use of branching process approximations have been documented by Marschner (1992), Ball and Donnelly (1995), Clancy and O'Neill (1998), Caraco et al. (1998), Muller and Kirkilionis (2000), Ball and Lyne (2001). More directly, branching processes have been proposed as statistical models on which to base inferences about the reproduction number R, represented by the offspring mean (Becker (1974), Farrington and Grant (1999), Yanev and Tsokos (1999)). Nevertheless, in contrast to the wide spread use of other types of models, particularly the deterministic models of Anderson and May (1991b), statistical models based on branching processes are seldom used in practice for infectious disease control.

2.2.3 Susceptible Exposed Infective Recovery Epidemic Model

A common class of epidemiological models developed for the spread of infectious diseases is the Kermack-McKendrick model and its variations McKendrick (1926). These models are represented as systems of ordinary differential equations. Many variations of the original Kermack-McKendrick model have been described in many epidemiological works and it typically using names based on acronyms of the involved

compartments. Some of these models incorporate a fourth class (Exposed) within the population, accounting for diseases with a latent period. Also some of these models account for non-permanent immunity, thus allowing individuals to again move to the susceptible compartment. Hethcote (2000) investigated a variety of mathematical models whose classes and interactions are a subset of the SEIRS model. In the SEIR model, the population is usually categorized into four different compartments (or epidemiological classes) at each point of time. These compartments are susceptible, exposed, infective, and recovered. The susceptible compartment contains those individuals who do not have any type of immunity to the disease and can become infected. When there is an adequate contact of a susceptible with an infective so that transmission occurs, the susceptible enters the exposed compartment of those in the latent period, who are infected but not yet infectious. After the latent period ends, the individual enters the infective compartment of those who are infectious in the sense that they are capable of transmitting the infection. When the infectious period ends, the individual enters the recovered compartment consisting of those with permanent infection-acquired immunity. This model has been used extensively by a number of authors (Hethcote (1976), Li et al. (1999), Anderson and Tom (2000)).

2.3 Branching Processes

The major characteristic of biological populations is that individuals undergo birth and death and that individuals carry information passed on from their parents at birth. Furthermore, there is a randomness in this process in that the number of births that an individual gives rise to is in general not deterministic but random. Branching processes model this process under simplifying assumptions but nevertheless provide the starting point for the modelling and analysis of such populations.

Branching process is a Markov process that models a population in which each individual in generation n produces some random number of individuals in generation $n+1$, according to a fixed probability distribution that does not vary from individual to individual. Branching processes are used to model reproduction; for example, the individuals might correspond to bacteria, each of which generates 0, 1, or 2 offspring with some probability in a single time unit. Branching processes can also be used to model other systems with similar dynamics for example the spread of disease in a community. Genarally, Galton-Watson processes Watson and Galton (1875) are used to describe a family tree from the perspective of generations. But in statistical epidemiology one may be more interested in the real time development of the number of infective individuals. Usually it is not possible to observe the size of an "infection generation", but it may be possible to observe the number of infective individuals at a certain time.

Branching processes in discrete time were first used by Galton and Watson in order to model the survival of surnames Mode (1971), but they have come to have many other applications, particularly in biology Jagers (1975).

At time zero, i initial ancestors begin their lifetimes. During each of its lifetime, each individual in the process will independently gives birth to offspring at the points of a Poisson process with rate β, and these offspring start their own lifetimes immediately. Each lifetime has length distributed according to the random variable Y_i, all instances of which are independent. Let $I(t)$ denote the number of individuals alive in the branching process at time t, so that overall, births occur at a rate $\beta I(t)$.

2.4 Small Area Estimation

Small area has been used to refer to a population for which reliable statistics of interest cannot be produced due to certain limitations in the available data. The small areas used could be provincial, county, sub-county, demographic groups (e.g. age $\times$ sex), a demographic group within a geographic region. The demand for data or estimates for such small areas has greatly increased during the past few years. This is primarily due to increase in usefulness of these data in government policy and program development, allocation of various funds and planning, Rao (2003).

Small area estimation (SAE) has in the recent past gained prominence and is a topic of great importance due to the growing demand for reliable small area statistics even when only very small samples are available for these areas. Over the years, a number of statisticians have introduced vigorous programs to meet this demand. Extensive research on the theoretical and practical aspects of SAE is carried out and many international conferences and workshops are held in order to share the results of this research effort. Interest in small area estimation methods has further enhanced in recent years due to the tendency of many countries to base future censuses on administrative record systems. Recognizing the inaccuracies of the administrative data and the fact that even the richest records cannot cover all the detailed information required for small census tracts, the idea is to test, correct and supplement the administrative information by sample data. In the more recent past, it has been used in Nepal to estimate poverty Nepal Central Bureau of Statistics (2013). SAE can be used in the study of disease epidemiology by establishing relative risks given the different variable effects.

The models used for describing the distribution of an observation in space and time

are usually formulated within a hierarchical Bayesian framework with two main approaches namely empirical and full Bayes. Empirical Bayes approach provides parameter estimates by maximizing their posterior distribution using penalized quasi-likelihood (PQL) techniques while the full Bayes provides the posterior distribution of the target parameters by assigning hyperpriors which take care of model uncertainties Gomez and Lopez (2006). In this study we consider the Bayesian Hierarchical Generalized Linear Mixed Models (BHGLMMs) which are used in small area estimation because of their ability to incorporate multiple levels of model dependencies (Cnaan et al. (1997) and Fong et al. (2009)). The BHGLMMs falls in subclass of structured additive regression (STAR) models known as latent Gaussian models and their response variable is usually non Gaussian and belongs to an exponential family (Umlauf et al. (2012); Klinker (2010)). The posterior marginals of the latent Gaussian models are not available in closed form for non-Gaussian observation models (Rue et al. (2009)). In such models the common approach to inference is the Markov Chain Monte Carlo (MCMC) which exhibits poor performance in terms of convergence and computation time. Integrated nested Laplace approximation (INLA) developed by Rue et al. (2009) and based on nested Laplace approximations is a new approach for Bayesian inference on latent Gaussian models and it has an excellent performance in terms of good accuracy and reduced computational time (Grilli et al. (2014), Taylor and Diggle (2013), Cameletti et al. (2013) and Rue et al. (2009)) .

Chapter 3

METHODOLOGY

3.1 Introduction

Modelling is basically about describing the relationship between a response variable and one or more other variables (explanatory variables) in a simplified way. First, explore the data to detect patterns and relationships after which you visualize the relationship by fitting a curve to the data. This is the empirical approach in which you choose a model based on the data. Alternatively, if there exist a well-established theory, you choose the model based on the context and see if it fits the data well. The deterministic growth models in biology can be used to model consumer behaviour in economics and the majority of models in physics and chemistry.

Epidemic modelling has three main aims. The first is to understand the transmission mechanism of the disease. The essential part is a mathematical structure (equations that gives threshold values and other constants which can be used to describe the behavior of the disease). The second aim is to predict the future direction of the epidemic. The third is to understand how control measures can be put in place to curtail the spread of the epidemic (treatment education, immunization, isolation etc) and lastly to develop a reliable model and generate sound and sensitive predictions. So it is important to validate models by checking whether they fit the observed data. In deterministic models, population size of the compartments are assumed to be functions of discrete times $t = 0, 1, 2, ...$ or differentiable functions of continuous time $t \geq 0$. This enables derivation of a set of difference-differential equations governing the process. The evolution of this process is deterministic in the sense that no randomness is allowed. In order to make a model for a disease in a population, the population is divided in to few classes and we study the change of their numbers

in time. The choice of which compartments to use in the model depends on the characteristics of a particular disease and the purpose of the model.

The threshold for many epidemiological models is the basic reproduction number R_0 . It reflects the average number of infected people when one infected individual is introduced into a population where everyone is susceptible. It is a threshold quantity which determines whether the epidemic will occur or not. So if the number of infected individuals is higher than this value, then the epidemic spreads across the population.

3.2 Mathematical Formulation of the Epidemiological Models

In this section, we present standard epidemiological models which are used to study the spread of disease in host populations under different standard conditions. We utilize some standard notations as used by, Hethcote (2000). If $S(t)$ denotes the number of susceptibles at time t, $I(t)$ denotes the number of infectives at time t, N denotes the population size, then we can write: $s(t) = \frac{S(t)}{N}$ and $i(t) = \frac{I(t)}{N}$ where $s(t)$ and $i(t)$ are the fractions of respective populations. If β is the average number of contacts (sufficient for transmission) of a person per unit time, then $\beta i = \frac{\beta I}{N}$ is average number of contacts with infectives per unit time of one susceptible. Therefore,

$$\beta N = \frac{\beta I}{N} S \tag{3.2.1}$$

is the number of new cases per unit time (because $S = Ns$). In this case the horizontal incidence is called standard incidence. The Simple Mass Action Principle

$$\eta IS = \eta(Ni)(Ns) \tag{3.2.2}$$

with η as a mass action coeffient, is a standard for horizontal incidence. Comparing the L.H.S of Equation 3.2.1 to the R.H.S of Equation 3.2.2 we get $\eta N = \beta$. So contact rate β increases linearly with population size. Therefore we can write: $\frac{\eta^v N^{SI}}{N}$ is the standard incidence if $v = 0$ and it is a mass action incidence if $v = 1$.

In the development of the models we considered two probability models: the first is the probability that the epidemic has died out at time t given that no cases have been observed in the interval (s, t). This is calculated for two models: a small population epidemic model and its large population branching process approximation. The second is the probability that the epidemic is over at the end of the first time period of length t in which no cases are observed.

3.2.1 Susceptible Infective Model

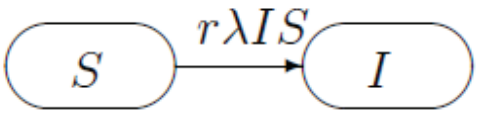


Figure 3.2.1: Susceptible Infective Model

The SI model is the simplest one among the epidemic models. This model was originally formulated by McKendrick (1926). That is why it is also called the Simple Model. In this model, the population is divided into just two compartments namely; the susceptible compartment $S(t)$ and the infectious compartment $I(t)$. We do assume the disease to be highly infectious but not serious, which means that the infectives remain in contact with susceptibles for all time $t \geq 0$. It is assumed that the infectives continue to spread the disease till the end of the epidemic, the population size to be constant $(S(t) + I(t) = N)$ and homogeneous mixing of population. Infection rate is proportional to the number of infectives, i.e.

$$\beta = r\lambda I$$

where r =fraction of those recovered with acquired immunity, and λ =force of infection.

The pair of ordinary differential equations for this model are

$$\frac{dS(t)}{dt} = -r\lambda I(t)S(t) \tag{3.2.3}$$

$$\frac{dI(t)}{dt} = r\lambda I(t)S(t) \tag{3.2.4}$$

where

$N = S(t) + I(t)$

$S(t) = N - I(t)$

and therefore we get

$$\frac{dI(t)}{dt} = r\lambda I(t)\left[N - I(t)\right], \tag{3.2.5}$$

which is known as the *logistic growth equation.*

The translation between the susceptibles and infectives $S \rightarrow T$ denoted by the quantity $r\lambda IS$ which is a separable linear ordinary differential equation given by:

$$\frac{1}{I\left(t\right)\left(N - I\left(t\right)\right)}\frac{dI}{dt} = r\lambda \tag{3.2.6}$$

Integrating both sides of Equation 3.2.6 with respect to t we obtain:

$$
\begin{aligned}
\int_0^t \frac{1}{I(t)\left(N-I(t)\right)}\frac{dI}{dt}dt &= \int_0^t r\lambda dt \\
\int_{I(0)}^{I(t)} \frac{1}{u\left(N-u\right)}du &= \int_0^t r\lambda dt \\
\frac{1}{N}\int_{I(0)}^{I(t)} \left[\frac{1}{u}+\frac{1}{N-u}\right]du &= \int_0^t r\lambda dt \\
\ln\left(u\right)-\ln\left(N-u\right)\big|_{u=I(0)}^{I(t)} &= r\lambda Nt \\
\left[\ln I(t)-\ln\left(N-I(t)\right)\right]-\left[\ln I(0)-\ln\left(N-I(0)\right)\right] &= r\lambda Nt \\
\ln\frac{I(t)}{\left(N-I(t)\right)}\cdot\frac{\left(N-I(0)\right)}{I(0)} &= r\lambda Nt \\
e^{r\lambda Nt}=\frac{I(t)\left(N-I(0)\right)}{I(0)\left(N-I(t)\right)} &= \frac{NI(t)-I(t)I(0)}{NI(0)-I(t)I(0)} \\
e^{r\lambda Nt}\left[NI(0)-I(t)I(0)\right] &= NI(t)-I(t)I(0) \\
Ne^{r\lambda Nt}I(0)-e^{r\lambda Nt}I(t)I(0) &= NI(t)-I(t)I(0) \\
I(t)\left[I(0)-N-I(0)e^{r\lambda Nt}\right] &= Ne^{r\lambda Nt}I(0) \\
I(t) &= \frac{Ne^{r\lambda Nt}I(0)}{\left[I(0)-N-I(0)e^{r\lambda Nt}\right]} \\
I(t) &= \frac{I(0)N}{I(0)+\left(N-I(0)\right)e^{-r\lambda Nt}}.
\end{aligned}
$$

As we can see I approaches N asymptotically with $t\to\infty$. Therefore, every susceptible joins the infectious compartment in this model, everybody becomes infected which is, in fact, the "end" of the epidemic in mathematical sense. This model has the challenge that it infers that in every epidemic all the individuals who are susceptible will eventually become infectious which is not true in virtually almost all epidemics since data obtained from most surveys do not always fit this model.

3.2.2 Susceptible Infective Recovery Models

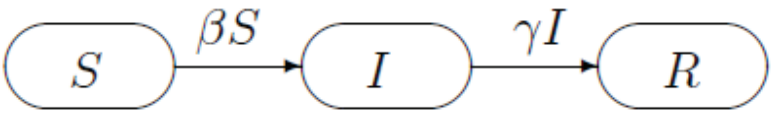


Figure 3.2.2: Susceptible Infective Recovered Model

There are two SIR models which were formulated by, McKendrick (1926). The two models are described as the foundations for the mordern mathematical epidemiology and are currently still widely used in practice. The models describe either an epidemic(that is a rapid outbreak of an infectious disease) or an endemic (a disease present in the population for a long period of time where the class of susceptibles is being nourished by new income from births or recovered individuals who lost their temporal immunity).

In these models the assumption is that the population size is large and constant (except for death from the disease) and homogeneously mixing for continuous time $t \geq 0$. Any person who has completely recovered from the disease acquired permanent immunity and the disease has a very short incubation period (so an individual who contracts the disease becomes infective immediately afterwards). This enables the division of the population into three compartments: $S(t)$ - susceptibles, $I(t)$ - infectives, $R(t)$ - recovered. Infection rate is proportional to the number of infectives, This can be illustrated in the equation below:

$$\beta = r\lambda I$$

where r =fraction of those recovered with acquired immunity, and λ =force of infection.

The system of equations governing this model (SIR) according to McKendrick (1926) are given by:

$$\frac{dS(t)}{dt} = -r\lambda S(t)I(t) \tag{3.2.7}$$

$$\frac{dI(t)}{dt} = r\lambda S(t)I(t) - \gamma I(t) \tag{3.2.8}$$

$$\frac{dR(t)}{dt} = \gamma I(t) \tag{3.2.9}$$

$$\begin{cases} S\left(0\right) = S_0 > 0 \\ I\left(0\right) = I_0 > 0 \\ R\left(0\right) = R_0 = 0 \end{cases} \tag{3.2.10}$$

It can be deduced that the derivative function $\frac{d}{dt}\left[S(t) + I(t) + R(t)\right] = 0$, it can therefore be seen that the size of the population remains constant as shown in the equation 3.2.11

$$S(t) + I(t) + R(t) = N \tag{3.2.11}$$

From the linear system of equations(3.2.7 - 3.2.10), we can obtain:

$$
\begin{aligned}
\frac{dI}{dS} &= \frac{\gamma - r\lambda S}{r\lambda S} = \frac{\gamma}{r\lambda S} - 1 \\
dI &= \left[\frac{\gamma}{r\lambda S} - 1\right] dS \\
\int_0^t 1dI &= \int_0^t \left[\frac{\gamma}{r\lambda S} - 1\right] dS \\
I|_{t=0}^{t} &= \left[\frac{\gamma}{r\lambda} \ln S - S\right]_{t=0}^{t} \\
I(t) - I(0) &= \frac{\gamma}{r\lambda} \ln S(t) - \frac{\gamma}{r\lambda} \ln S_0 - S(t) + S_0 \\
I(t) + S(t) - \frac{\gamma}{r\lambda} \ln S(t) &= S_0 + I_0 - \frac{\gamma}{r\lambda} \ln S_0 = constant
\end{aligned}
\tag{3.2.12}
$$

Thus the final equation can be obtained as shown in the equation 3.2.13:

$$
I(t) = S_0 + I_0 - S(t) + \rho \ln \frac{S(t)}{S_0} \tag{3.2.13}
$$

where $\rho \equiv \frac{\gamma}{\beta} \equiv \frac{\gamma}{r\lambda}$. The parameter ρ is called the relative removal rate.

3.2.3 Model Modification by an Intervention

The description of models that have been introduced in sections 3.2.1 and 3.2.2 have not taken into account factors that would affect or impede the progress of the disease, like public education, nutritional, comorbidity with other diseases such as HIV, diabetes, vaccination etc. We now consider a closed, homogeneous and homogeneously mixing population of individuals partitioned at any time $t \geq 0$ into three categories: susceptible, infective and removed as first described in Anderson and Tom (2000). Susceptible individuals become infectives when they come into

contact with an infective individual. In the study they assumed that the newly infected individual undergoes a latent period (during which they are infected but not infectious) but instead, they become infective when they develop the disease and instantly begin their infectious period. When their infectious period is over, an individual infective moves into the removed category - they either die, put on treatment or become immune to the infection - and can play no further part in the epidemic. In this section consider mathematical modelling by three families of non-negative, integer valued random variables indexed by a continuous time parameter. The number of the susceptible is modelled by the process $S(t) : t \geq 0$, with $S(0)$ denoted by n. It is assumed that no individuals can enter the susceptible category during an epidemic, the number of susceptibles is non-increasing. This is so because those who are susceptible once they get infected are removed from the susceptible category. The number of infectives is represented by the process $I(t) : t \geq 0$, and $I(0) = i$. Finally, the number of removed individuals is modelled by the process $R(t) : t \geq 0$ and without loss of generality, $R(0) = 0$. Since no individuals may leave the removed category once they have entered it, $R(t)$ is non-decreasing. Since the population is fixed, for all $t \geq 0, S(t) + I(t) + R(t) = n + i$. Any pair of individuals make contact at the points of a Poisson process with rate $\frac{\beta}{n}$, and all of the $(n + i/2)$ Poisson processes are assumed to be independent. This implies that at time t, infections occur at a rate of $\frac{\beta}{n}S(t)I(t)$. The non-negative constant beta encapsulates both the infectiousness of the disease and the susceptibility of the population, as a 'contact' is defined to be a meeting between two individuals sufficient to transfer the disease from an infective to a susceptible. Alternatively, contacts could be defined to occur at a rate $\frac{\beta}{n}^1(n + i)S(t)I(t)$. In this formulation when an infective makes a contact, they choose an individual at random from the whole population to contact,

whereas in our formulation contacts occur at $\frac{n}{n+i}$ times the rate, but contacts are chosen from only the initial susceptibles. This formulation is used as it highlights the connection with the branching process approximation: when $S(t) \approx S(0) = n$ then infections occur at rate $\beta I(t)$, as in the branching process. At the instant of infection, the new infective is allocated an infectious period according to a non-negative random variable TI. All instances of the random variable TI are assumed to be independent of each other, and of the Poisson processes governing the infections. When the infectious period distribution TI has an exponential distribution, the SIR epidemic model has been referred to as the General Stochastic Epidemic model, and in this case the process $\{S(t); I(t) : t \geq 0\}$ has the Markov property.

If we assume that the suscesptibles have sound knowledge on how to protect themselves at a particular rate v which could be through any intervention e.g. vaccination, proportional to their number, then we have a modified system of equations as:

$$\begin{aligned} \frac{dS}{dt} &= -r\lambda SI - vS \\ \frac{dI}{dt} &= r\lambda SI - \gamma I. \end{aligned} \tag{3.2.14}$$

We then can obtain the system of equations as:

$$\begin{aligned}
\frac{dI}{dS} = \frac{r\lambda SI - \gamma I}{-r\lambda SI - vS} = \frac{\gamma I - r\lambda SI}{r\lambda SI + vS} &= \frac{I(\gamma - r\lambda S)}{S(r\lambda I + v)} \\
\int_0^I \left[r\lambda + \frac{v}{I}\right] dI &= \int_0^I \left[\frac{\gamma}{S} - r\lambda\right] dS \\
\left[r\lambda I + v \ln I\right]_0^I &= \left[\gamma \ln S - r\lambda S\right]_{S(I_0)}^{S(I)} \\
r\lambda I(S) + v \ln \frac{I(S)}{I_0} &= \gamma \ln \frac{S}{S_0} - r\lambda S + constant \qquad (3.2.15)
\end{aligned}$$

where the constant is equal to $r\lambda S_0 + r\lambda I_0$.

If we take into account the recovered group in the epidemiological model, it can be shown that $S(t)$ tends to zero as t tends to infinity for every solution of $(S(t), I(t))$ of the equations, that is,

$$\begin{aligned}
\frac{dS}{dR} &= \frac{-r\lambda SI - vS}{\gamma I} = \frac{S(-r\lambda I - v)}{\gamma I} \\
\frac{1}{S} dS &= \frac{-(r\lambda I + v)}{\gamma I} dR \\
\int_0^t \frac{1}{S} dS &= \int_0^t \frac{-(r\lambda I + v)}{\gamma I} dR \\
\ln S|_{t=0}^t &= \left.\frac{-(r\lambda I + v)}{\gamma I} R\right|_{t=0}^t \\
\ln \frac{S(t)}{S_0} &= -\frac{r\lambda I - v}{\gamma I}(R(t) - R_0) \\
S(t) &= S_0 e^{-\frac{r\lambda I - v}{\gamma I}(R(t) - R_0)}, \qquad (3.2.16)
\end{aligned}$$

so we have Equation 3.2.16 as what approaches zero when we get to infinity with time.

The interventions being put in place to decelerate the progression of the disease for example a massive ART program to ensure that all those who are HIV positive and in need of ARVs access them could dramatically alter the number of those who become infective with tuberculosis. If we assume that the rate of ARV uptake is v and it is proportional to the product of their numbers and the square of the infectives $I(t)$

$$\frac{dS}{dt} = -r\lambda SI - vSI^2 \tag{3.2.17}$$

$$\frac{dI}{dt} = I\left(r\lambda S - \gamma\right) \tag{3.2.18}$$

Dividing Equation 3.2.17 by 3.2.18 we obtain the system of equations as:

$$\begin{aligned}
\frac{dI}{dS} = \frac{I\left(r\lambda S - \gamma\right)}{-r\lambda SI - vSI^2} = \frac{r\lambda S - \gamma}{-r\lambda S - vSI} &= \frac{\gamma - r\lambda S}{S\left(r\lambda + vI\right)} \\
\left(vI + r\lambda\right) dI &= \frac{\gamma - r\lambda S}{S} dS \\
\left[\frac{vI^2}{2} + r\lambda I\right]_{I(S_0)}^{I(S)} &= \left[\gamma \ln S - r\lambda S\right]_{S_0}^{S}
\end{aligned} \tag{3.2.19}$$

$$\begin{aligned}
\frac{vI^2\left(S\right)}{2} + r\lambda I\left(S\right) - \frac{vI_0^2}{2} - r\lambda I_0 &= \gamma \ln S - r\lambda S - \gamma \ln S_0 + r\lambda S_0 \\
\frac{vI^2\left(S\right)}{2} + r\lambda I\left(S\right) &= \gamma \ln S - r\lambda S + \underbrace{\frac{vI_0^2}{2} + r\lambda I_0 - \gamma \ln S_0 + r\lambda S_0}_{=const}
\end{aligned}$$

If we apply the same procedure we obtain the equations indicated below. It shows that at the end of the epidemic outbreak there will be some individuals who will still

be susceptible in the population:

$$\frac{dS}{dR} = \frac{-r\lambda SI - vSI^2}{\gamma I} = \frac{-r\lambda S - vSI}{\gamma} = \frac{S\left(-r\lambda - vI\right)}{\gamma}$$

$$\begin{aligned} \frac{1}{S}dS &= \frac{-r\lambda - vI}{\gamma}dR \\ \left[\ln S\right]_{S_0}^{S(t)} &= \left[e^{\frac{-r\lambda - vI}{\gamma}R}\right]_{t_0}^{t} \end{aligned}$$

Therefore,

$$S\left(t\right) = S_0 e^{\frac{-r\lambda - vI}{\gamma}\underbrace{\left[R\left(t\right) - R_0\right]}_{<N - vS_0}} \geqslant S_0 e^{\frac{-r\lambda - vI}{\gamma}\left[N - vS_0\right]} > 0.$$

We now proceed to identify an important tool for comparing the epidemic strikes and its defined as an *intensity i*. We thus define the fraction of the total number of the susceptibles that eventually contract the disease. Hence,

$$i = \frac{I_0 + \left(S_0 - S_\infty\right)}{S_0}$$

where S_∞ is the root of the equation $S = S_0 e^{\frac{S - S_0 - I_0}{\rho}}$. We have already showed that $S(R) = S_0 e^{-\frac{R(t)}{\rho}}$ and we know that $R(\infty)$ is equal to the total population without the left susceptibles and infectives (this term is zero), so we have $R_\infty = S_0 + I_0 - S - 0$ where $S_0 + I_0 = N$. Recall that $S_\infty = S_0 e^{(S - S_0 - I_0)/\rho}$.

3.2.4 The General Endemic Model

Epidemic ceases to exist due to depletion of susceptibles below the threshold value $\rho = \frac{N}{R_0}$. Therefore, in case of an endemic presence, the susceptibles have to be kept over this value. There are two ways of achieving this: the first one is the case of non-immunizing diseases, the other is taking into consideration the vital dynamics of the system (births and deaths).

The former is so called SIS model with the system of equations:

$$\frac{dS}{dt} = -r\lambda SI + \gamma I \tag{3.2.20}$$

$$\frac{dI}{dt} = r\lambda SI - \gamma I \tag{3.2.21}$$

with initial conditions:

$$S(0) = S_0$$

$$I(0) = I_0$$

The R-compartment is missing because an infective individual goes back to the class of susceptibles after recovery. It is so due to the fact, that this individual can not acquire immunity for the disease. This is one possibility to get an endemic model but it is not the SIR model.

The general endemic Model is the SIR model with vital dynamics given by the system of equations:

$$\frac{dS}{dt} = \mu N - \mu S - \beta IS \tag{3.2.22}$$

$$\frac{dI}{dt} = \beta IS - \gamma I - \mu I \tag{3.2.23}$$

$$\frac{dR}{dt} = \gamma I - \mu R \tag{3.2.24}$$

$$S(0) = S_0 \geq 0, I(0) = I_0 \geq 0, R_0 \geq 0 \tag{3.2.25}$$

Again, it is assumed that for the population size N holds the relation: $N = S(t)+I(t)+R(t)$. This model is almost the same as general epidemic model, except that it has an inflow of newborns into the susceptible compartment and we also assume deaths (vital dynamics). The parameter μ denotes the per capita death rate (in case of μN birth rate) and therefore the life expectancy is μ^{-1}. But as it can be seen from the first two Equations 3.2.20 and 3.2.21, that this is a closed system (R is not on the right-hand side of those equations) and therefore we can disregard R from our analysis.

Let us now study the "infection rate" state $(N, 0)$. We can observe, in Equation 3.2.20 and Equation 3.2.21, that the right-hand side of these differential equations for I has a factor I, and a factor $\beta S - \mu - \gamma$. By the process of linearization this amounts to replacing S by N in the second factor, and leads to the following equations:

$$\frac{dI}{dt} = (\beta N - \mu - \alpha\gamma)\, I.$$

Hence in the above equation there can only be stability if $\beta N - \mu - \alpha\gamma < 0$ and

instability occurs if $\beta N - \mu - \alpha\gamma > 0$.

The basic reproduction number R_0 is obtained as indicated below

$$R_0 = \frac{\beta N}{\gamma + \mu}.$$

This implies that we have stability for $R_0 < 0$ (that is, whenever $\gamma + \mu > \beta N$) and instability for $R_0 > 0$. (that is, whenever $\gamma + \mu < \beta N$)

For $I \neq 0$, $\frac{dI}{dt} = 0$ requires $\beta S - \mu - \gamma = 0 \Rightarrow \bar{S} = \frac{\mu+\gamma}{\beta}$. We can rewrite this in the following form: $\frac{\bar{S}}{N} = \frac{\mu+\gamma}{\beta N} = \frac{1}{R_0}$. The same observation also shows that $S = \frac{\mu+\gamma}{\beta} = \frac{N}{R_0}$ is an isocline, so long orbits the variable I takes its maxima and minima on this line. In particular, the steady-state has to lie on this line. The steady state is a case which has to produce, on average, one secondary case and the expected number of secondary cases is R_0 multiplied by the reduction fraction $\bar{S}/N$. So in an endemic steady state $(S, I) = \left(\bar{S},\bar{I}\right)$ with $\bar{I} > 0$ necessarily

$$\frac{\bar{S}}{N} = \frac{1}{R_0}.$$

Note, that if we can estimate $\bar{S}/N$ (from blood samples taken at random), we can estimate $R_0 = N/\bar{S}$.

If we put $\frac{dS}{dt} = 0$ in Equation 3.2.22, we find

$$\bar{I} = \frac{\mu N - \mu\bar{S}}{\beta\bar{S}} = \frac{\mu}{\beta}\left(\frac{\mu N}{\mu S} - 1\right) = \frac{\mu}{\beta}\left(R_0 - 1\right).$$

In this subsection the same argument applies to the minima of I at which S is increasing. In other words, $S = \frac{N}{R_0}$ is an isocline.

So in endemic steady state

$$\bar{I} = \frac{\mu}{\beta}(R_0 - 1) \tag{3.2.26}$$

For Equation 3.2.26 and the fact that $R_0 = \frac{\beta N}{\gamma+\mu}$ we can have

$$\begin{aligned}
\bar{I} &= \frac{\mu}{\beta}\left(\frac{\beta N}{\gamma+\mu} - 1\right) \\
&= \frac{\mu}{\beta}\left(\frac{\beta N - \gamma + \mu}{\gamma+\mu}\right) \\
&= \frac{\mu}{\gamma+\mu}\left(\frac{\beta N - \gamma + \mu}{\beta}\right), \quad \text{but } \bar{S} = \frac{\mu+\gamma}{\beta} \\
&= \frac{\mu}{\gamma+\mu}\left(N - \bar{S}\right)
\end{aligned}$$

Dividing through by N we obtain Equation 3.2.27;

$$\frac{\bar{I}}{N} = \frac{(\gamma+\mu)^{-1}}{\mu^{-1}}\left(1 - \frac{\bar{S}}{N}\right). \tag{3.2.27}$$

Equation 3.2.27 expresses the relative steady-state incidence in terms of the life expectancy μ^{-1}, the expected length of the infectious period $(\gamma+\mu)^{-1}$ and the steady-state fraction of susceptibles $\bar{S}/N$.

From the sum, $\bar{S} + \bar{I} + \bar{R} = N$ the system admits the following steady states:

1. When $\bar{S} = N$, then $\bar{I} = 0$, $\bar{R} = 0$
2. When $\bar{S} = \frac{N}{R_0}$, then $\bar{I} = \frac{\mu N}{\gamma+\mu}(1 - R_0)$,$\bar{R} = \frac{\gamma N}{\gamma+\mu}\left(1 - \frac{1}{R_0}\right)$

and while the disease free equilibrium always exists, the endemic one stands only for $R_0 > 1$

From conditions (1) and (2) the steady-state is obtained, however this steady-state does not guarantee that the balance between constant inflow of new susceptibles and

the removals from deaths and/or infection is exact at every instant. There has to be a balance but it may be over a longer period of time interval. So fluctuations around the steady-state are not necessarily damped. In order to find out what happens in this model, we linearise around the steady-state.

The linearised system has solutions which depend (by factor $e^{\lambda t}$) for some values of λ. When this value is real, it denotes the growth/decay rate. When it is a complex number, then $Re\{\lambda\}$ denotes the growth/decay rate and $Im\{\lambda\}$ denotes the frequency of the oscillations which accompany the growth/decay. The principle of the linear stability guarantees that, provided that $R_e\{\lambda\}$ are non-zero , the information about solutions of the non-linear system (as long as these stay in a close neighbourhood of the steady-state).

The linearised system is fully characterised by Jacobi matrix J and the λ_s are the eigenvalues of this matrix. They are found by solving the characteristic equation $det(\lambda I - J) = 0$, which is a polynomial of degree n, where n is the dimension of the system. For us $n = 2$, so the characteristic equation is of the form

$$\lambda^2 - T\lambda + D = 0. \tag{3.2.28}$$

where T is the trace of the Jacobian matrix $J = (j_{i,j})_{1\leq i,j\leq 2}$ (i.e. the sum of the diagonal elements) and D, is determinant of J. Then we get:

$$\lambda = \frac{T \pm \sqrt{T^2 - 4D}}{2} \tag{3.2.29}$$

So we have that $T < 0$ and $D > 0$ is the condition for linearised stability (or so called decaying exponential) and $T^2 < 4D$ is the condition for the oscillations appearing around the steady state.

If we look at our system of equations outlined in Equations 3.2.22 and 3.2.23, by calculating the Jacobian matrix and evaluate its element for $S = \bar{S}$ and $I = \bar{I}$:

$$\frac{\partial}{\partial S}(\mu N - \beta SI - \mu S) = -\beta I - \mu = -\beta \bar{I} - \mu \tag{3.2.30}$$

$$\frac{\partial}{\partial I}(\mu N - \beta SI - \mu S) = -\beta S = -\beta \bar{S} \tag{3.2.31}$$

$$\frac{\partial}{\partial S}((\beta S - \mu - \gamma)I) = \beta I = \beta \bar{I} \tag{3.2.32}$$

$$\frac{\partial}{\partial I}((\beta S - \mu - \gamma)I) = \beta S - \mu - \gamma = 0 \tag{3.2.33}$$

so the corresponding Jacobian matrix is obtained as:

$$\begin{pmatrix} -\beta \bar{I} - \mu & -\beta \bar{S} \\ \beta \bar{I} & 0 \end{pmatrix}$$

The trace of the matrix is

$$T = -\beta \bar{I} - \mu < 0 \tag{3.2.34}$$

and that its determinant is

$$D = \beta^2 \bar{S} \bar{I} > 0 \tag{3.2.35}$$

and therefore the roots of the characteristics equation have negative real parts. According to the principle of linearised stability the endemic steady state is locally asymptotically stable.

In fact, the endemic steady state is globally asymptotically stable. Using Lyapunov function we show that the endemic steady state is globally asymptotic and stable.

From the steady states that are admitted by the Lyapunov function (Severo (1969)), we consider:

$$V(S,I) = S - \bar{S}\ln S + I - \bar{I}\ln I \tag{3.2.36}$$

The derivatives of Equation 3.2.36 are given by:

$$\begin{aligned}
\frac{dV}{dt} &= \frac{\partial V}{\partial S}\frac{dS}{dt} + \frac{\partial V}{\partial I}\frac{dI}{dt} \\
&= \left(1 - \frac{\bar{S}}{S}\right)(\mu N - \beta SI - \mu S) + \left(1 - \frac{\bar{I}}{I}\right) I(\beta S - \gamma - \mu) \\
&= (S - \bar{S})\left(\frac{\mu N}{S} - \frac{\mu N}{\bar{S}}\right) \\
&= \frac{\mu N}{S\bar{S}}(S - \bar{S})^2 = -\frac{\mu R_0}{S}\left(S - \frac{N}{R_0}\right)^2
\end{aligned}$$

where $\bar{S} = \frac{\alpha+\mu}{\beta}$ and $\mu N - \beta\overline{SI} - \mu\bar{S} = 0$. Hence, we can see that $\frac{dV}{dt} < 0$ except on the line $S = \frac{N}{R_0}$, where it equals zero. At the line we have $\frac{dI}{dt} = 0$ and $\frac{dS}{dt} = \mu N - (\mu + \beta I)\frac{N}{R_0}$. So $\frac{dS}{dt} > 0$ for $I < \bar{I}$ and $\frac{dS}{dt} < 0$ for $I > \bar{I}$. Orbits cannot stay on the "line", unless we consider the steady-state. According to the LaSalle's Invariance principle allows us to conclude, that all orbits which stay bounded do converge to the steady state. On the other hand, the boundedness of the orbits is a direct consequence of our assumptions (a constant population birth rate, while the per capita death rate is constant). Mathematically, this is reflected in the invariance of the region $\{(S,I) : S \geq 0, I \geq 0, S + I \leq N\}$ (note that $\frac{d}{dt}(S+I) = \mu N - (\mu(S+I) - \gamma I \leq \mu N + \mu(S+I))$.

Note **1**. $\omega-$limit of a point $\bar{x}$: $\omega(\overline{x}) = \{\bar{y} \in R^n : \bar{y}(t_k) \to \bar{y}$ for some sequence $t_k \to -\infty\}$

Note **2.** $\alpha-$limit of a point $\bar{x}$: $\alpha(\bar{x}) = \{\bar{y} \in R^n : \bar{y}(t_k) \to \bar{y}$ for some sequence $t_k \to -\infty\}$

Note **3.** (Lyapunov Second Theorem on stability)

If $f = (f_1, f_2, ..., f_n) : R^n \to R^n$ is a continuous differentiable function, $f(0) = 0$ and there exists continuous differentiable function $V(x) : R \to R$, such that it is true:Then the following are also adjudged also to hold true

(i) $V(x) \geq 0$ and $V(x) = 0 \Longleftrightarrow x = 0$

(ii) $\exists$ continuous function $W(x) : R \to R$ such that $W(x) \geq 0$ $\forall x \in R$ and $W(x) = 0 \Leftrightarrow x = 0$

(iii) $V(x) = \Sigma_{j=1}^{n} \frac{\partial V(x)}{\partial x_j} f_j(x) \leq -cW(x)$ $\forall x \in R$ where $c \geq 0$ is a constant

and the system is said to be asymptotically stable in the sense of Lyapunov and the function $V(x)$ is called the Lyapunov function for this system.

Note **4.** (LaSalle's Invariance Principle)

Suppose there is a neighbourhood D of 0 and a continuous differentiable (time-independent) positive definite function $V : D \to R$, whose orbital derivative V with the respect to the autonomous system $x = f(x)$ is negative semi-definite. Let I be the union of all complete orbits contained in

$$\{x \in D : V(x) = 0\}$$

$\Rightarrow$there is a neighbourhood U of 0, such that $\forall x_0 \in U$ $w(x_0) \subseteq I$.

In a real life, setting the life expectancy $\mu^{-1}$is usually much bigger than the duration of the infectious period γ^{-1}. Next we shall proceed to illustrate that the model predicts damped oscillations around the steady state. Further we will determine the relations between time and frequency.

Using Equation 3.2.34 and Equation 3.2.35 in the characteristic Equation 3.2.28 we can rewrite it as:

$$\lambda^2 + (\beta\bar{I} + \mu)\lambda + \beta^2\overline{SI} = 0$$

With relations $R_0 = \frac{\beta N}{\gamma+\mu}$ and Equation 3.2.25, and dividing the equation by μ^2 we obtain:
$\left(\frac{\lambda}{\mu}\right)^2 + (R_0 - 1 + 1)\frac{\lambda}{\mu} + (R_0 - 1)\frac{\beta}{\mu}\frac{N}{R_0} = \left(\frac{\lambda}{\mu}\right)^2 + R_0\frac{\lambda}{\mu} + \frac{\gamma+\mu}{\mu}(R_0 - 1) = 0.$

When $\frac{1}{\mu} \geq \frac{1}{\gamma}$,we have $\frac{\gamma}{\mu} \geq 1$, and consequently we can approximate the last term by $\frac{\gamma}{\mu}(R_0 - 1)$. The equation

$$y^2 + R_0 y + \frac{\gamma}{\mu}(R_0 - 1) = 0$$

has roots:

$$y = \frac{-R_0 \pm \sqrt{R_0^2 - 4\frac{\gamma}{\mu}(R_0 - 1)}}{2}$$

If we use the assumption $\frac{\gamma}{\mu} \geq 1$ again, we notice that the expression under the square root is negative ($R_0 > 1$ in endemic steady state) and that the roots are, in the first approximation,
$\frac{\gamma}{\mu} = y = -\frac{R_0}{2} \pm i\sqrt{\frac{\gamma}{\mu}(R_0 - 1)}.$
So we have:

1. relaxation time $\frac{1}{|Re\{\lambda\}|}$ equals to $\frac{2}{\mu R_0}$
2. frequency equals to $\sqrt{\gamma\mu(R_0 - 1)}$ with respect to the small parameter $\frac{\mu}{\gamma}$.

For $\frac{\gamma}{\mu} \geq 1$ the relaxation time is of the order $\frac{1}{\mu}$ but the period $\frac{2\pi}{\sqrt{\mu\gamma(R_0-1)}}$ is of the order of $\frac{1}{\sqrt{\mu}}$, so the ratio between the two does to infinity for $\mu \downarrow 0$.Therefore, we should see many oscillations before the steady state is reached.

3.2.5 SEIR model

The SEIR model contains one more compartment, as it is apparent from its name. The new compartment is called *exposed compartment* E. These are the people who are infected but the symptoms of the disease are not yet visible. They cannot communicate the disease either. These people are in a phase called the *latent period*. For some diseases, it takes certain time for an infective agent to multiply inside the host up to the critical level so that the disease actually manifests itself in the body of the host. This is called an *incubation period* or in the case of tuberculosis referred to as a latent period. We have the same assumptions as in the previous models, that is homogeneous mixing (mass action principle), constant population size and the rates of change from one compartment to the other follow the system below:

$$\frac{dS}{dt} = \mu N - \beta SI - \mu S \tag{3.2.37}$$

$$\frac{dE}{dt} = \beta SI - \mu E - \theta E \tag{3.2.38}$$

$$\frac{dI}{dt} = -\mu I - \theta E - \gamma I \tag{3.2.39}$$

$$\frac{dR}{dt} = -\mu R + \gamma I \tag{3.2.40}$$

In this model we shall assume that the probability to survive the latency period and to enter the infectious period equals to $\frac{\theta}{\theta+\mu}$. Therefore the basic reproductive number in this case will be $R_0 = \frac{\theta}{\theta+\mu}\frac{\beta N}{\gamma+\mu}$.

Next we proceed to review the steady states. Putting $\frac{dI}{dt} = 0$ we get $\bar{E} = \frac{\gamma+\mu}{\theta}\bar{I}$, while $\frac{dE}{dt} = 0$ leads to $\beta\bar{S}I = (\mu+\theta)\,\bar{E}$. Combining these relations, we obtain, after dividing out of a factor $\bar{I}$ (we are determining the endemic steady state so we are not interested in $I = 0$), that $\beta\bar{S} = \frac{1}{\theta}(\mu+\theta)(\gamma+\mu)$ or, equivalent, $\bar{S} = \frac{N}{R_0}$. From $\frac{dS}{dt} = 0$ we get that $\bar{I} = \frac{\mu N - \mu\bar{S}}{\beta S} = \frac{\mu}{\beta}(R_0 - 1)$.

Analogously to the derivation in Equations 3.2.30 to 3.2.33 we derive the Jacobian matrix and the characteristic equation. The linearised system is now described by the 3×3 matrix

$$\begin{pmatrix} -\left(\beta\bar{I}+\mu\right) & 0 & -\beta\bar{S} \\ \beta\bar{I} & -\left(\mu+\theta\right) & \beta\bar{S} \\ 0 & \theta & -\left(\mu+\gamma\right) \end{pmatrix} \tag{3.2.41}$$

and the eigenvalues are the roots of the characteristic equation

$\lambda^3 + (\mu R_0 + 2\mu + \gamma + \theta)\lambda^2 + \mu R_0(\gamma + \mu + \theta)\lambda + \mu(R_0 - 1)(\gamma + \mu)(\theta + \mu) = 0.$

When γ and θ are relatively close to μ and μR_0 then we can approximate the roots of the characteristic equation by the equation

$$\lambda^3 + (\gamma+\theta)\lambda^2 + \mu R_0(\gamma+\theta)\lambda\mu(R_0-1)\gamma\theta = 0 \tag{3.2.42}$$

This can be further re-written in the form as shown below

$$\lambda^3 + (\gamma+\theta)\left(\lambda^2 + \mu R_0\lambda + \mu(R_0-1)\frac{\gamma^\theta}{\gamma+\theta}\right). \tag{3.2.43}$$

It can thus be concluded that this cubic equation has one root $\lambda \approx -(\gamma+\theta)$ (corresponding to perturbations that decay rapidly) and two other roots given approximately by the roots of the quadratic equation in the braces. Hence, the period of the oscillations is given by $\pi\sqrt{\bar{a}\frac{\gamma+\theta}{\gamma\theta}}$, in other words $\frac{1}{\gamma}n$ has to be replaced by $\frac{\gamma+\theta}{\gamma\theta} = \frac{1}{\gamma}+\frac{1}{\theta}$, which is still in the expected duration of the "infection" , in the sense of the period between being infective and becoming immune.

3.2.6 Epidemic Size at any Given Time

When studying the spreading of the epidemics and its mathematical models, one asks questions concerning how long will it take for the epidemic to diminish. It seems that there is an observation which suggests, that we can find N for which infection tends to maintain in the population, whereas for smaller N it would die out and reintroducing of the infectious agent would be necessary in order to the spread of the infection. The idea of the critical community size appeared. This section is included in this thesis to give more insight on the spreading of an infectious agent in time and it shows relation between the endemic-epidemic occurrence.

Within the stochastic models the agent will go extinct with certainty. So we cannot define the critical community size just on the basis of the extinction criterion alone, the expected time until extinction has to be taken into consideration as well. This will be an increasing function of the population size N.

If for illustration purposes we chose an arbitrary constants $T > 0$ and $p \in (0, 1)$ and declare that the population size is above criticality if the probability of extinction before time T is less than p (p is dependent on the initial condition). The number of constants can be reduced to one by taking into consideration the limits. That is, concentrating on the limit $N \to \infty$ we can only see that expected time to extinction

tends to infinity, taking astronomical values of the order of e^{cN} for N large. Therefore, we have to consider more assumptions.

What we do, is that we consider approaching the infinity in a two-parameter plane, spanned by the N axis and the γ/μ axis (ratio of the two time scales). So we have concluded, that along the N axis the expected time till extinction goes infinity. Along the γ/μ, for $\gamma/\mu \to \infty$, we find opposite behaviour which is instantaneous extinction after the first outbreak.

We are interested in so called "phase transition", i.e. the way of approaching infinity in this plane such that the expected time till extinction neither blows up nor diminishes to zero; instead, it stays bounded away from zero. We try to determine the paths in the plane which provide the transition: $\frac{\gamma}{\mu}\frac{1}{\sqrt{N}}$ should be bounded. We cannot really take the limit because we do not exactly know how extinction time "bounded. So we have to make an arbitrary choice for a constant. Therefore we "define" $\frac{\gamma}{\mu}\frac{1}{\sqrt{N}} = C$ as the critical relationship, determine γ/μ , choose $C = 1$ and compute $C_{crt.}$

We consider the system Equations 3.2.20 and 3.2.21 while putting $\beta = \frac{\delta}{N}$ to show the dependence on the population size. As we have already seen, the steady state value for I is given by

$$\bar{I} = \frac{\mu(N-\bar{S})N}{\beta N S} = \frac{\mu N}{\mu+\gamma}\left(1-\frac{\bar{S}}{N}\right) = \frac{\mu}{\mu+\gamma}N\left(1-\frac{1}{R_0}\right), \tag{3.2.44}$$

since $\bar{S} = \frac{\mu+\gamma}{\beta N} = \frac{N}{R_0}$.

For births processes, it is well known that demographic stochasticity leads to fluctuations of the order of $\sqrt{N}$, with N the population size Nisbet and Gurney (1982), Goel and Richter-Dyn (1974) and Taylor and Karlin (1984). Assume $R_0 = O(0)$

what means that changes in N do not substantially influence R_0. Also assume that $\frac{\mu}{\mu+\gamma} = O(\frac{1}{\sqrt{N}})$.These assumptions were made such that they imply $\bar{I} = O\left(\sqrt{N}\right)$ (the average level of infected individuals in the population lies within the range of natural fluctuations). The agent will extinct, sooner or later. When $\mu \geq \gamma$ we have that $\frac{\mu}{\mu+\gamma} = \frac{\mu}{\gamma}$, this is as a result of the assumption that $\frac{\mu}{\mu+\gamma} = O(\frac{1}{\sqrt{N}})$ which implies that $\frac{\mu}{\gamma} = O(\frac{1}{\sqrt{N}})$ and therefore $\frac{\gamma}{\mu} = O\left(\sqrt{N}\right)$.Since $R_0 = \frac{\beta N}{\mu+\gamma} = O(1)$, this requires also $\frac{\beta N}{\mu} = O\left(\sqrt{N}\right)$.

So there is not a critically directly in the community size, but rather a critical relationship between population size and the ratio of the two time scales involved (that of demography and that of transmission). When both $\frac{1}{\sqrt{N}}\frac{\gamma}{\mu}$ and $\frac{1}{\sqrt{N}}\frac{\beta N}{\mu}$ are really small, we expect a single outbreak of an epidemic and when they are large we expect an endemic situation; everything in between is considered to be critical. The approximate formula for expected extinction time under critical conditions according to NASELL (1999) is

$$\bar{t}_{extinction} = \frac{(R_0 - 1)N}{2\left(\frac{\gamma}{\mu}\right)^2}\frac{1}{\mu} \tag{3.2.45}$$

Note that the right hand side is $O\left(\frac{1}{\mu}\right)$when $\frac{\gamma}{\mu}$ is $O\left(\sqrt{N}\right)$, what is completely in line with the text above.

3.3 Mathematical Model of the Branching Process

The model proposed by Watson and Galton (1875) was of the following form:

1. A population starts with one individual at time $n = 0 : Z_0 = 1$
2. After one unit of time (at time $n = 1$) the sole individual produces Z_1 identical clones of itself and dies. Z_1 is an N_0 valued random variable.

3. (a) If Z_1 happens to be equal to 0 the population is dead and nothing happens at any future time $n \geq 2$

   (b) If $Z_1 > 2$, a unit of time later, each of Z_1 individuals gives birth to a random number of children and dies. The first one has $Z_{1,1}$ children, the second one $Z_{1,2}$ children, etc. The last, Z_1^{th} one, gives birth to Z_{1,Z_1} children. It is assumed that the distribution of the number of children is the same for each individual in every generation and independent of either the number of individuals in the generation and of the number of children the others have. This distribution, shared by all $Z_{n,i}$ and Z_1, is called the o*ff spring distribution*. The total number of individuals in the second generation is now $Z_2 = \sum_{k=1}^{Z_1} Z_{1,k}$

4. The third, fourth, etc. generations are produced in the same way. If it ever happens that $Z_0 = 0$ for some n, then $Z_m = 0$ for all $m \geq n$ -the population is extinct otherwise $Z_{n+1} = \sum_{k=1}^{Zn} Z_{n,k}$

Stochastic proceses with the properties described in (1), (2),(3) and (4) above are referred to as branching process.

3.3.1 Branching Process Approximation

This section describes an approximation to the SIR epidemic model valid for large population sizes and/or during the early stages of the epidemic. First the approximating branching process is defined and then a coupling argument is set out to make clear the relationship between the two processes. The term 'branching process' is used in several different ways in the literature, however in this thesis it is used exclusively to describe the continuous time branching process defined in section 3.3.

Let $I(t)$ denote the number of individuals alive in the branching process at time t, so that overall, births occur at a rate $\beta I(t)$. This branching process is an approximation to the epidemic process in the early stages or when n is large as in these circumstances

$S(t) \approx S(0) = n$ and so infections will occur in the epidemic process at a rate approximately equal to $\beta I(t)$ while the infectious periods in the epidemic process have the same distribution as the lifetime distribution in the branching process. The relationship between the processes can be made much more precise by constructing them on the same probability space, so that there is a coupling between them. This will allow obtaining precise instant at which the two processes first diverge, and to observe that after this instant the number of individuals alive in the branching process is an upper bound for the number of infectives in the epidemic process.

3.3.2 Coupling the Branching Process and the Epidemic

A coupling between the epidemic process and the branching process is constructed from the algorithm by Anderson and Tom (2000). First, we will construct the branching process and then enlarge its probability space $(\delta; F; P)$ to include an infinite sequence $\{U_j\}, j \geqslant 1$ of independent uniform random variables on $(0, 1)$. Next, we construct the epidemic process from the branching process as follows. Labelling the n susceptibles in the epidemic process from $1, ..., n$ and associate each initial infective with an initial ancestor in the branching process. We let a contact occur in the epidemic process when a birth occurs in the branching process. The $j - th$ individual to be contacted is defined to be the initially susceptible individual with label $\{{}_nU_j\}$, and if this individual is still susceptible then they become an infective. However, if they have already been contacted then the contact has been unsuccessful

and the epidemic process does not change. The individual born in the branching process at this instant and all of their offspring are subsequently ignored in the construction of the epidemic process and are called 'ghosts'. When a death occurs or an individual has been put on treatment and now no longer infective i.e. got cured in the branching process, the corresponding infective in the epidemic is removed. By constructing the epidemic process from the branching process in this particular manner it will be clear that the number of infectives in the epidemic and the number of individuals in the branching process must coincide until the time of the first 'ghost', after which there must be more individuals (by the number of ghosts) in the branching process. Also note that in any time interval $(0, t)$ for $t \in \mathbb{R}$, there can be only finitely many births and so the probability that the two processes coincide throughout $(0, t)$ tends to 1 as n tends to infinity.

3.4 Mathematical Illustration of Small Area Estimation

Literature on Small Area Estimation has received significant attention in the last decade. Rao (2003) published a comprehensive book on SAE that covers all the main developments in this topic up to that time. The book was written about ten years after the review paper of Ghosh and Rao (1994), published in Statistical Science, which stimulated much of the early research in SAE. Since 2003, a few other review papers have been published; for example, Rao (2005), Jiang and Lahiri (2006), Datta (2009) and Lehtonen and Veijanen (2009). In this thesis we illustrate SAE based on the framework that has been developed by Pfeffermann (2013). The study illustrated that if we consider a population U of size N distinct elements identified through the labels $j = 1, ..., N$. If then a sample s is selected from U with probability $p(s)$, and the probability of including the $j - th$ element in the sample is π_j. The design

weight for each selected unit $j \in s$ is defined as $w_j = {}^{1}/_{\pi_j}$. If then the population U is divided into M exclusive and exhaustive areas $U_1 \cup ... \cup U_M$ with N_i units in area i, $\sum_{i=1}^{M} N_i = N$. Suppose that samples are available for $m \leq M$ of the areas, and let $s = s_1 \cup ... \cup s_m$ define the overall sample, where s_i of size n_i is the sample observed for sampled area i, $\sum_{i=1}^{m} n_i = n$. Note that n_i is random unless a planned sample of fixed size is taken in that area. Let y define the characteristic of interest, and denoted by y_{ij} the response value for unit j belonging to area $i, i = 1, ..., M$, $j = 1, ..., N_i$ with sample means $\bar{y}_i = \sum_{j=1}^{n_i} {y_{ij}}/{n_i}$, it is assumed without loss of generality that the sample consists of the first n_i units. Let $x_{ij} = (x_{1ij,...,}x_{pij})'$ denote the covariate values associated with unit (i, j) and by $\bar{x}_i = \sum_{j=1}^{n_i} \frac{x_{ij}}{n_i}$ the column vector of sample means.The corresponding vector of true area means is $\bar{X}_i = \sum_{j=1}^{N_i} \frac{x_{ij}}{N_i}$. The area target quantity is denoted by θ_i; where, $\theta_i = \bar{Y}_i = \sum_{j=1}^{N_i} \frac{y_{ij}}{N_i}$, the response area mean. Estimating a proportion is a special case where y_{ij} is binary.

3.5 Generalized Linear Mixed Models

The main assumption of GLMMs is that the distribution of the response variable Y_{i} belongs to an exponential family of the form Y_i/θ_i , ϕ_1 $p(\cdot)$

where $p(\cdot)$ is a member of the exponential family defined as,

$$p(y_i/\theta_{\mathrm{i}}, \phi_1) = \exp\left(\frac{y_i\theta_i - b(\theta_i)}{a(\phi_1)} + c(y_i, \phi_1)\right) \tag{3.5.1}$$

for $i = 1, ..., n$ observations and θ_{i} is the scalar canonical parameter. The mean $\mu_{\mathrm{i}} = \mathrm{E}\left(y_{\mathrm{i}}/\beta, f^{\mathrm{i}}(\cdot), \phi_1\right)$ can be linked to the structured additive predictor which

accounts for various covariates η_{i} by a monotonic link function $g(\cdot)$ such that

$$g(\mu_{\mathrm{i}}) = \eta_{\mathrm{i}} = \beta_0 + \sum_{i=1}^{n} f_i(\mu_{\mathrm{i}}) + \sum_{k=1}^{m} \beta_{\mathrm{k}} x_{ki} + \varepsilon_{\mathrm{i}} \tag{3.5.2}$$

where $f_i(\cdot)$ are unknown functions of the covariates used to model temporal and spatial dependencies and also used to relax the linear relationships of the covariates. The $\beta'_{\mathrm{k}}s$ represents the linear effects of the covariates x's and $\varepsilon'_{\mathrm{i}}$s are unstructured terms. Latent Gaussian model a flexible and large class of statistical models obtained by assigning a Gaussian prior to β_0, $f_i(\cdot)$, β_{k} and ε_{i}. This can be represented as $\Theta = (\beta'_{\mathrm{k}}\mathrm{s}, f'_i\mathrm{s},)$ where Θ is unobserved multivariate Gaussian random variable, whose density $\pi(\Theta/\phi)$ is controlled by a vector of hyperparameters Φ (Rue and Martino (2007a)). The latent Gaussian field Θ is assumed to have a Gaussian distribution with zero mean and variance covariance matrix $Q(\phi_2)$; with vector of hyper-parameters defined as $\Phi = (\phi_1, \phi_2)$ which are not necessarily Gaussian (Martins et al. (2013), Rue et al. (2009) and Fong et al. (2009)). Latent Gaussian model is composed of three elements namely; the likelihood of the data $\pi(y/\Theta)$, the Gaussian density of the random vector Θ, $\pi(\Theta/\Phi)$ and the prior distribution of the parameter vector $\pi(\Phi)$. The posterior is therefore defined as

$$\pi(\Theta, \Phi/y) \propto \pi(\Phi)\,\pi(\Theta/\Phi) \prod_{\mathrm{i}=1} \pi(y_i/x_i, \Phi) \tag{3.5.3}$$

The main inferential interest involves computing the posterior marginals for x_i and posterior marginals for Φ or some Φ_j . The approaches for Bayesian inference on latent Gaussian models are MCMC sampling and INLA. The high dimensionality of the latent field Θ and the strong correlation within Θ and between Θ and Φ especially when the number of observations are many leads to problems in convergence and

computation time. INLA developed by Rue et al. (2009) bypasses MCMC entirely by basing inferences on closed form approximations making it computationally efficient compared to MCMC. In the next section we describe INLA methodology.

3.6 INLA Methodology

This is an approximate inference based method for approximating the posterior marginals of the latent Gaussian field $\pi\left(x_i/y\right), i = 1....n$ in three steps. The posterior marginals of the latent effects Θ and hyper-parameters Φ are written as

$$\pi\left(x_{\mathrm{i}}/y\right) = \int \pi\left(x_i/\Phi, y\right)\pi\left(\Phi/y\right) d\Phi \tag{3.6.1}$$

$$\pi\left(\Phi_{\mathrm{i}}/y\right) = \int \pi\left(\Phi/y\right) d\Phi_{-j} \tag{3.6.2}$$

The posterior marginals $\tilde{\pi}\left(x_{\mathrm{i}}/y\right)$ and $\tilde{\pi}\left(\Phi_j/y\right)$ can be approximated using the Laplace approximation. The first approximation$\pi\left(\Phi/y\right)$ to using Gaussian distributions is constructed as follows

$$\tilde{\pi}\left(\Phi/y\right) \propto \left.\frac{\pi\left(\Theta, \Phi, y\right)}{\tilde{\pi}_G\left(\Theta/\Phi, y\right)}\right|_{\Theta=\Theta^*(\Phi)} \tag{3.6.3}$$

where $\tilde{\pi}_G\left(\Theta/\Phi, y\right)$is a Gaussian approximation to the full conditional of Θ and $\Theta^*\left(\Phi\right)$ is the mode of the full conditional for Θ, for a given value of Φ (Rue et al. (2009)). It involves locating the mode of $\tilde{\pi}\left(\Phi/y\right)$ which is used to integrate out the uncertainty with respect to Φ when approximating the posterior marginal of x_{i}. The posterior marginals of the latent field are supposed to start from $\tilde{\pi}_G\left(x_i/\Phi, y\right)$ and approximate

the density of $x_i/\Phi, y$ with the Gaussian marginal derived from $\tilde{\pi}_G(\Theta/\Phi, y)$, i.e.

$$\pi(x_i/\Phi, y) = N\left(x_i; \mu_i(\Phi), \sigma_{ii}^2(\Phi)\right) \tag{3.6.4}$$

The marginals of the interest can be computed using numerical integration over a multidimensional grid of values of Φ

$$\pi(x_i/\Phi, y) = N\left(x_i; \mu_i(\Phi), \sigma_{ii}^2(\Phi)\right) \tag{3.6.5}$$

where the sum is over the values of Φ with area weights Δ_k (Rue and Martino (2007b)). The first step in INLA computation involves approximating the posterior marginal of by using Laplace approximation in Equation 3.6.3 The second step involves computing the Laplace approximation of$\tilde{\pi}(x_{\mathrm{i}}/\Phi, y)$ for selected values of Φ which improves the Gaussian approximation in Equation 3.6.4.

$$\tilde{\pi}_{LA}(x_i/\Phi, y) \propto \left.\frac{\pi(\Theta, \Phi, y)}{\tilde{\pi}_{GG}(\Theta_{-i}/x_i, \Phi, y)}\right|_{\Theta_{-i}=\Theta^*_{-i}(x_i,\Phi)} \tag{3.6.6}$$

where $\tilde{\pi}_{GG}(\Theta_{-i}/x_i, \Phi, y)$ is a Gaussian approximation to $\Theta_{-\mathrm{i}}/x_i, \Phi, y$around its mode $\Theta_{-i}(x_i, \Phi)$. An improved version of $\tilde{\pi}_{LA}(x_i/\Phi, y)$ known as Simplified Laplace approximation was developed by Rue et al. (2009). It involves a series of expansion of $\tilde{\pi}_{LA}(x_i/\Phi, y)$around $x_i = \mu_i(\Phi)$ which corrects for skewness and location and it is also less computationally expensive (Rue et al. (2009)). The third step involves combining steps 1 and 2 using numerical integration in Equation 3.6.5 (For more details, Rue et al. (2009)).

3.7 Spatial Effects

Suppose that the index $s \in (1, ..., S)$ represents the geographically connected regions. The spatially correlated effects in INLA are introduced by assuming that neighbouring regions are more alike than two arbitrary regions. Two regions and are neighbours if they share a common boundary. The spatial smoothness prior for the function evaluation $f(s)$ is given by

$$\frac{f(s)}{f(s')} \text{ for } s \neq s' \text{ and } \tau^2 \sim N\left(\frac{1}{N_s}\sum_{s\sim s'} f(s'), \frac{\tau^2}{N_2}\right) \tag{3.7.1}$$

where N_s the number of neighbours of region is,indicates that two regions and are neighbours and is the variance parameter (Martino and Rue (2009); Brezger et al. (2005)).

3.8 TB Distribution in Kenya

In this study a generalized linear mixed model assuming a Poisson distribution with spatial and temporal random effects was used to characterize the relationships between TB cases and covariates. Poisson regression was used primarily because it has a strenth in modelling count data it assumes that the rate parameter μ is constant over all intervals. In this model the response variable is generated by a Poisson process: The Poisson regression model is $E(y_i) = \mu_i = \exp(\mathrm{X}_i\beta + offset_i)$ where y_i is the observed TB cases, X_i are the covariates for the observation and offset term represented the population (**StataCorp, 2013**). The generalized linear mixed

model used to describe the TB cases y_i is of the form:

$$g(\mu_i) = \beta_0 + \sum_j \beta_j X_{ij} + f_{trend}(time) + f_{\text{str}}(S_i) + f_{\text{unstr}}(S_i) \tag{3.8.1}$$

where $g(.)$ is a monotonic link function, β_j represents the parameter vector of the covariates X_{ij}, f_{trend} is trend component, f_{str} and f_{unstr} are structured and unstructured spatial effects of the county. The f_{str} and f_{unstr} were assigned a Markov random field prior and Gaussian i.i.d respectively. The spatial effects were estimated at county level in which a household was located and Kenya counties boundaries were used to compute the neighbourhood information. The prior for f_{trend} was first order random walk. The covariates were assigned default INLA Gaussian priors while the default inverse gamma hyperpriors were assigned for the random effects.

Chapter 4

RESULTS AND DISCUSSION

4.1 Simulation Study

This section presents results based on simulated data on epidemic models discussed in chapter three for the deterministic and stochastic epidemic model with a view of developing algorithms which were later subject to the real tuberculosis epidemiological data in Kenya. The Poisson distribution was assumed for the birth and death rates spanning 63 years corresponding to the life expectancy in Kenya at present.

4.1.1 Deterministic Model Simulation Results

We carried out simulation using an initial population of 1000 individuals while assuming that transmissible acts can either take 1, 3, 5, 10 or 15 individuals being infected by one single individual at the start of the epidemic. The choice of transmissible acts was based on the existing literature WHO (2007) that one case of TB, if left untreated, can infect between ten and fifteen people per year. We further assumed that the recovery rates for those who get infected can be 0.80, 0.85 or 0.90.

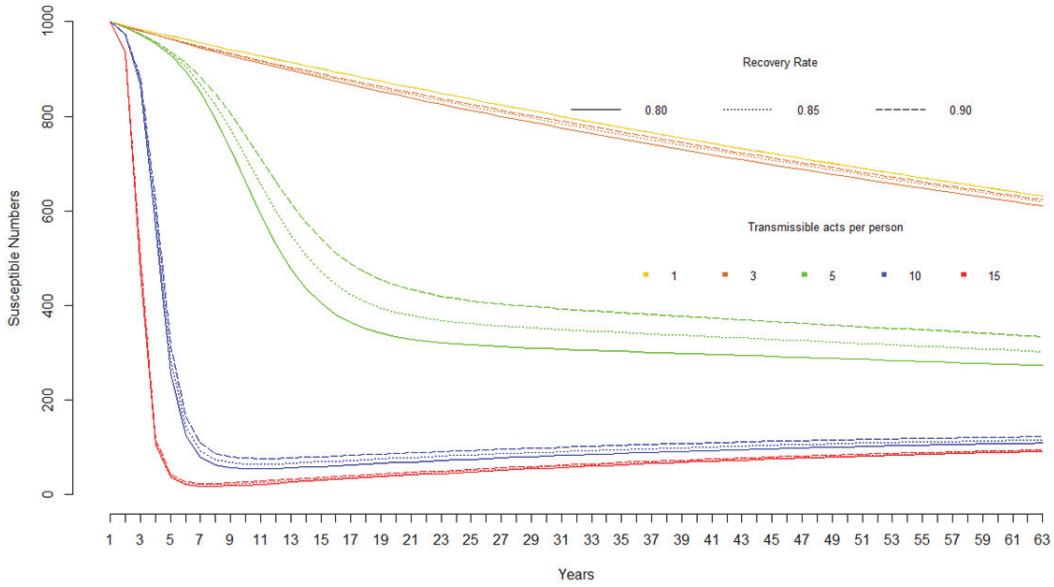


Figure 4.1.1: DCM model representation for the simulated susceptible number of cases

Figure 4.1.1, shows that when the transmissible act is large there is a significant drop in the number of cases up to the 5th year and begins to rise and remains close to 200 susceptible cases over all the duration. When the transmissible acts are few i.e. 1 and 3 cases the number of susceptible cases remains high throughout the period, while when transmissible acts are 5 the susceptible cases show and exponential decay over the period but remains high.

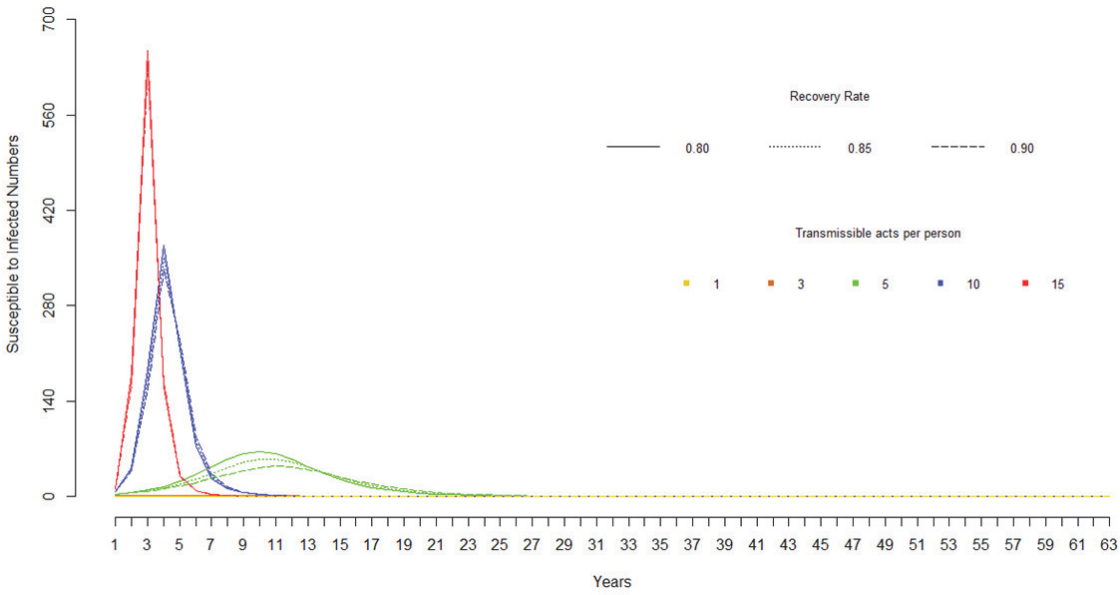


Figure 4.1.2: DCM model representation for the simulated susceptible to infected number of cases

Figure 4.1.2 shows that the number of cases that transition from the susceptible to group to the infection group is highest when the transmissible acts is large i.e. 10 and 15 and reduces dramatically by the 11th year while peaking by year 3 and 5 respectively. The transmissible act of 5 tends to peak by the 10th year and reduce to zero by the 27th year. While when transmissible acts are 1 or 3 the numbers which move from the susceptible pool to the infectious pool.

Figure 4.1.3 shows the number of infected cases at any particular year. It shows that when the transmissible acts is high the numbers rapidly peaks and also rapidly declines thereafter, when the transmissible acts are 15 and 10 respectively but when transmissible acts is 5 the number of infected numbers tends to assume a normal distribution peaking by year 11 and declines to zero by year 31. When transmissible acts are either 1 or 3 the numbers are barely noticeable and moves to zero by year 3.

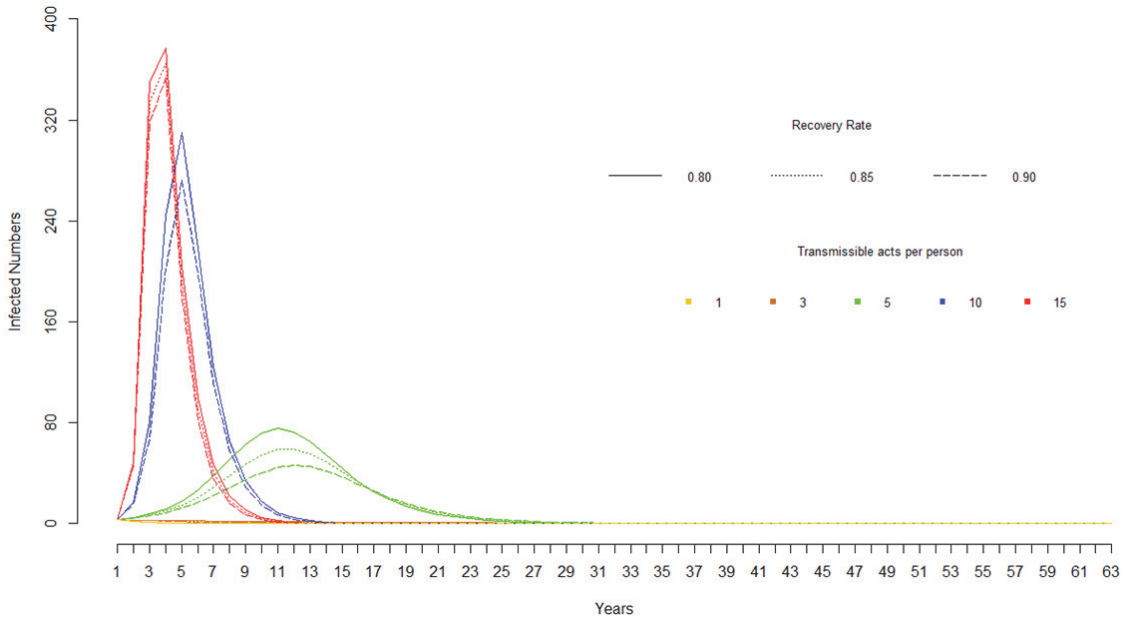


Figure 4.1.3: DCM model representation for the simulated infected number of cases

Figure 4.1.4 shows the number of infected cases which move to the recovery compartment at any particular year. It shows that when the transmissible acts is high the numbers rapidly peaks and also rapidly declines thereafter, when the transmissible acts are 15 and 10 respectively but when transmissible acts is 5 the number of infected numbers tends to assume a normal distribution peaking by year 11 and declines to zero by year 31. When transmissible acts are either 1 or 3 the numbers are barely noticeable and moves to zero by year 3.

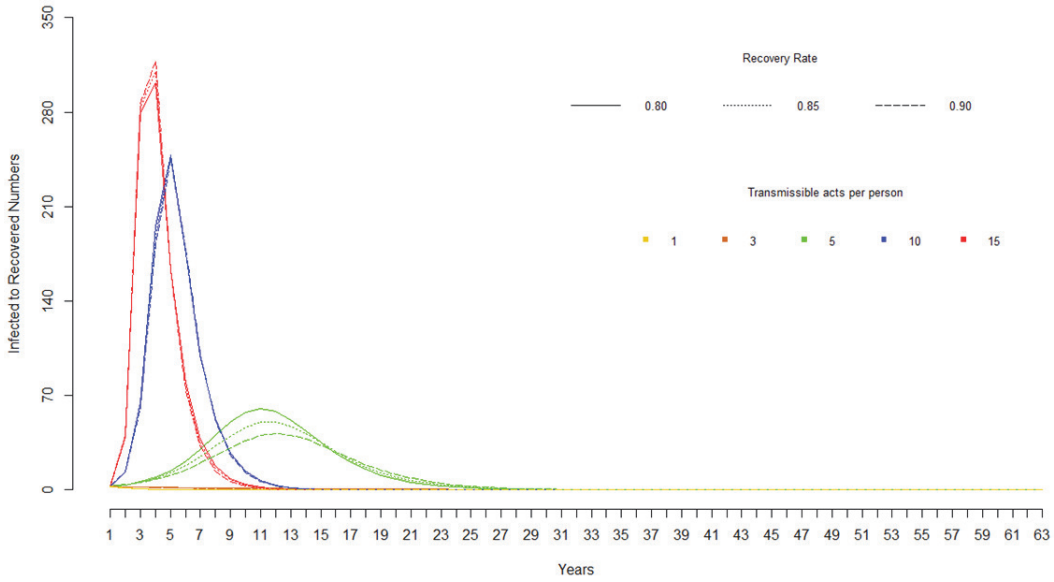


Figure 4.1.4: DCM model representation for the simulated infected to recovered number of cases

4.1.2 Stochastic Simulation Results

The Stochastic Individual Contact Models (SICM) belongs to a novel class of microsimulation models were developed to mirror the deterministic models but add random variation in all components of the transmission dynamics system, from infection to recovery to vital dynamics (births and deaths). We implemented ICM model using EpiModel in R. The results we present below relate to simulated data with 1000 individuals while the parameters of interest were transmissible acts per person (1,3,5,10,15) this refers to the number of individuals one individuals could potentially infect if left untreated. A total of 5,000 simulations were carried out to obtain the parameter estimates.

Figure 4.1.5 shows that regardless of the transmissible acts there is a uniform decline in the number of susceptible number of cases from the hypothetical 1,000 number of susceptible cases. Only when the transmissible acts is 15 is when we realize a more pronounced decline.

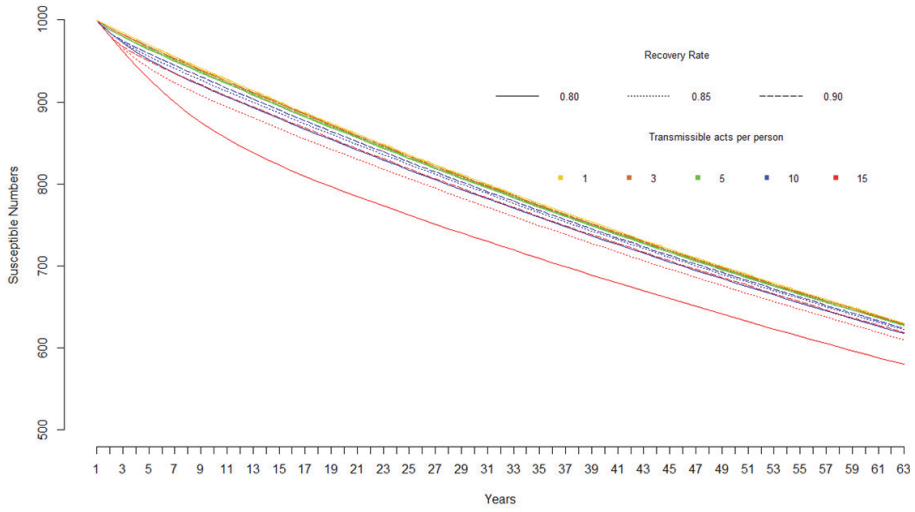


Figure 4.1.5: ICM model representation for the simulated susceptible number of cases

Figure 4.1.6 shows that the transition numbers from susceptible to infected shows a rise for all transmissible and peaks after 2 years. The rate of decline is not highly pronounced when the transmissible acts is 15 and declines to zero after 27 years. While the other transmissible acts comes to a close after 13 years.

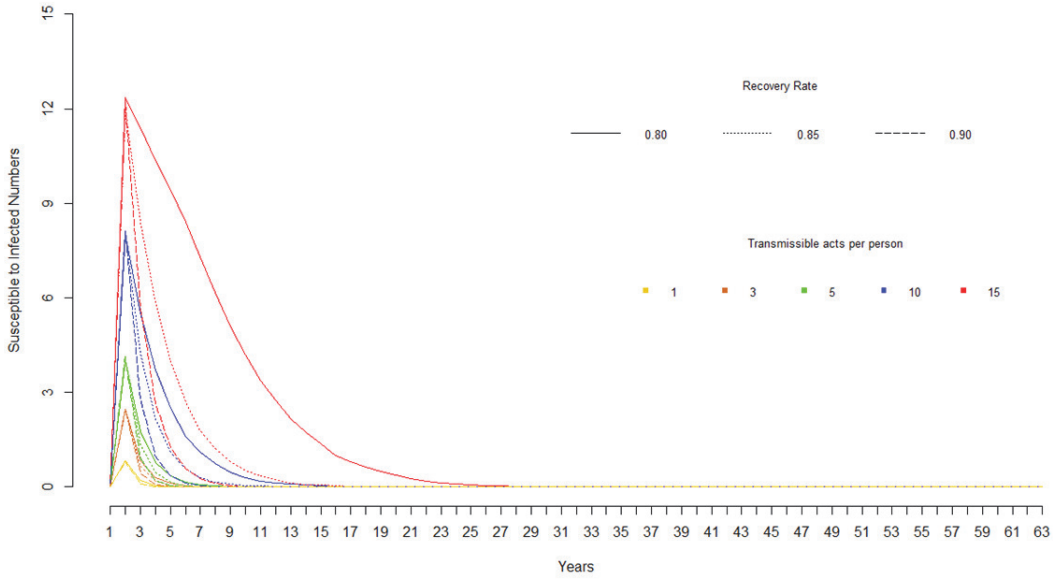


Figure 4.1.6: ICM model representation for the simulated susceptible to infected number of cases

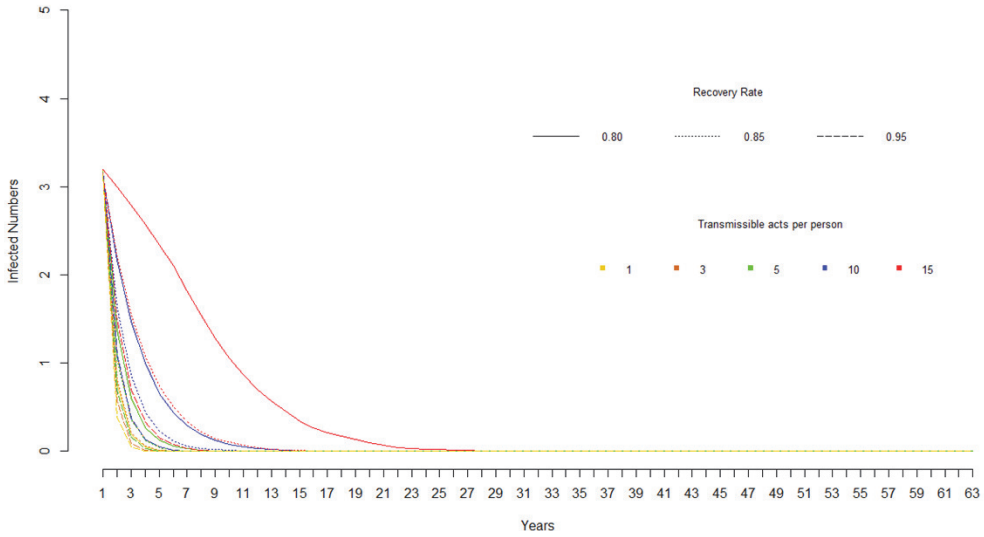


Figure 4.1.7: ICM model representation for the simulated infected number of cases

The infected number of cases for all the transmissible acts from Figure 4.1.7 shows a uniform rate of decline for all the transmissible acts with exception of 15 transmissible acts. This shows that the number of infected numbers shows an exponential decay with the number coming to zero by the end of 28 years.

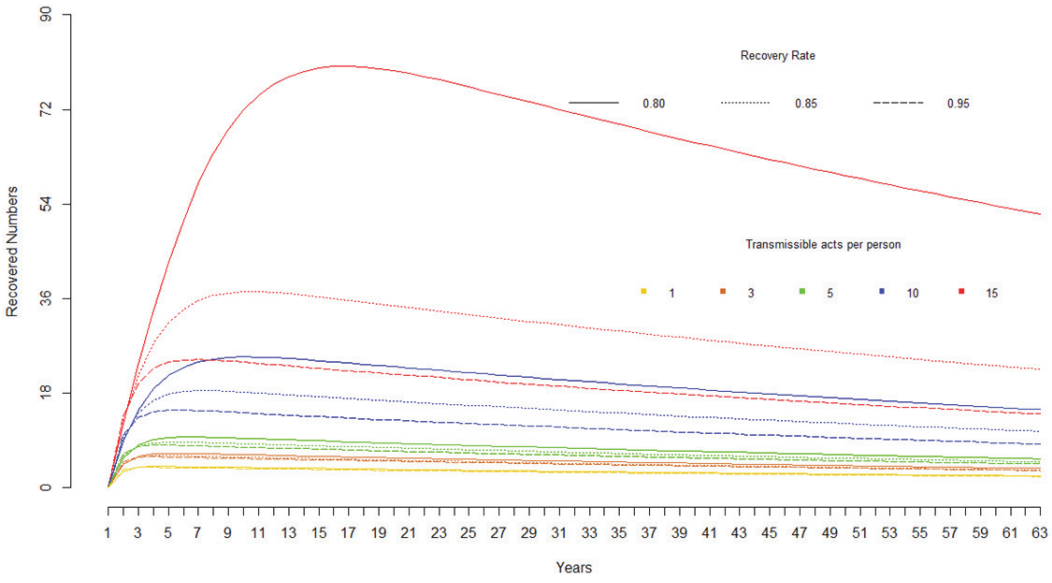


Figure 4.1.8: ICM model representation for the simulated recovered number of cases

The Figure 4.1.8 shows that the number of cases that recover increases sharply up to year 2 after which it begins to decline the number of cases which recover decline in a stable manner. The number of cases which recover remains lower for the 1, 2 and 3 transmissible acts.

4.1.3 Branching Process Simulation

The plot shown in figure 4.1.9 gives a sample output where X~binom(3,0.26) where 3 is the initial number infected and 0.26 is the probability of the infection. There were 10000 simulations over 63 generations. The number of infections seems to be increasing up to generation 10 and then starts to decline and die out by generation

48.

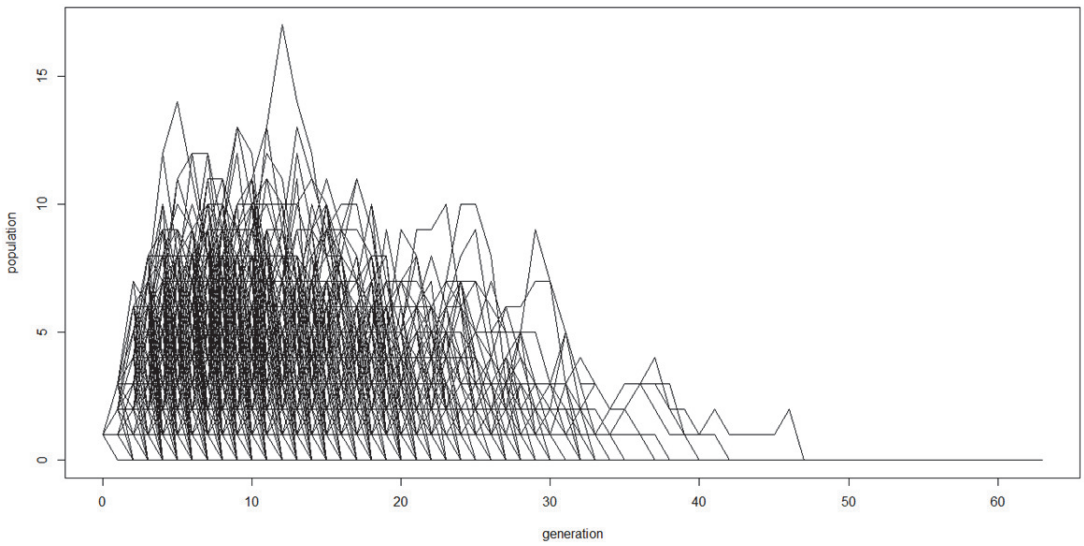


Figure 4.1.9: Branching process simulation

4.2 Empirical Results

4.2.1 Deterministic Model Results

In this section we present results when the deterministic model was subjected to the tuberculosis data reported to the National tuberculosis program for the period 2012-2014.

The results presented in this sub-section relate to the application of the DCM model on the case notification data in Kenya. The assumption we took was that 1/3 of the population are infected with the tuberculosis albeit it could be in latent form and some may never develop TB in their lifetime. Thus in Kenya with an approximate population of 44 million people it is estimated that close to 15 million individuals would be susceptible of developing TB in a lifetime.

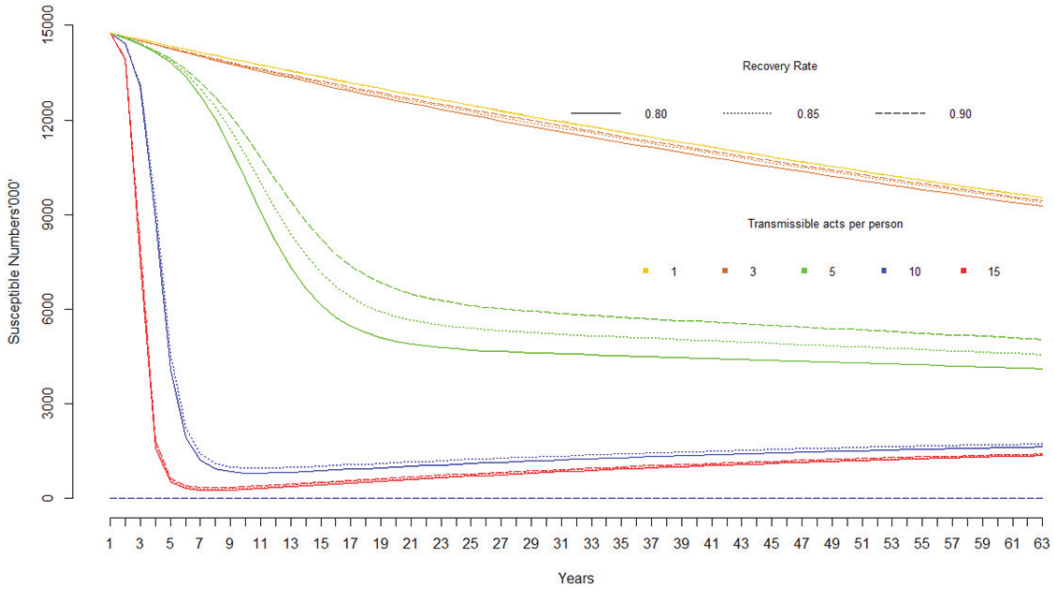


Figure 4.2.1: DCM model representation for the susceptible number of cases in Kenya

Figure 4.2.1 shows that when the transmissible act is large there is a significant drop in the number of cases up to the 5th year and begins to rise and remains close to 1.5 million susceptible cases over all the duration. When the transmissible acts are few, that is, 1 and 3 cases the number of susceptible cases declines uniformly and remains high throughout the period at approximately 7-8 million susceptible cases, while when transmissible acts are 5 the susceptible cases show and exponential decay over the period but remains high. The observed data shows consistent behavior as the simulated data.

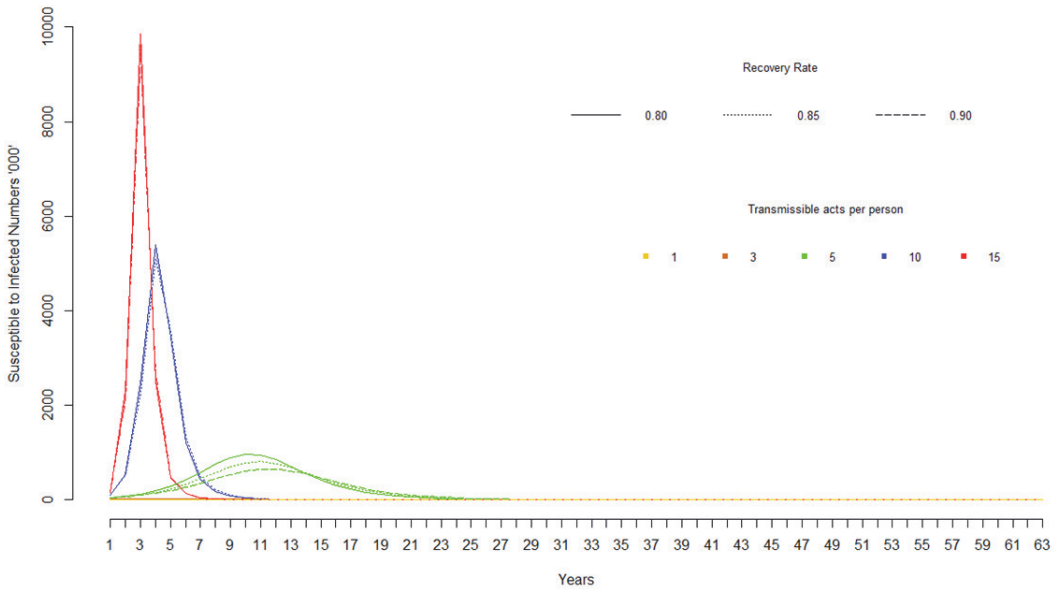


Figure 4.2.2: DCM model representation with observed data for the susceptible to infected number of cases in Kenya

Figure 4.2.2 shows that the number of cases that transition from the susceptible to group to the infected pool, the numbers which transition is highest when the transmissible acts is large i.e. 10 and 15 and reduces dramatically by the 11th year while peaking by year 3 and 5 respectively. The transmissible act of 5 tends to peak by the 11th year and reduce to zero by the 28th year. While when transmissible acts are 1 or 3 the numbers which move from the susceptible pool to the infectious pool remain fairly stable at less than 1 million cases all through the period.

Figure 4.2.3 shows the number of infected cases at any particular year. It shows that when the transmissible acts is high the numbers rapidly peaks and also rapidly

declines thereafter this could be attributed to the high recovery rates, when the transmissible acts are 15 and 10 respectively but when transmissible acts is 5 the number of infected numbers tends to assume a normal distribution peaking by year 12 and declines to zero by year 31. When transmissible acts are either 1 or 3 the numbers are barely noticeable and moves to zero by year 3.

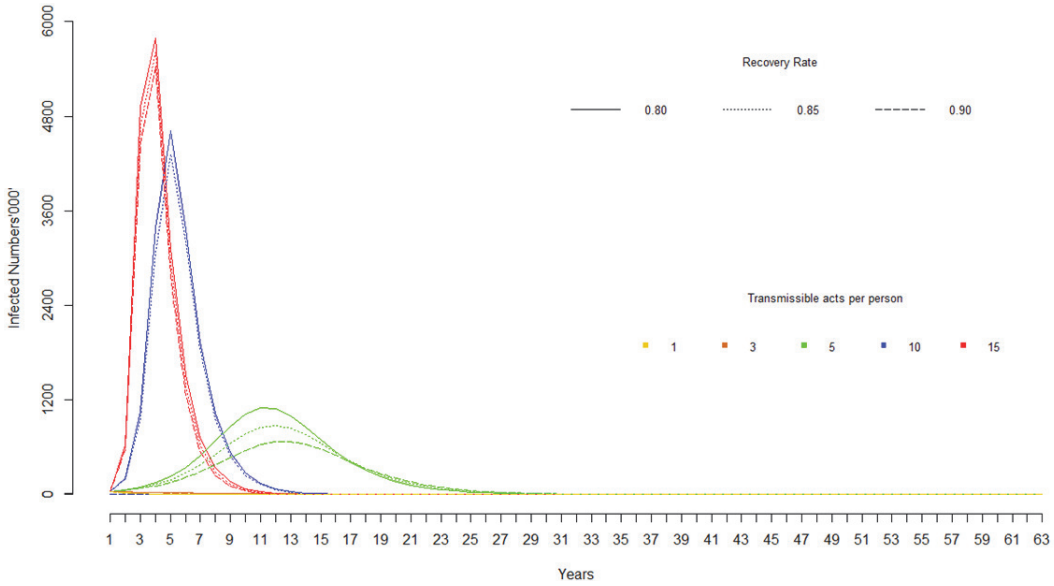


Figure 4.2.3: DCM model representation for the infected number of cases in Kenya

Figure 4.2.4 shows the number of infected cases which move to the recovery compartment at any particular year. It shows that when the transmissible acts is high the numbers rapidly peaks and also rapidly declines thereafter, when the transmissible acts are 15 and 10 respectively but when transmissible acts is 5 the number of infected numbers tends to assume a normal distribution peaking by year 11 and declines to zero by year 32. When transmissible acts are either 1 or 3 the numbers are barely noticeable and moves to zero by year 3.

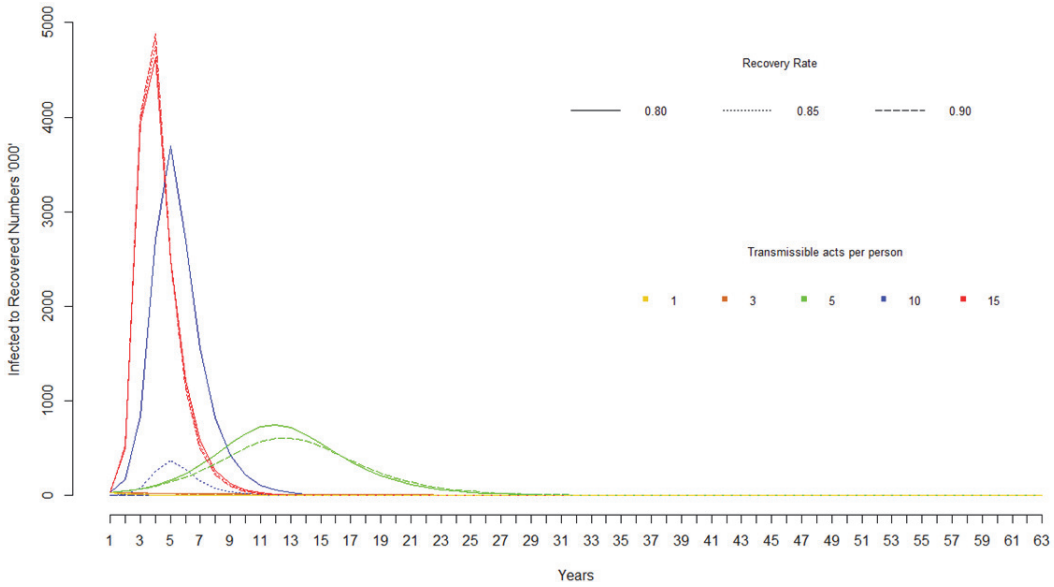


Figure 4.2.4: DCM model representation with observed data for the infected to recovered number of cases in Kenya

Figure 4.2.5 shows the number of infected cases which eventually recover the numbers rapidly increase when the transmissible acts is large i.e. 10 and 15, this number peaks by year 10 and 7 respectively. But when transmissible acts are 5 the number of infected numbers tends rise exponentially and peaks by the 17th year and begins to decline throughout the remaining period. When transmissible acts are either 1 or 3 the numbers are barely noticeable and remains constant throughout the period after the 15th year.

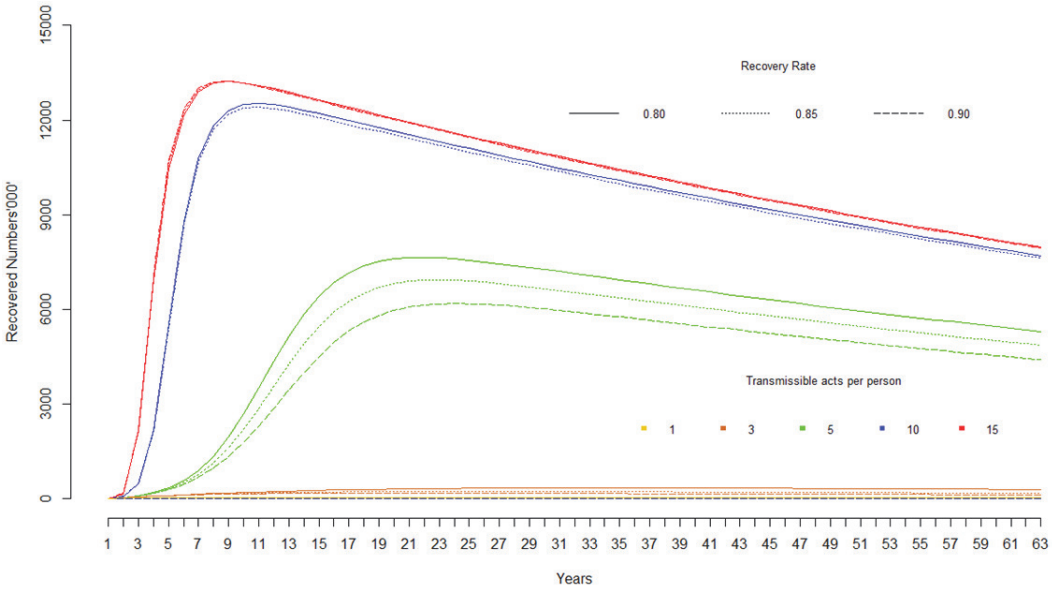


Figure 4.2.5: DCM model representation with observed data for the recovered number of cases in Kenya

4.2.2 Stochastic Model Results

The results we present here relates to the models that is obtained when it is subjected to the actual notifications reported in Kenya. The assumption we took was that 1/3rd of the population are infected with the tuberculosis albeit it could be in latent form and some may never develop TB in their lifetime. Thus in Kenya with an approximate population of 44 million people it is estimated that close to 15 million individuals would be susceptible of developing TB in a lifetime.

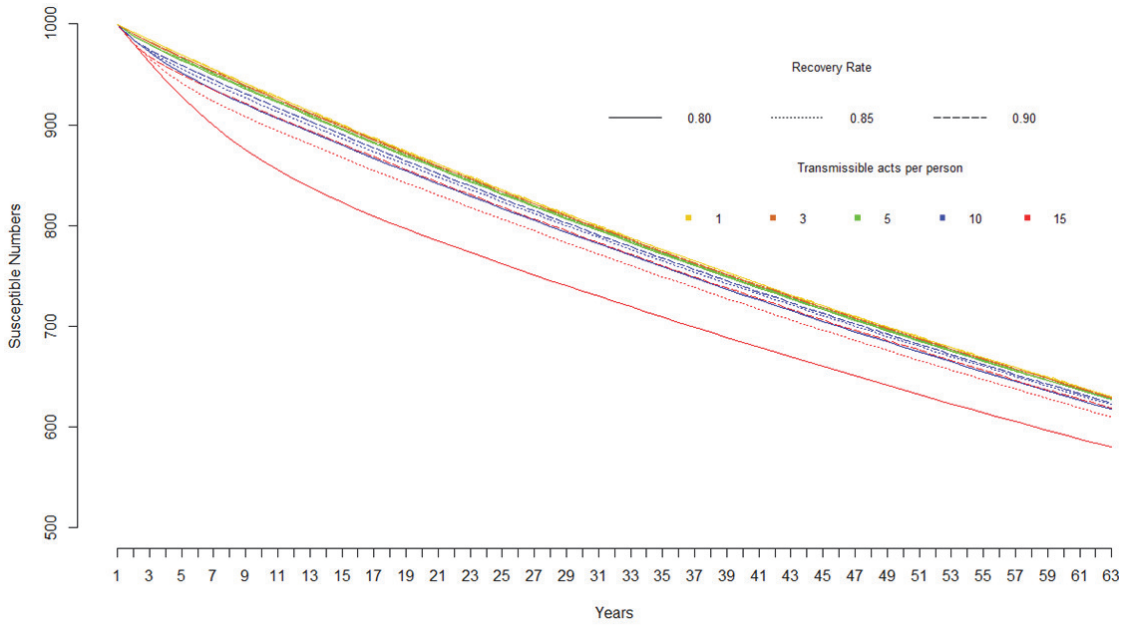


Figure 4.2.6: ICM Model representation with Observed data for the Susceptible Number of Cases

The susceptible number of cases from the estimated Kenyan population was taken to be approximately 15 million people. With an assumed recovery rate of 80, 85 and 90% with transmissible acts per person being (1, 3,5,10 and 15), when the observed data was used it showed that there is close concordance with the simulated model. Figure 4.2.6 shows a smooth decline for the susceptible numbers with exception when

the transmissible act is 15, the rate of decline is higher than any other transmissible acts.

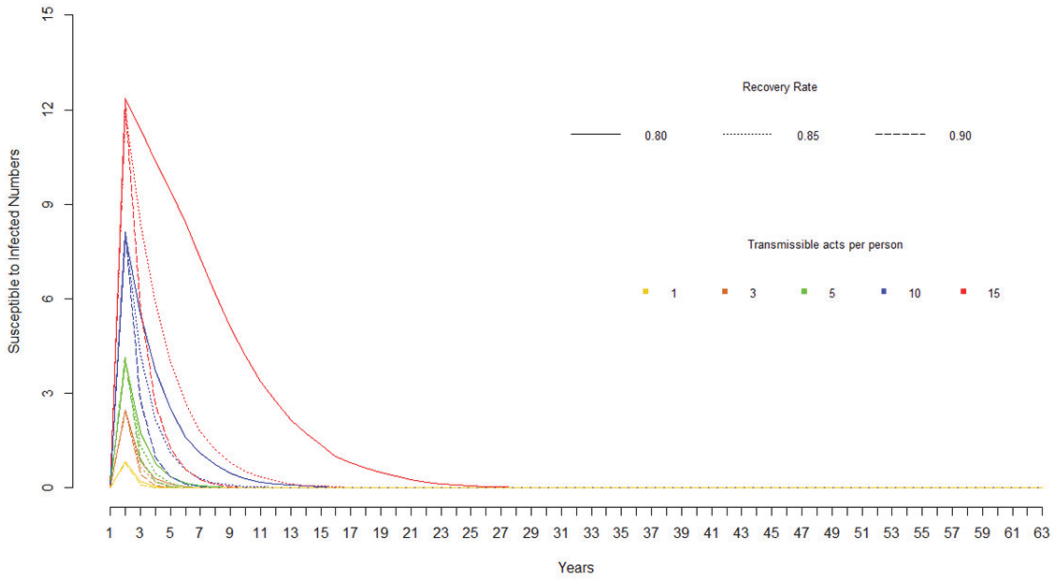


Figure 4.2.7: ICM model representation with observed data for the susceptible to infected number of cases in Kenya

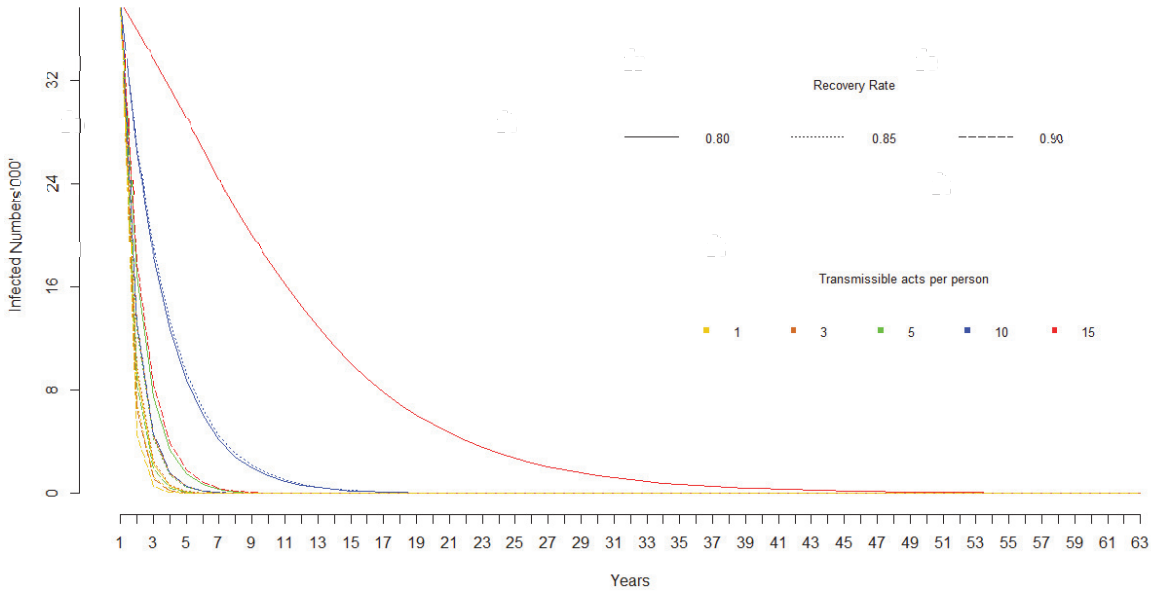


Figure 4.2.8: ICM model representation with observed data for the infected number of cases in Kenya

Figure 4.2.8 represents the infected number of cases for all the transmissible acts

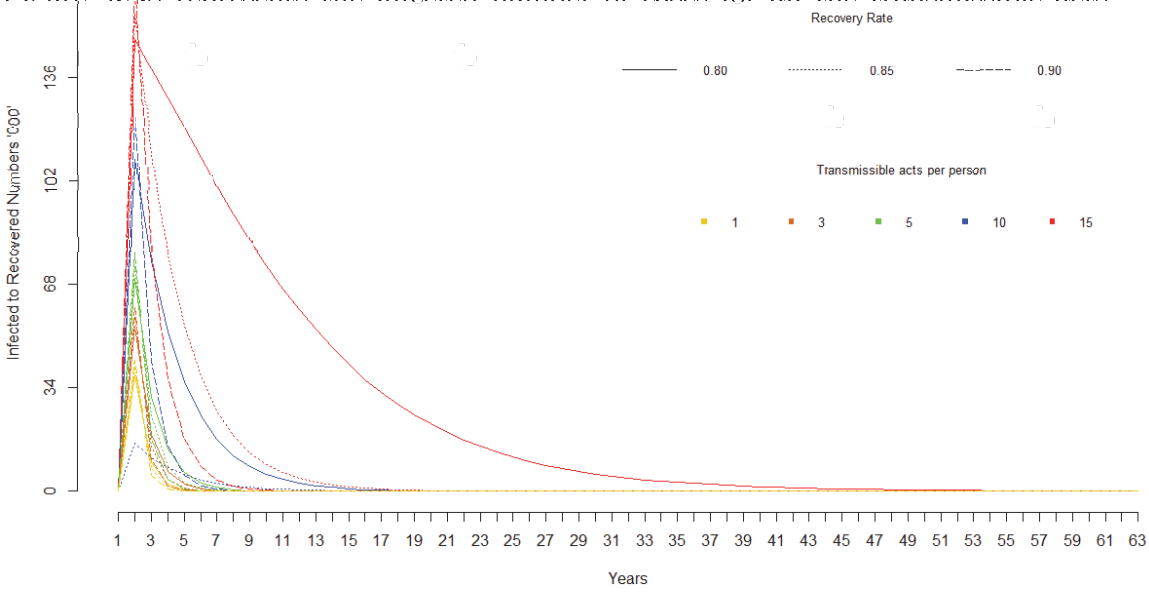


Figure 4.2.9: ICM model representation with observed data for the infected to recovered number of cases in Kenya

Figure 4.2.9 shows that the number which transition from the infected to recovered class dramatically increases and declines after 4 years with the rate for all the transmissible acts except when the recovery rate is 0.80 when the rate of decline is exponential in nature.

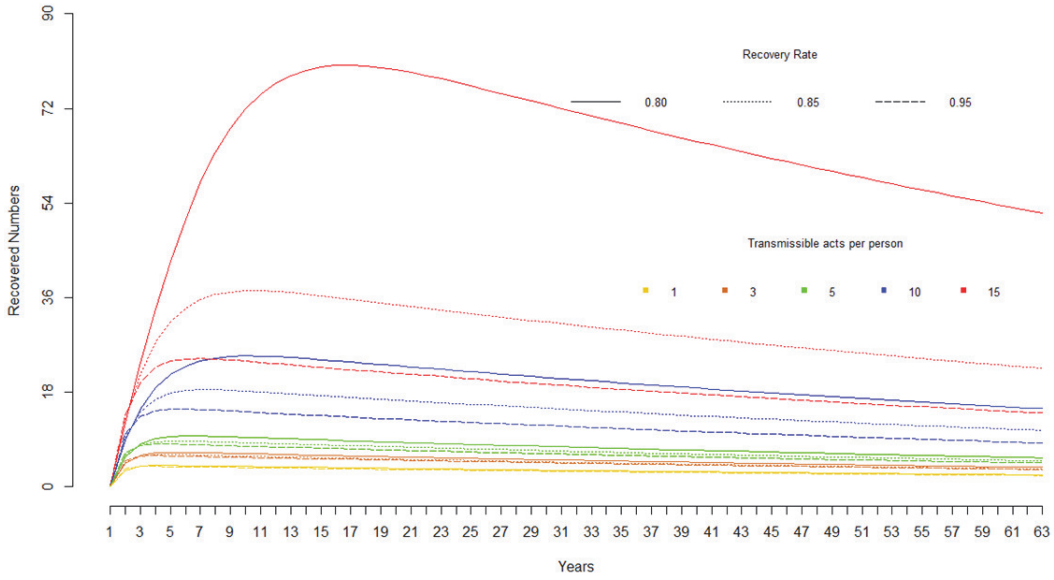


Figure 4.2.10: ICM model representation with observed data for the recovered number of cases in Kenya

Figure 4.2.10 shows that the number of cases that recover increases sharply up to year 7 after which it begins to decline the number of cases which recover decline in a stable manner. The number of cases which recover remains lower for the 1, 2 and 3 transmissible acts. While for the 15 transmissible acts when the recovery rate is 0.80 it remains higher throughout the years.

4.2.3 Small Area Estimation

To have a clear understanding of regional localized variations in the burden of Tuberculosis, variations from 3 different perspectives were looked at namely; distribution

of TB cases, smear positive cases and the treatment outcomes in the presence of risk factors. To obtain the covariates that were included in the spatial temporal model used to obtain relative risks, a generalized linear model was used in each of the three models.

(i) Model Results for All TB cases

Table 4.1 shows that only average BMI was excluded from the spatial temporal model since it was not statistically significant ($p - value > 0.05$). The variables gender, HIV+, DOTs by, Weight, BMI and age were included in the spatial temporal model. Table 4.2 shows the predicted mean risk, 2.5% quadrant, 97.5% quadrant and its associated standard deviation. Table 4.3 shows the model with, temporal effect and both structured and unstructured spatial effects. Table 4.4 shows exponentiated model results of fixed effects.

Table 4.1: GLM model summary for all forms of TB cases

	Estimate	**Std Error**	**Z-value**	**p-value**
Intercept	-13.5957	3.986021	-3.411	0.001
Gender (props)	4.320850	0.232397	18.593	0.001
HIV+(props)	1.244945	0.101580	12.256	0.001
Dots_byprops	8.399804	4.003454	2.098	0.036
Avg_Weight	-0.023816	0.005760	-4.135	0.001
Avg_BMI	0.004818	0.003980	1.211	0.226
Avg_Age	-0.104979	0.005246	-20.013	0.001

Table 4.2: Spatio-temporal fixed effects model summary for all forms of TB cases

	Mean	**sd**	**0.025quant**	**0.975quant**
Intercept	-6.4284	0.8476	-6.4201	-4.7870
Gender (props)	0.5154	0.5544	0.5154	1.6032
HIV+(props)	0.3462	0.3531	0.3516	1.0254
Avg_Weight	0.0044	0.0130	0.0043	0.0301
Avg_Age	-0.0414	0.0150	-0.0415	-0.0418
DOTS_BY	0.0001	0.0000	0.0002	0.0001

Table 4.3: Spatial effect model summary for all forms of TB cases

Model Hyperparameters	**Mean**	**Sd**	**0.025quant**	**0.975quant**
Spatial effect	4.077	1.002	3.971	6.338
Time	5616.823	8358.246	27073.973	555.377
Unstructured spatial effect	18406.828	18286.296	66605.713	3358.016

Table 4.4: Exponentiated model summary for all forms of TB cases

	Mean	**sd**	**0.025quant**	**0.975quant**
Intercept	0.00162	2.3340	0.0002	0.0083
Gender (props)	1.67434	1.7408	0.5638	4.9690
HIV+(props)	1.41367	1.4235	0.6953	2.7882
Avg_Weight	1.00440	1.0131	0.9792	1.0305
Avg_age	0.95944	1.0151	0.9319	0.9884
DOTS_byprops	1.00011	1.0002	1.0001	1.0002

(ii) Model Results for Smear Positive Tuberculosis Cases

Smear positive cases are the infectious forms of TB and it is important to understand the variations in the relative risks among the different counties. From table 4.5 it shows that gender, HIV+ and average age covariates were statistically significant to be included spatial models ($p - value < 0.05$). While average BMI, average weight and DOTS were not statistically significant ($p - value > 0.05$). Table 4.6 shows the predicted mean 2.5% quadrant and 97.5% quadrant, and its associated standard deviation. Table4.7 shows the model with both structured and unstructured random effects, while table 4.8 shows the model results when the fixed effects models results in table 4.6 were exponentiated.

Table 4.5: GLM model summary for Smear Positive TB cases

	Estimate	**Std Error**	**z-value**	**P-value**
Intercept	-0.388251	4.232537	-0.092	0.927
Gender (props)	4.237643	0.216150	19.605	0.001
HIV+(props)	1.320209	0.104949	12.580	0.001
DOTs_by props	-6.360081	4.233056	-1.502	0.133
Avg_Weight	-0.006348	0.005485	-1.157	0.247
Avg_BMI	0.004890	0.004434	1.103	0.270
Avg_Age	-0.091879	0.005843	-15.725	0.001

Table 4.6: Spatial-temporal fixed effects model summary for smear positive TB cases

	Mean	**sd**	**0.025quant**	**0.975quant**
Intercept	-5.9763	0.5845	-7.1292	-4.8351
Gender (props)	0.0205	0.4808	-0.9237	0.9623
HIV+(props)	0.6257	0.3961	-0.1712	1.3866
Avg_Age	-0.0382	0.0155	-0.0383	-0.0074

Table 4.7: Spatial effect model summary for smear positive TB cases

Model Hyper parameters	**Mean**	**s.d**	**0.025quant**	**0.975quant**
Structured spatial effect	3.064	0.7245	1.87	4.7
Time	9403.734	41500	182.19	60883.48
Unstructured spatial effect	18854.859	18500	1258.555	67555.95

Table 4.8: Exponentiated model summary for smear positive TB cases

	Mean	**s.d**	**0.025quant**	**0.975quant**
Intercept	0.00250	1.794080	0.0025431	0.007946
Gender (props)	1.020683	1.617332	1.0209915	2.6176453
HIV+(props)	1.869636	1.486003	1.8817462	4.0014110
Avg_age	0.962516	1.015668	0.9623974	0.9926654

(iii) Model Results for Favourable Treatment Outcomes

In Tuberculosis control favorable outcomes play an important part in providing an understanding on the performance of the TB control programs. It is important to

understand the distribution of favorable outcomes throughout the county. From table 4.9 it shows that all the covariates were statistically non-significant ($p-value > 0.05$). Thus all the covariates were omitted from the spatial temporal model. Table 4.10 presents the results of the 2.5% quadrant and 97.5% quadrant, its associated mean and standard deviation. Table 4.11 shows the model with both structured and unstructured random effects. Table 4.12 shows the model results when the fixed effects models results in table 4.10 were exponentiated.

Table 4.9: GLM model summary for favourable TB cases outcomes

	Estimate	**Std Error**	**Z-value**	**p-value**
Intercept	0.0340207	0.4182700	0.081	0.935
Gender (props)	-0.0791200	0.2903074	-0.273	0.785
HIV+(props)	-0.1482350	0.1230142	-1.205	0.228
Avg_Weight	-0.0002505	0.0067971	-0.037	0.971
Avg_BMI	0.0005053	0.0042712	0.118	0.906
Avg_Age	-0.0031634	0.0069327	-0.456	0.648
Nutrition props	0.0285180	0.0591302	0.482	0.630

Table 4.10: Spatial-temporal fixed effects model summary for favourable outcomes of TB cases

	Mean	**sd**	**0.025quant**	**0.975quant**
Intercept	-0.1502	0.0112	-0.1722	-0.1282

Table 4.11: Spatial effect model summary for favourable outcomes of TB cases

Model Hyper parameters	**Mean**	**s.d**	**0.025quant**	**0.975quant**
Structured spatial effect	19777.28	18826.60	69638.15	2.913
Time	20138.93	20142.33	2649.11	75694.55
Unstructured spatial effect	23138.93	19033.72	1665.00	70555.62

Table 4.12: Exponentiated model summary for favourable outcomes of TB cases

	Mean	**sd**	**0.025quant**	**0.975quant**
Intercept	-0.1502	0.0112	-0.1722	0.861

(iv) Spatial Temporal Distribution Maps

To display the distribution of notified cases, case notification rates, relative risks and TB/HIV distribution spatial temporal maps were produced. The calculations of the relative risks were premised on the significant covariates found to be significant and used in spatial temporal models used to generate the relative risks, posterior probability means and the associated standard deviations which facilitated the generation of the spatial temporal maps which was overlaid on the Kenya county administrative map.

Figure 4.2.11: Distribution of counties in Kenya

Figure 4.2.11 shows the distribution of administrative counties in Kenya.

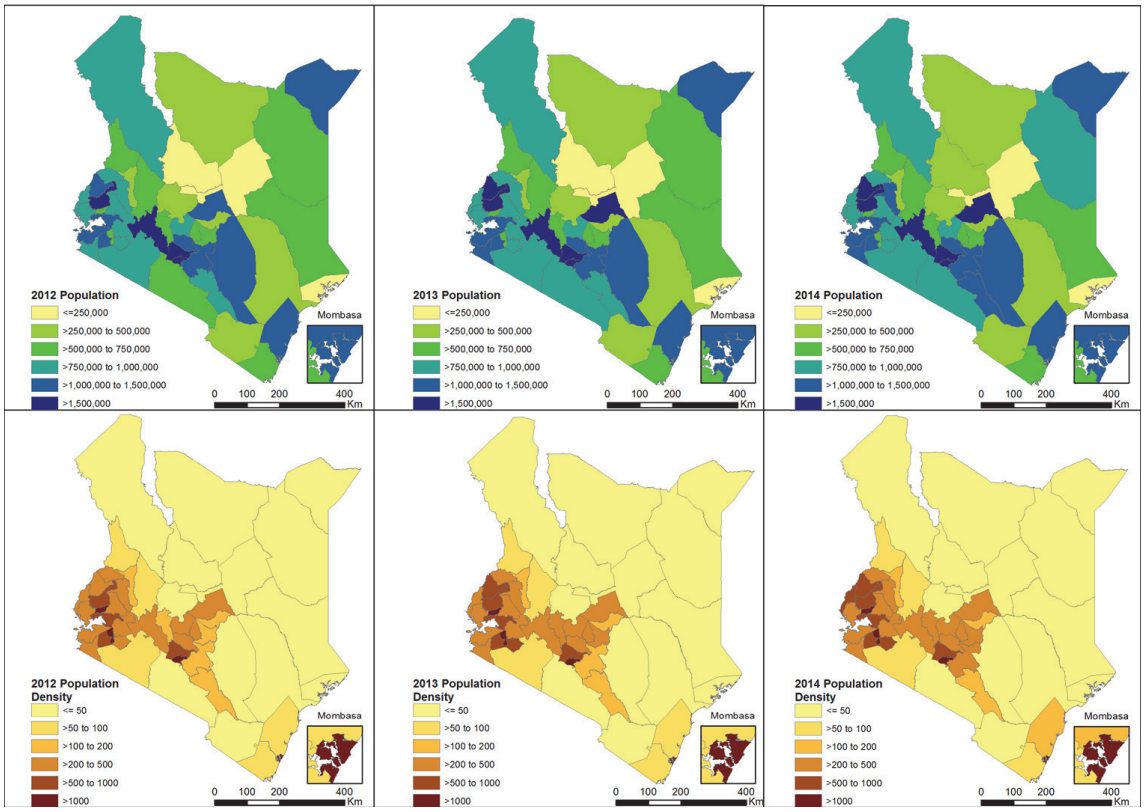


Figure 4.2.12: Kenya Population Distribution

Figure 4.2.12 above shows the population distribution in Kenya with respect to population and population density.

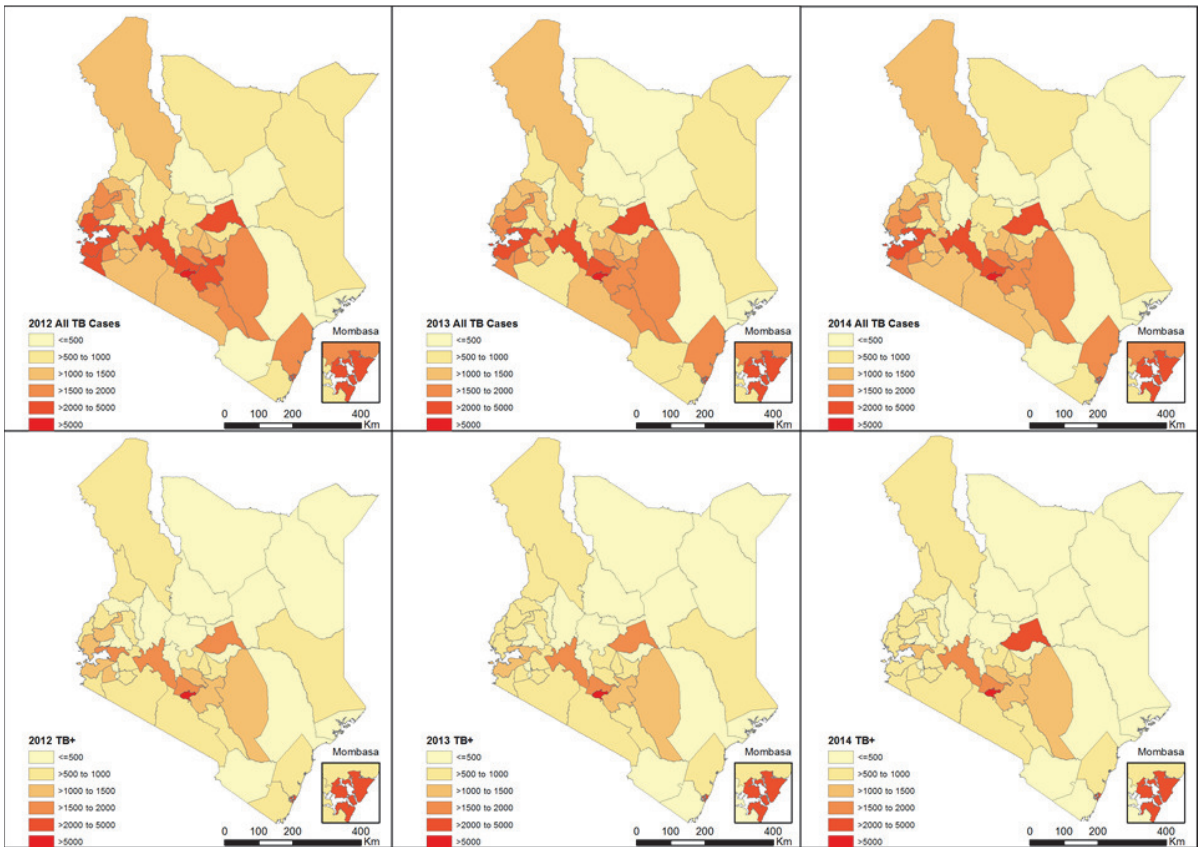


Figure 4.2.13: Distribution of all TB cases and smear positive TB cases in Kenya, 2012-2014

Figure 4.2.13 above shows the spatial temporal distribution over the three years (2012-2014) for all forms of TB and smear positive TB show similar pattern for the notified TB cases with close to 10 counties showing a huge and remaining constant TB burden over the three years. This counties could benefit from more focused TB control strategies.

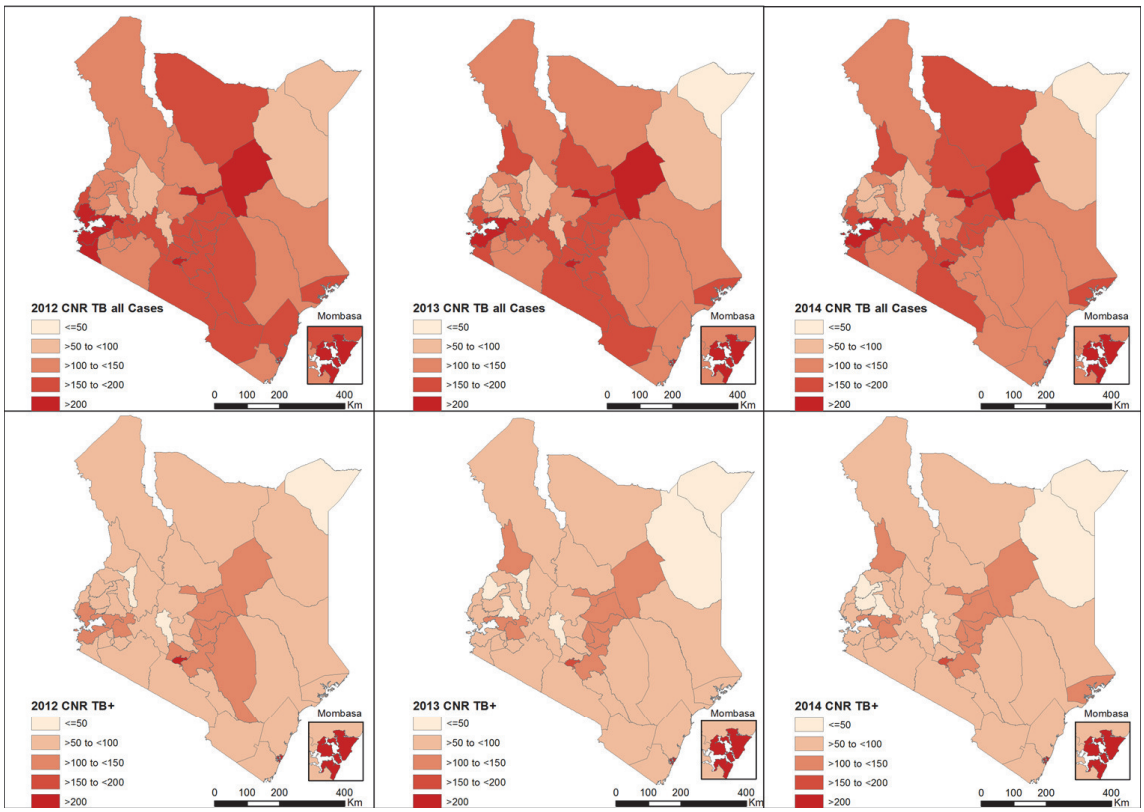


Figure 4.2.14: Case notification rate for all TB cases and smear positive TB cases in Kenya, 2012-2014

When case notification rates for both all forms and smear positive were considered the picture changes significantly for additional number of counties, clearly showing where more effort should be put by the national TB control program. It shows that the highest case notification rates were in the following counties; Nairobi, Mombasa, Lamu, Machakos, Kisumu, Isiolo, West Pokot increasing in the last two years.

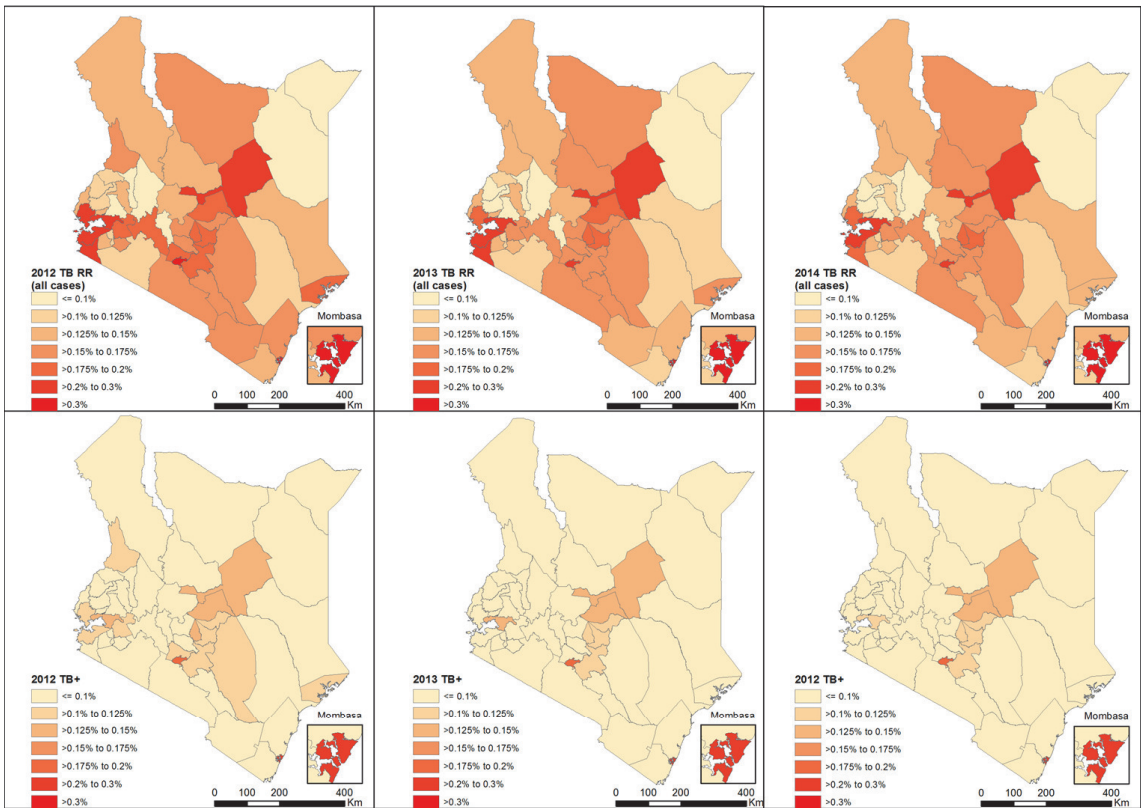


Figure 4.2.15: Relative risk for all TB cases and smear positive TB cases in Kenya, 2012-2014

From Figure 4.2.15, above shows that the estimated risk of case notification rates per 100,000 was highest in the following counties Marsabit, Isiolo, Nairobi, Lamu, Mombasa, Machakos, Kajiado, Makueni, Kisumu, Siaya and Homabay and it shows that over the three years it seems the dynamics of TB disease has not been addressed in these counties.

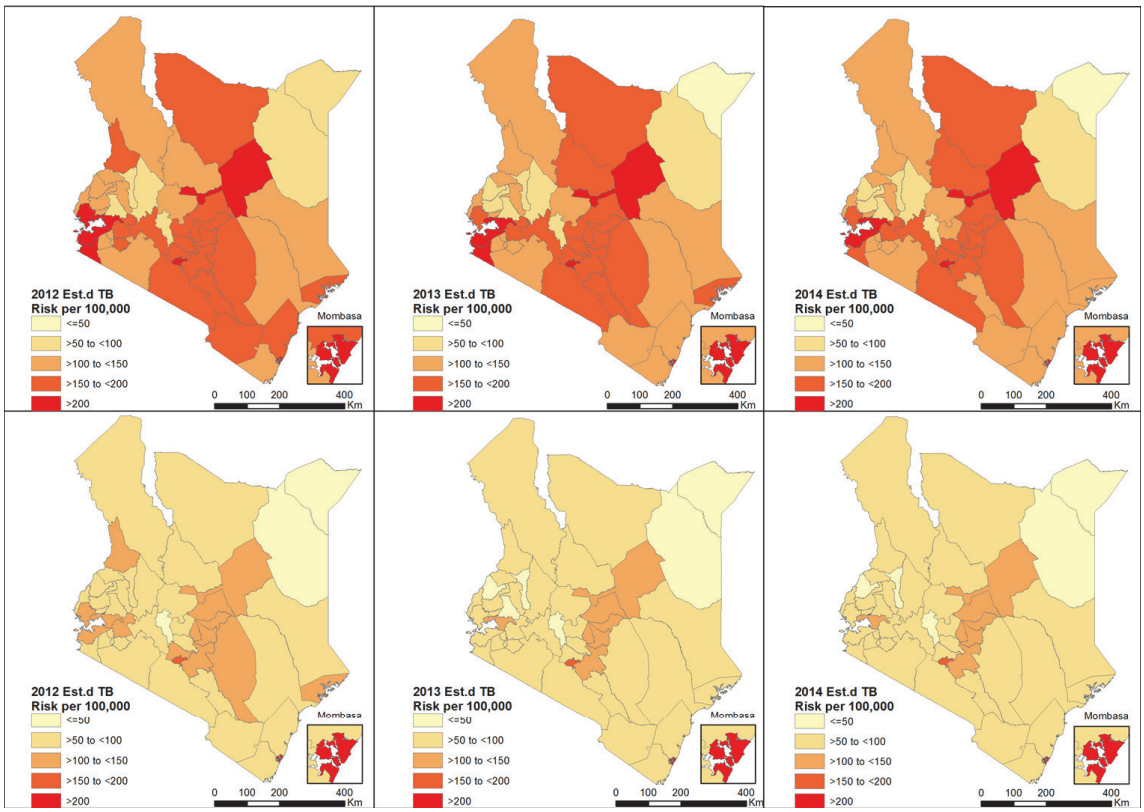


Figure 4.2.16: Relative risk for case notification rates for TB cases and smear positive TB cases in Kenya, 2012-2014

Figure 4.2.16, shows that the estimated risk all forms of TB and smear positive TB were consistent with the results of relative risk for the case notification rates shown in figure 4.2.14 with still the highest relative risk being reported in the following counties Marsabit, Isiolo, Nairobi, Lamu, Mombasa, Machakos, Kajiado, Makueni, Kisumu, Siaya and Homabay and it shows that over the three years it seems the dynamics of TB disease has not been addressed in these counties.

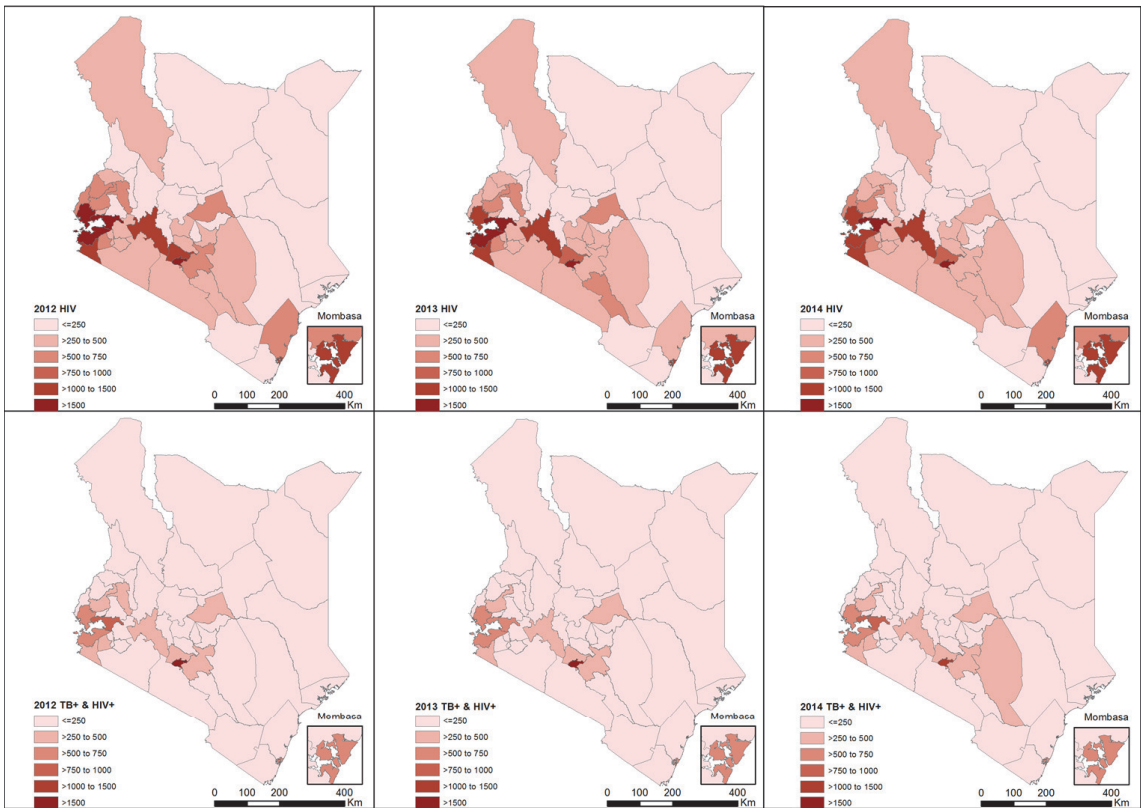


Figure 4.2.17: Distribution of the number of all TB-HIV+ cases and HIV + smear positive TB in Kenya, 2012-2014

In Figure 4.2.17, above it shows that HIV continues to be a major factor in spatial temporal distribution of tuberculosis given its major contribution to the burden of TB on HIV patients. The distribution shows that most of the counties with high burden of TB have also high burden of HIV. It is however notable that although not all high burden counties are high HIV it is important to then understand the dynamics of what is driving the TB epidemic in those counties

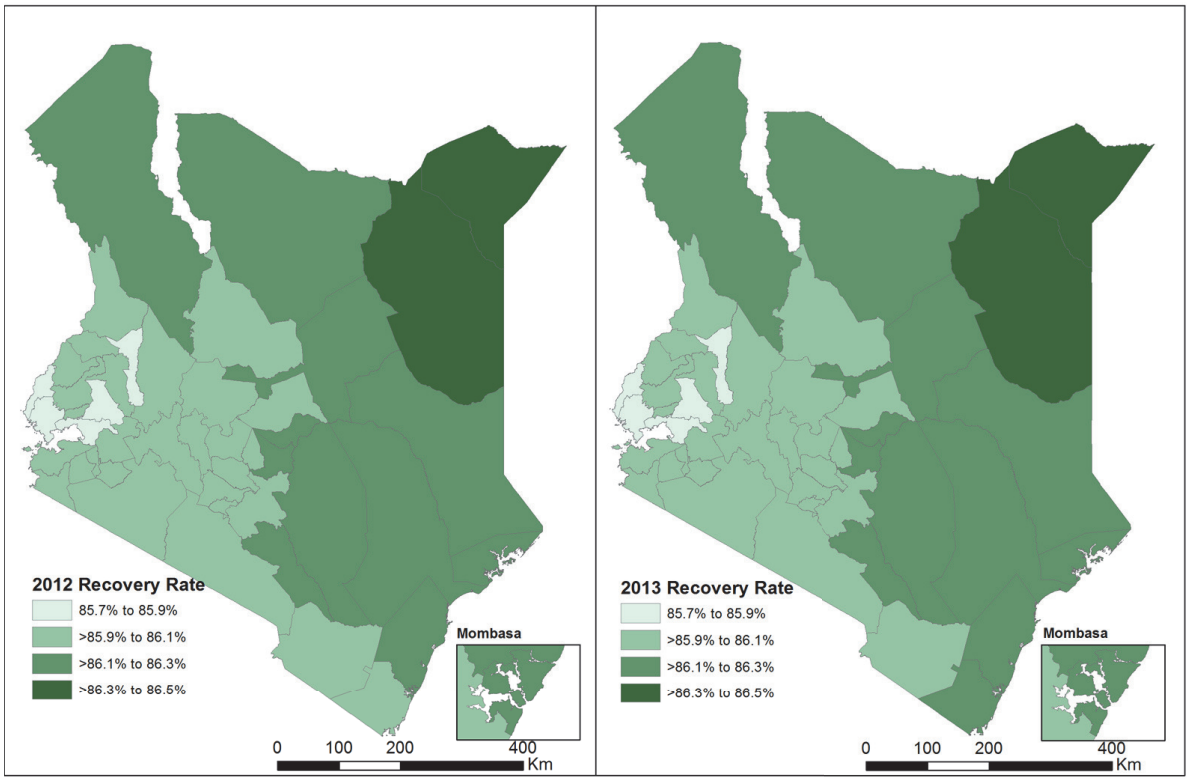


Figure 4.2.18: Distribution of TB recovery rates in Kenya, 2012-2013

In this Figure 4.2.18, it shows that all the counties continue to post very good treatment outcomes herein referred to as recovery rates. In the two years where the treatment outcomes were available it showed that none of the counties reported recovery rate less than 85%. This result is very encouraging given the fact that for successful TB control you require to cure or successful treat almost all you TB patients if you are to significantly cut down the rates of transmission in the communities.

4.2.4 Branching Process Results

In this section we present results from the framework on looking at the reproduction numbers R_0. This quatity is used to measure if the infection will die out eventually or not. When $R_0 < 1$, it shows that the infection will die out in the long run. But if $R_0 > 1$, the infection will be able to continue to spread in a population. And generally, the larger the R_0, the harder it is to control the epidemic. The approaches used to estimate initial reproduction numbers included Exponential growth, Maximum likelihood estimation and Sequential bayesian method. Act.rate in this context is used to represent the number of cases a that are infected from a single case in a year.

(i) Reproduction Numbers for Smear Positive Tuberculosis Monthly notified Cases : 2012-2014

Table 4.13 shows that when monthly data was considered both methods (exponenetial and maximum likelihood) gave R_0 values close to unity. It is notable however that the maximum likelihood method had wider confidence interval when the transmissible acts per person were more than 3 new infections per each contact case. The best fitting values of R_0 were realised when the transmissible acts were 1 and 3 respectively as shown in table 4.14. figure 4.2.19 shows that the value of R_0 was variable throughout the months with some months showing values of R_0 below unity. It was notable that wide confidence levels were realised towards the end of the time period.

Table 4.13: Reproduction numbers for the monthly notified smear positive cases: 2012-2014

	Method	**RO**	**2.5% CI**	**97.5%CI**
Act.rate=1	Exponential Growth	0.9998	0.9998	0.9999
	Maximum Likelihood	1.0095	1.0014	1.0178
Act.rate=3	Exponential Growth	0.9997	0.9996	0.9997
	Maximum Likelihood	1.0508	1.0422	1.0594
Act.rate=5	Exponential Growth	0.9995	0.9994	0.9999
	Maximum Likelihood	1.1203	1.1112	1.1295
Act.rate=10	Exponential Growth	0.9990	0.9989	0.9992
	Maximum Likelihood	1.3281	1.3174	1.3390
Act.rate=15	Exponential Growth	0.9986	0.9983	0.9989
	Maximum Likelihood	1.6113	1.5982	1.6245

Table 4.14: Best fit for reproduction numbers for the monthly notified smear positive cases: 2012-2014

Act.rate	**Time**	**Begin**	**End**	**R**	**Growth Rate**	**R squared**	**2.5%CI**	**97.5%CI**
1	7	18	25	0.905	-0.048	0.339	0.894	0.915
3	7	18	25	0.845	-0.048	0.3393	0.829	0.861
5	7	18	25	0.768	-0.0478	-0.048	0.746	0.792
10	7	18	25	0.606	-0.048	0.339	0.572	0.641
15	7	18	25	0.477	-0.048	0.339	0.439	0.519

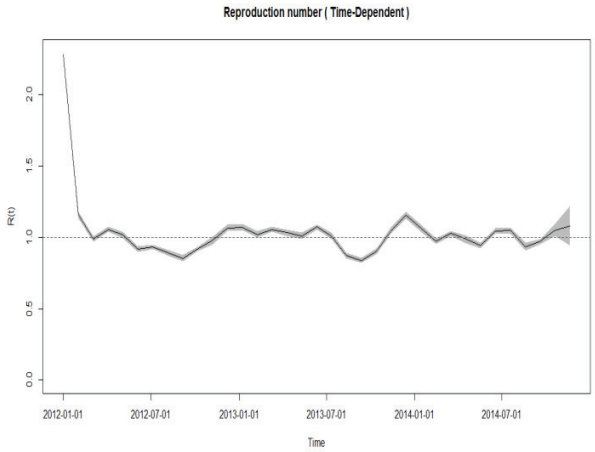


Figure 4.2.19: Trend of Reproduction Numbers for Monthly notified TB cases: 2012-2014

(ii) Reproduction Numbers for Total Tuberculosis yearly notified Cases : 1987-2014

Table 4.15 shows that when year data was considered for the period (1987-2014) both methods (exponenetial and maximum likelihood) gave R_0 values slightly higher than unity. It is notable however that the maximum likelihood method had wider confidence interval when the transmissible acts per person were more than 3 new infections per each contact case. The best fitting values of R_0 were realised for all the transmissible acts as shown in table 4.16. Figure 4.2.20 shows that the value of R_0 has been declining over the years with the value of R_0 falling slightly below unity since the year 2005. It was notable that over the last few years it has shown signs of rising again.

Table 4.15: Reproduction numbers for the yearly notified tuberculosis cases: 1987-2014

	Method	**RO**	**2.5% CI**	**97.5%CI**
Act.rate=1	Exponential Growth	1.0003	1.0003	1.0003
	Maximum Likelihood	1.0753	1.0717	1.0789
Act.rate=3	Exponential Growth	1.0006	1.0006	1.0006
	Maximum Likelihood	1.1665	1.1626	1.1704
Act.rate=5	Exponential Growth	1.0009	1.0009	1.0009
	Maximum Likelihood	1.3284	1.3241	1.3329
Act.rate=10	Exponential Growth	1.0017	1.0017	1.0018
	Maximum Likelihood	2.0852	2.0783	2.0921
Act.rate=15	Exponential Growth	1.0026	1.0025	1.0026
	Maximum Likelihood	4.2773	4.2631	4.2915

Table 4.16: Best fit for reproduction numbers for the yearly notified tuberculosis cases: 1987-2014

Act.rate	Time	Begin	End	R	Growth Rate	R squared	2.5%CI	97.5%CI
1		6	18	1.197	0.088	0.996	1.192	1.202
3		6	18	1.356	0.088	0.996	1.346	1.365
5	6	12	18	1.617	0.088	0.996	1.599	1.636
10	6	12	18	2.514	0.088	0.996	2.461	2.570
15	6	12	18	3.909	0.088	0.996	3.786	4.037

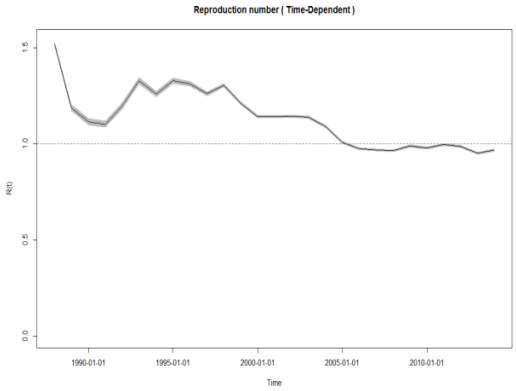


Figure 4.2.20: Trend of Reproduction Numbers for yearly notified TB cases: 1987-2014

4.3 Discussion of the Results

The burden of TB disease in Kenya has shown consistent decline in the last 8 years. This result is consistent with the findings by Sitienei et al. (2013). The key disease burden measures incidence, prevalence and mortality has shown marked decline over the last couple of years, however prevalence and mortality have high levels of uncertainty WHO (2012). Although mortality estimates have high levels of uncertainty, there has been marked decline in the in the estimates of TB deaths among the HIV patients. This in part could be attributed to high HIV testing and initiation of ART treatment among HIV positive patients WHO (2012). It is expected that with a

robust surveillance system, notifications should provide good approximations to TB incidence. Kenya over the last 14 years has notified cumulatively 1,378,592 TB cases. Year by year trend has shown that Kenyan TB epidemic peaked in the year 2007 and has been on a decline since then with an average decline of 5% over the last 8 years. The highest decline in the TB case notification was reported between the years 2012-2013 at 12%. There has been consistent trend in the case notification rates from both the Kenyan TB data and WHO estimates DLTLD (2013). In this thesis, it has been further demonstrated that the proportion of childhood TB has remained fairly consistent over the years. With more attention currently being directed at childhood TB through better diagnostic tools and improvement of surveillance systems it could reduce the gap in understanding childhood TB epidemiology. (Glaziou et al. (2015), WHO (2013)). In Kenya, a large proportion of TB has been those who have been confirmed to have TB using smear microscopy examination with this proportion ranging (37-43) %, the extra-pulmonary ranges from (15-18)% and the smear negative TB cases ranging between (33-42)%. Kenya began the disaggregated reporting for age and sex in the year 2008, while adopting the case based surveillance system in the year 2012. DLTLD (2013). The disaggregation has clearly shown that tuberculosis continues to take greater toll on those in the age groups 24-44 years, this being the most economically productive group. The average age of a TB patient has increased from 33.7 to 37.7 years in the years 2012 and 2013 respectively. Fewer children aged 0-4 years have consistently been reported to have tuberculosis. From the age sex case notification rates it shows that tuberculosis continues to be a male disease with rates in males being approximately 1.5 times higher in particular the most productive age groups 25-54 years. This trend is similar for both all forms of TB and smear positive. It was however noted that the age sex distribution for smear

positive TB was bimodal for females while unimodal for males. This result is consistent with results in Sitienei et al. (2013). However for all forms of TB the peakedness of the distribution is unimodal at the same age groups 25-34 years. At present there is no plausible explanation for the difference in the age sex distribution peakedness for smear positive TB Sitienei et al. (2013). It can however be hypothesized that it could be because of early sexual activity in female which predisposes them to HIV and consequently tuberculosis. There is thus need for further studies to explain the difference. Other possible hypothesis that have been advanced include biological differences in the risk of infection and subsequent TB disease Thorson et al. (2007), differences due to gender roles in different communities which predisposes men to TB disease (Borgdorff et al. (2000), Connolly and Nunn (1996)) and the hypothesis that it could be that women access services to a greater extent than men in some settings or that disease progression could be slower on average in women Glaziou et al. (2015). Sitienei et al. (2013) showed that HIV is the single most important factor that caused resurgence of HIV in Kenya and there is a twin relationship between HIV and tuberculosis. This historical twin relationship has been shown by K'Oyugi and Muita (2002). To understand TB epidemiology in Kenya, understanding the close relationship between TB and HIV is critical. The data from the surveillance system show that upon the adoption in Kenya of the WHO interim policy on TB/HIV collaborative activities in 2004, WHO (2004) with Kenya adopting the policy in 2005. The HIV testing rose from 83 to 94% with HIV sero prevalence declining from 45% to 38% by the end of 2012. It had further been shown that those countries with high HIV prevalence have also high burden of tuberculosis Narain and Lo (2004). Given the role HIV plays in the development of TB disease provision of ART and INH prophylaxis has increasingly played the role in the prevention of TB in individuals

with HIV and latent TB infection. (Glaziou et al. (2015),Straetemans et al. (2010), Glaziou et al. (2011)). Kenya has made significant strides in addressing the dual burden with 74% of the HIV positive TB patients accessing ART within TB clinics by the end of 2012 DLTLD (2013).

The association between poverty and tuberculosis has been well established and widespread in literature of TB risk factors. It has been shown that TB prevalence is significantly higher among people living below the poverty line compared with those above the poverty line and the situation is even more dire among the marginalised people, where TB could be 1.5 times more prevalent Muniyandi and Ramachandran (2008). Further, it has been shown TB can contribute to moving individuals into poverty by reducing patients' physical strength and ability to carry on normal routine work (Paton and Ng (2006); Sagbakken et al. (2008); WHO (2005); Hansel et al. (2004)).

Modelling of tuberculosis has had a long history and among the first model to be developed first in the 1960s (Waaler et al. (1962); Brogger (1967)). Key TB model formulations appeared quickly thereafter. Waaler et al. (1962) developed a linear model for TB based on three compartments: susceptible, infected non-cases and infectious cases. ReVelle et al. (1967) clearly formulated the connection between TB prevalence and infection rate in his model using a set of differential equations. The studies on modelling tuberculosis have tried to provide insights into disease transmission dynamics. Among the earlier work that mark the beginning of modern approach to modelling and thoughtful consideration of probabilities in modelling have be highlighted by (Waaler et al. (1962); Waaler and Piot (1969); Azuma (1975); Styblo (1990)). In our study we constructed both deterministic and stochastic models which can enable us obtain insights and obtain the integral quantities description of

TB transmission dynamics. The results we obtained indicate that both deterministic and stochastic fit well to the Kenyan TB epidemic model however with varying time periods. The models show that for deterministic model the number of infected individuals increases dramatically within three years and begins to fall quickly when the transmissible acts are 10 and 15 and falls to close to zero by 15 years but when the transmissible act is 5 the number infected peaks by the 11th year and declines to zero by year 31, Figure 4.2.3. while for stochastic models the number infected falls exponentially but when the transmissible acts is 15 the decline is slow and will get to zero by the 53rd year while for 10 transmissible acts to declines to zero by the 18th year. While the other transmissible acts (1,3,5) decline to zero by the 9th year. Figure 4.2.8. The results obtained are consistent with other findings in other studies which have attempted to investigate the broad-level transmission dynamics of tuberculosis, Porco and Blower (1998) by conducting a time-dependent uncertainty and sensitivity analysis based on previous work. Model formulation framework, individuals can fully recover from TB, and the model does not include treatment parameters. Their results indicated that most parameters do not significantly affect the severity of the TB epidemic; those that do so include: disease reactivation rate, fraction of infected individuals who develop TB soon after infection (instead of a prolonged latent period), number of individuals that an infectious individual infects per year, disease death rate, and population recruitment rate i.e. the transition from the susceptible pool to those who are infected. Model results were in rough agreement with historical case rate data and developing country data. Because transmission is a function of both contact rate and infectivity, Aparicio et al. (2000) formulated a deterministic cluster model to specifically explore the impact of intense and long exposure to individuals with active TB on population level transmission

dynamics. In contrast to Porco and Blower (1998) which had consistent assumptions that there is an average number of people who get infected each year from the untreated case, this model does not assume an average number of individuals infected per year from one infectious case. Specifically, this model differentiates between epidemiologically active clusters (defined as active when one member has active TB infection) and casual infections. Model results indicate that casual infections may be as or more important than cluster-generated secondary infections at a population level. In our study we considered the assumption that the population is randomly mixing i.e. homogeneous and there is a constant recovery rate of those who are infected at 80, 85 or 90% because of chemotherapy and those who remain continue within the community to infect others. From our results we hypothesize that the transmissible acts can range between 1 to 5 for any infectious case if left untreated. The results we obtained have congruence with the spatial stochastic model, developed by Schinazi (2002) which explored the role of social clusters in disease transmission. Similar to Aparicio et al. (2000), their results indicate that three parameters that influence the transmission of TB: the size of each individual's social cluster, and the infection rates within and outside of the cluster. When the infection rate is low outside the cluster, an epidemic is only possible when the average cluster size and within-cluster infection rate are large enough. They then compared this to the mean field model with corresponding parameters (homogenous mixing, except by cluster), and discover that the qualitative model behavior is unchanged, indicating that the model results are robust to mixing heterogeneity. The results we present here however did not explicitly focus on the effects of heterogeneity in demographically distinct populations. More work must be done to better understand the dynamics of disease spread in heterogeneous populations including the exogenous factors that

affect disease transmission, and implicit population heterogeneity. This could hopeful augment more clearer understanding of TB epidemiology.

The results show the inequality of TB disease distribution both all TB forms and smear positive in Kenya as shown in figures 4.2.15, 4.2.14 and 4.2.17. The results clearly show the counties at risk of severe TB disease in the future. In this study it has showed that some of the key covariates that were found to be significant were HIV+, gender and age. These covariates are some of the known risk factors for developing TB disease. Although this study did not examine the role of socio economic status it showed that TB tends to be more pronounced in those areas where population is highest and also where there is high incidence of HIV. In other studies, they have shown that there is a strong correlation between the measures of income, education and social vulnerability. (Roza et al. (2012),. Corbett et al. (2003);Houben et al. (2009)). While several studies have attempted to utilise spatial analytical tools using the GIS framework to describe the pattern of various infectious diseases in the African region (Rogers and Williams (1993); Beyers et al. (1996); Munch et al. (2003); Gaudart et al. (2005);Gaudart et al. (2006)), only a handful of studies have attempted to obtain spatial temporal review of tuberculosis.

It has been noted that with improved surveillance and improvement in the access to and utilization of TB control services could lead to decrease in the transmission ongoing in the communities since with access and utilization of services, individuals who are sick are able to access care quickly hence cutting down the rates of transmission. One possible hypothesis is that those areas showing high burden to TB could be having challenges of access and utilization of TB services. (Tanrikulu et al. (2007); Vargas et al. (2004)). In our study we have demonstrated that national programs can utilize spatial temporal tools to identify hot spots for TB hence better focusing

of the interventions. With this illustration of cluster analysis and mapping of the disease distribution it adds value beyond that which can be obtained by presenting the disease notification rates or notification numbers reported in a table, as a cluster analysis helps identify areas with unusually high disease rates, which have less likely occurred by chance Cromley and McLafferty (2012). It has been documented that that improving TB control efforts could help reduce the transmission and change the geographic distribution of TB, Jacobson et al. (2005). Further, despite different intervention programs aimed at reducing disease transmission and improving case detection over many years, the unusually high rates of the disease could persisted in the same places, with the most likely spatial clusters showing a stable pattern if the actual dynamics of disease transmission are not addressed. Dangisso et al. (2015). Further, it is clear that the reproduction numbers as shown figures 4.2.19 and 4.2.20 are just around unity meaning than it seems only infectious forms just replace itself and more intesified TB case finding needs to be implemented to cut down the transmissions which continue to take place in the community. Despite the interesting results, our study was faced with a number of challenges which included understanding the relationship between the economic status and tuberculosis distribution. Further there was need for understanding the health seeking behavior of patients, access to and utilization of health services in particularly TB control services. These are the areas that require further research to further understand the dynamics with govern the occurrence of tuberculosis transmission.

Chapter 5

SUMMARY, CONCLUSION AND RECOMMENDATIONS

5.1 Summary

The burden of TB disease in Kenya has shown consistent decline in the last 8 years. The key disease burden measures incidence, prevalence and mortality has shown marked decline over the last couple of years, however prevalence and mortality have high levels of uncertainty. Although mortality estimates have high levels of uncertainty, there has been marked decline in the in the estimates of TB deaths among the HIV patients. This in part could be attributed to high HIV testing and initiation of ART treatment among HIV positive patients. It is expected that with a robust surveillance system, notifications should provide good approximations to TB incidence. Kenya over the last 14 years has notified cumulatively 1,378,592 TB cases. Year by year trend has shown that Kenyan TB epidemic peaked in the year 2007 and has been on a decline since then with an average decline of 5% over the last 8 years. The highest decline in the TB case notification was reported between the years 2012-2013 at 12%. The disaggregation has clearly shown that tuberculosis continues to take greater toll on those in the age groups 24-44 years, this being the most economically productive group. The average age of a TB patient has increased from 33.7 to 37.7 years in the years 2012 and 2013 respectively. Fewer children aged 0-4 years have consistently been reported to have tuberculosis. From the age sex case notification rates it shows that tuberculosis continues to be a male disease with rates in males being approximately 1.5 times higher in particular the most productive age groups 25-54 years.

It has been shown that TB prevalence is significantly higher among people living below the poverty line compared with those above the poverty line and the situation

is even more dire among the marginalised people, where TB could be 1.5 times more prevalent. The results from this study indicate that both deterministic and stochastic fit well to the Kenyan TB epidemic model however with varying time periods. The models show that for deterministic model the number of infected individuals increases dramatically within three years and begins to fall quickly when the transmissible acts are 10 and 15 and falls to close to zero by 15 years but when the transmissible act is 5 the number infected peaks by the 11th year and declines to zero by year 31.

While for stochastic models the number infected falls exponentially but when the transmissible acts is 15 the decline is slow and will get to zero by the 53rd year while for 10 transmissible acts to declines to zero by the 18th year. While the other transmissible acts (1,3,5) decline to zero by the 9th year. Model results indicate that casual infections may be as or more important than cluster-generated secondary infections at a population level. In our study we considered the assumption that the population is randomly mixing i.e. homogeneous and there is a constant recovery rate of those who are infected at 80, 85 or 90% because of chemotherapy and those who remain continue within the community to infect others. From our results we hypothesize that the transmissible acts can range between 1 to 5 for any infectious case if left untreated.

The results show the inequality of TB disease distribution both all TB forms and smear positive in Kenya. The results clearly show the counties at risk of severe TB disease in the future. In this study it has showed that some of the key covariates that were found to be significant were HIV+, gender and age. These covariates are some of the known risk factors for developing TB disease.

In our study we have demonstrated that national programs can utilize spatial temporal tools to identify hot spots for TB hence better focusing of the interventions.

5.2 Conclusion

Kenya's TB epidemiology has evolved over time and it has been characterised by a period where there was increase in the TB cases reaching a peak in the year 2007 after which there was a decline which began to accelerate in the year 2011. The gains which have been witnessed should be sustained. What is equally notable is the clear epidemiologic shift in age indicating reduced transmission in the younger age groups. To sustain the gains made in containing TB in Kenya, efforts must also be made in addressing the risk factors majorly HIV and what has been demonstrated that the burden of HIV on TB patients has been addressed in the country but focus should be on universal access to ART for those who are eligible and access to the INH prophylaxis. Further, to sustain the gains social determinants in health must be addressed given that TB affects the most vulnerable in the population and should include strengthening partnership across all the sectors that are involved. Further systematic analysis should be carried out to understand the socio economic dynamics of TB patients in order to address the well-being of the TB patients. Kenya is on course to achievement of millennium development goals (MDGs) of halting and beginning to reverse the incidence and prevalence to tuberculosis. To better understand mortality, greater focus should be given to improvement of mortality statistics though strengthening of collaboration with the civil registration system in the country and supporting the improvement in the capacity to certify and code cause of death in health facilities.

From this study we can conclude that if the national control program continues with the current interventions it could take them about upto the next 31 years to being the infection numbers to zero if the deterministic model is considered while in the stochastic model with accelerated interventions and high recovery rate and assuming

that there is no change in the risk factors it could take them upto 11 years to bring the infections to zero. Despite the significant increase in the information from molecular TB studies, some major questions remain unresolved because the natural history of the disease makes its comprehensive study difficult coupled with surveys based on tuberculin skin test are hard to interpret because of the cross-reactivity of BCG vaccine; case notification data underestimates TB burden in the country. Thus there is need to have well detailed models which include both environmental and biological factors into the understanding of the TB transmission dynamics.

It is very clear from our study that the growing field of small area estimation epidemiology brings in person, place, and time aspects in the risk factor analysis. The results show that small area estimation enables the drawing of risk maps, which can be used by policy makers to target and develop interventions which address the real challenges which occur in the public health arena. The results indicate that for Kenya there exists several geographically differentiated TB epidemics rather than a single one, implying the existence of TB hot spots which should also have differentiated interventions. The study joins a small but growing number of studies with similar spatial specificity utilizing small area estimation including mapping the spatial variation of proximate and underlying factors that influence tuberculosis dynamics.

5.3 Recommendations

To sustain the gains made in containing TB in Kenya, efforts must also be made in addressing the risk factors majorly HIV and what has been demonstrated that the burden of HIV on TB patients has been addressed in the country but focus should be on universal access to ART for those who are eligible and access to the INH prophylaxis. Further, to sustain the gains social determinants in health must be

addressed given that TB affects the most vulnerable in the population and should include strengthening partnership across all the sectors that are involved. Further systematic analysis should be carried out to understand the socio economic dynamics of TB patients in order to address the well-being of the TB patients.

To better understand mortality, greater focus should be given to improvement of mortality statistics though strengthening of collaboration with the civil registration system in the country and supporting the improvement in the capacity to certify and code cause of death in health facilities and consider initiation of verbal autopsies to obtain an understanding of cause of death amongst those deaths which occur at home. The national TB control program should consider utilizing spatial models at all levels (National, County and Sub-county levels) in monitoring the unfolding TB epidemic as it provides more insights as to wheather the interventions are working or not. More interventions particularly those that cut down transmission of TB should be agressively implmentented in those counties with the highest burden of smear positive TB. This will be achieved by ensuring that there are tailor made interventions for for different counties.

The National TB control program should monitor the value of R_0 at all levels as this particular measure informs as to whether the war against TB control is being won or not. Despite the significant increase in the information from molecular TB studies, some major questions remain unresolved because the natural history of the disease makes its comprehensive study difficult coupled with surveys based on tuberculin skin test are hard to interpret because of the cross-reactivity of BCG vaccine; case notification data underestimates TB burden in the country. Thus there is need to have well detailed models which include both environmental and biological factors into the understanding of the TB transmission dynamics.

5.4 Areas for Further Research

The results from this study did not explicitly focus on the effects of heterogeneity in demographically distinct populations. More work must be done to better understand the dynamics of disease spread in heterogeneous populations including the exogenous factors that affect disease transmission, and implicit population heterogeneity. This could hopeful augment more clearer understanding of TB epidemiology.

There is need to carry out further studies to examine the relationship and the role between of socio economic status in determining the burden of Tuberculosis

Further there was need for understanding the health seeking behavior of patients, access to and utilization of health services in particularly TB control services and to deduce why Tuberculosis continues to disproportinately affect more male than female. These are the areas that require further research to further understand the dynamics with govern the occurrence of tuberculosis transmission.

REFERENCES

Anderson, H. and Tom, B. (2000). *Stochastic Epidemic models and their Applications*. Stockholm.

Anderson, R. and May, R. M. (1991a). *Infectious Diseases of Humans: Dynamics and Control.* Oxford University Press, London.

Anderson, R. and May, R. M. (1991b). *Infectious Diseases of Humans: Dynamics and Control.* Oxford: Oxford University Press.

Aparicio, J., Capurro, A., and Castillo-Chavez, C. (2000). Transmission and dynamics of tuberculosis on generalized households. *Journal of Theoretical Biology*, 206:327-341.

Azuma, Y. (1975). A simple simulation model of tuberculosis epidemiology for use without large-scale computers. *Bull World Health Organ*, 52:313-22.

Bailey, N. (1975). *Mathematical Theory of Infectious Diseases 2nd edn*. Griffin, London.

Ball, F. and Donnelly, P. (1995). Strong approximations for epidemic models. *Stochastic Processes and their Applications*, 55(1):1 - 21.

Ball, F. and Lyne, O. D. (2001). Stochastic multi-type sir epidemics among a population partitioned into households. *Advances in Applied Probability*, 33(1):99-123.

Bartlett., M. (1949). Some evolutionary stochastic processes. *J. R. Statist. Soc*., 11:193-202.

Bernoulli, D. (1766). Essai d'une nouvelle analyse de la mortalite ecausee par la petite verole.me. *Math.Phys.Acad.Roy.Sci.*, 1:1–29.

Beyers, N., Gie, R., Zietsman, H., and et al. (1996). The use of a geographical information system (gis) to evaluate the distribution of tuberculosis in a high-incidence community. *South African Medical Journal*, 86:41–44.

Borgdorff, M., Nagelkerke, N., Dye, C., and Nunn, P. (2000). Gender and tuberculosis: a comparison of prevalence surveys with notification data to explore sex differences in case detection. *International Journal of Tuberculosis and Lung Disease*, 4(2):123–132.

Brezger, A., Kneib, T., and Lang, S. (2005). Bayesx: Analyzing bayesian structured additive regression models. *Journal of statistical software*, 14(11):1–22.

Brogger, S. (1967). Systems analysis in tuberculosis control: A model. *American Review of Respiratory Disease.*, 95:419–434.

Cadmus, S., Yakubu, M., Magaji, A., Jenkins, A. O., and van Soolingen, D. (2010). Mycobacterium bovis, but also mycobacterium africanum present in raw milk of pastoral cattle in north central nigeria. *Tropical Animal Health and Production*, 42(6):1047–8.

Cameletti, M., Lindgren, F., Simpson, D., and Rue, H. (2013). Spatio-temporal modeling of particulate matter concentration through the spde approach. *AStA Advances in Statistical Analysis*, 97(2):109–131.

Caraco, T., Duryea, M., Gardner, G., Maniatty, W., and Szymanski, B. (1998). Host spatial heterogeneity and extinction of an sis epidemic. *Journal of Theoretical Biology*, 192:351–361.

Clancy, D. and O'Neill, P. (1998). Approximation of epidemics by in homogeneous birth-and-death processes. *Stochastic Processes and their Applications*, 73(2):233 – 245.

Cnaan, A., Laird, N., and Slasor, P. (1997). Using the general linear mixed model to analyze unbalanced repeated measures and longitudinal data. *Statistics in Medicine.*, 16:2349–2380.

Connolly, M. and Nunn, P. (1996). Women and tuberculosis. *World Health Stat Q*, 49:115–119.

Corbett, E., Watt, C., Walker, N., Maher, D., Williams, B., and Raviglione, M. (2003). The growing burden of tuberculosis: global trends and interactions with the hiv epidemic. *Archives of internal medicine.*, 163(9):1009–21.

Cromley, E. and McLafferty, S. (2012). *GIS and Public Health, Second edition ed.* The Guilford press.

D'Alembert, J. (1761). *Sur L'application du calcu ldes probabilites l'a Innoculation de la petite vlerole.* Number 9. Opuscules Math lematiques t.2, David., Paris.

Dangisso, M., Datiko, D., and Lindtjorn, B. (2015). Spatio-temporal analysis of smear-positive tuberculosis in the sidama zone, southern ethiopia. *PLoS ONE*, 10(6).

Datta, G. S. (2009). Model-based approach to small area estimation. *In Handbook of Statistics: Sample Surveys: Inference and Analysis*, 29B(D.PfeffermannandC.R.Rao,eds.)(29B):251–288.

DLTLD (2009). *Division of Leprosy, TB and Lung Disease Annual Report*. Kenya Ministry of Health, Nairobi.

DLTLD (2013). *Division of Leprosy, TB and Lung Disease Annual Report*. Kenya Ministry of Health, Nairobi.

Dooley, S., G.K., C., M., H. D., J.M., M., A.J, P., and E.D, S. (1990). Guidelines for preventing transmission of tuberculosis in health-care settings, with special focus on hiv-related issues. *Morbidity and Mortality Weekly Report*, 39(RR-17):1–29.

Eichner, M. and Dietz, K. (1996). Eradication of poliomyelitis: When can one be sure that polio virus transmission has been terminated? *American Journal of Epidemiology*, 143:816–822.

Farrington, C. P. and Grant, A. D. (1999). The distribution of time to extinction in subcritical branching processes: applications to outbreaks of infectious disease. *Journal of Applied Probability*, 36(3):771–779.

Fasina, F., Shittu, A., Lazarus, D., Tomori, O., Simonsen, L., and Viboud, C. (2014). Transmission dynamics and control of ebola virus disease outbreak in nigeria, july to september 2014. *Euro Surveill.*, 19:20920.

Fong, Y., Rue, H., and Wakefield, J. l. (2009). Bayesian inference for generalized linear mixed models. *Biostatistics*, 0:1–16.

Gaudart, J., Poudiougou, B., Dicko, A., and et al. (2006). Space-time clustering of childhood malaria at the household level: a dynamic cohort in a mali village. *BMC Public Health*, 6(286).

Gaudart, J., Poudiougou, B., Ranque, S., and Doumbo, O. (2005). Oblique decision trees for spatial pattern detection: optimal algorithm and application to malaria risk. *BMC Medical Research Methodology*, 5(22).

Ghosh, M. and Rao, J. N. K. (1994). [small area estimation: An appraisal]: Rejoinder. *Statist. Sci.*, 9(1):90–93.

Glass, K., Becker, N., and Clements (2007). Predicting case numbers during infectious disease outbreaks when some cases are undiagnosed. *Statistics in Medicine*, 26:171–183.

Glaziou, P., Floyd, K., Korenromp, E. L., Sismanidis, C., Bierrenbach, A. L., Williams, B., and Atun, Rand Raviglione, M. (2011). Lives saved by tuberculosis control and prospects for achieving the 2015 global target for reducing tuberculosis mortality. *Bulleting of the World Health Organization*, 89(8):573–582.

Glaziou, P., Sismanidis, C., Floyd, K., and Raviglione, M. (2015). Global epidemiology of tuberculosis. 5(2).

Goel, N. and Richter-Dyn, N. (1974). *stochastic models in Biology.* Academic Press.

Gomez, R. and Lopez, Q. A. (2006). Empirical and full bayes estimators for disease mapping. *In International workshop on spatio-temporal modelling (METMA3).*

Grilli, L., Metelli, S., and Rampichini, C. (2014). Bayesian estimation with integrated nested laplace approximation for binary logit mixed models. *Journal of Statistical Computation and Simulation*, pages 1–9.

Hamer, W. (1906). The milroy lectures on epidemic disease in england - the evidence of variability and persistence of type. *The Lancet*, 1:733– 739.

Hansel, N., Wu, A., Chang, B., and GB, D. (2004). Quality of life in tuberculosis: patient and provider perspectives. *Quality of life research: an international journal of quality of life aspects of treatment, care and rehabilitation*, 13(3):639–652.

Hethcote, H. (2000). The mathematics of infectious diseases. *SIAM Review*, 42(4):599–653.

Hethcote, H. W. (1976). Qualitative analyses of communicable disease models. *Math. Biosci.*, 28:335–356.

Houben, R., Crampin, A., Mallard, K., Mwaungulu, J., Yates, M., Mwaungulu, F., and et al. (2009). Hiv and the risk of tuberculosis due to recent transmission over 12 years in karonga district, malawi. *Transactions of the Royal Society of Tropical Medicine and Hygiene.*, 103(12):1087–99.

Jacob, C. (2010). Branching processes: Their role in epidemiology. *Int. J.Environ. Res. Public Health*, 2010(7):1186 – 1204.

Jacobson, L., deLourdes, G.-G., Hernandez-Avila, J., Cano-Arellano, B., Smal, l. P., Sifuentes-Osornio, J., and et al. (2005). Changes in the geographical distribution of tuberculosis patients in veracruz, mexico, after reinforcement of a tuberculosis control programme. *Tropical medicine and international health: TM and IH.*, 10(4):305–11.

Jagers, P. (1975). *Branching Processes with Biological Applications.* John Wiley and Sons., London.

Jiang, J. and Lahiri, P. (2006). Mixed model prediction and small area estimation. *TEST*, 15(1):1–96.

Keeling, M. J. and Danon, L. (2009). Mathematical modelling of infectious diseases. *British Medical Bulletin*, 92(1):33–42.

Klaus, D. and Heesterbeek, J. (2002). Daniel bernoulli's epidemiological model revisited. *Math Biosci.*, 180:1–21.

Klinker, F. (2010). Generalized linear mixed models for ratemaking: A means of introducing credibility into a generalized linear model setting. *In Casualty Actuarial Society E-Forum*, 2.

K'Oyugi, B. and Muita, J. (2002). "the impact of a growing hiv/aids epidemic on the kenyan children. *AIDS, public policy and child well-being*.

Laubenbacher, R. and Stigler, B. (2004). A computational algebra approach to the reverse engineering of gene regulatory networks. *Journal of Theoretical Biology*, 229(4):523–537.

Lehtonen, R. and Veijanen, A. (2009). Design-based methods of estimation for domains and small areas. *In Sample Surveys: Inference and Analysis: Handbook of Statistics*, 29B(D.PfeffermannandC.R.Rao,eds.)(29B):219–249.

Lewis, J. and Glass, L. (1991). Steady states, limit cycles, and chaos in models of complex biological networks. *International Journal of Bifurcation and Chaos*, 1(2):477–483.

Li, M. Y., Graef, J., Wand, L., and Karsai, J. (1999). Global dynamics of a seir model with varying total population size. *Mathematical Biosciences*, 160:191–215.

MACET (2009). *Latent Tuberculosis Infection:A Guide for Massachusetts Providers*. Massachusetts Medical Advisory Committee for the Elimination of Tuberculosis, Massachusetts.

Marschner, I. C. (1992). The effect of preferential mixing on the growth of an epidemic. *Mathematical Biosciences*, 109(1):39 – 67.

Martino, S. and Rue, H. (2009). Implementing approximate bayesian inference using

integrated nested laplace approximation: A manual for the inla program. *Department of Mathematical Sciences,*.

Martins, T. G., Simpson, D., Lindgren, F., and Rue, H. (2013). Bayesian computing with inla: new features. *Computational Statistics and Data Analysis*, 67(10):68–83.

McKendrick, A. (1926). Applications of mathematics to medical problems. *Proceedings of the Edinburgh Mathematical Society*, 14:98–130.

Mode, C. J. (1971). *Multi-Type Branching Processes-Theory and Application*. JAmerican Elsevier Publishing Company, Inc.

Molison, D. (1991). Dependence of epidemic and population velocities on basic parameters. *Mathematical Biosciences*, 107:255–287.

Muller, J. and Kirkilionis, M. . (2000). Ring vaccination. *Journal of Mathematical Biology*, 41:143–171.

Munch, Z., Van-Lills, W., Booysen, C., Zietsman, H., Enarson, D., and Beyers, N. (2003). Tuberculosis transmission patterns in a high incidence area: a spatial analysis. *The International Journal of Tuberculosis and Lung Disease*, 7:271–277.

Muniyandi, M. and Ramachandran, R. (2008). Socioeconomic inequalities of tuberculosis in india. *Expert Opinion on Pharmacotherapy*, 9(10):1623–1628.

Murray, J. (1989). *Mathematical Biology*. Springer Verlag., Berln.

Murray, J., Stanley, E., and Brown, D. (1986). On the spatial spread of rabies among foxes. *Proceedings of the Royal Society, London*, 1255:111–150.

Muyembe-Tamfum, J., Mulangu, S., Masumu, J., Kayembe, J., Kemp, A., and Paweska, J. (2012). Ebola virus outbreaks in africa: past and present. *Onderstepoort J Vet Res*, 79:6–13.

Narain, J. and Lo, Y. (2004). Epidemiology of hiv-tb in asia. *Indian Journal of Medical Research.*, 120(4):277–289.

NASCOP (2013). *Kenya Aids Indicator Survey.* Kenya Ministry of Health, Nairobi.

NASELL, I. (1999). On the time to extinction in reccurent epidemics. *J. Roy. Stat.Soc.*, 61B:309–330.

Nepal Central Bureau of Statistics, W. B. (2013). Nepal Central Bureau of Statistics.

Nisbet, R. M. and Gurney, W. (1982). Modelling fluctuating populations.

O'Neill, P. and Roberts, G. (1999). Bayesian inference for partially observed stochastic epidemics. *Journal of the Royal Statistical Society*, A162:121–129.

Paton, N. and Ng, Y. (2006). Body composition studies in patients with wasting associated with tuberculosis. *Nutrition (Burbank, Los Angeles County, Calif)*, 22(3):245–251.

Pfeffermann, D. (2013). New important developments in small area estimation. *Statist. Sci.*, 28(1):40–68.

Porco, T. and Blower, S. (1998). Quantifying the intrinsic transmission dynamics of tuberculosis. *Theoretical Population Biology.*, 54:117–132.

Rao, J. (2003). *Small Area Estimation.* Wiley, Inc.

Rao, J. N. K. (2005). *Small Area Models*, pages 75 –94. John Wiley and Sons, Inc.

ReVelle, C., Lynn, W., and Feldmann, F. (1967). Mathematical models for the economic allocation of tuberculosis control activities in developing nations. *American Review of Respiratory Disease.*, 96:893–909.

Rogers, D. and Williams, B. (1993). Monitoring trypanosomiasis in space and time. *Parasitology*, 106:77–92.

Roza, D., Caccia, B., and Martinez, E. (2012). Spatio-temporal patterns of tuberculosis incidence in ribeirao preto, state of sao paulo, southeast brazil, and their relationship with social vulnerability: a bayesian analysis. *Rev. Soc. Bras. Med. Trop.*, 45(5):605–615.

Rue, H. and Martino, S. (2007a). Approximate bayesian inference for hierarchical gaussian markov random field models. *Journal of statistical planning and inference*, 137(10):3177–3192.

Rue, H. and Martino, S. (2007b). Approximate bayesian inference for hierarchical gaussian markov random field models. *Journal of statistical planning and inference*, 137(10):3177–3192.

Rue, H., S, M., and N, C. (2009). Approximate bayesian inference for latent gaussian models by using integrated nested laplace approximations. *Journal of the royal statistical society: Series b (statistical methodology)*, 71(2):319–392.

Sagbakken, M., Frich, J., and G, B. (2008). Barriers and enablers in the management of tuberculosis treatment in addis ababa, ethiopia: a qualitative study. *BMC Public Health*, 8(11).

Schinazi, R. (2002). On the role of social clusters in the transmission of infectious diseases. *Theor. Popul. Biol.*, 61:163–169.

Severo, N. C. (1969). Generalizations of some stochastic epidemic models. *Mathematical Biosciences*, 4(3):395 – 402.

Shuaib, F., Gunnala, R., Musa, E., Mahoney, F., Oguntimehin, O., and Nguku, P. (2014). Centers for disease control and prevention (cdc). ebola virus disease outbreak - nigeria, july-september 2014. *MMWR Morb Mortal Wkly Rep*, 63:867–72.

Siettos, C., Anastassopoulou, C., Russo, L., Grigoras, C., and Mylonakis, C. (2015). Modeling the 2014 ebola virus epidemic-agent-based simulations, temporal analysis and future predictions for liberia and sierra leone. *PLOS Currents Outbreaks.*, 1:1–22.

Sitienei, J., Nyambati, V., and Borus, P. (2013). The epidemiology of smear positive tuberculosis in three tb/hiv high burden provinces of kenya (2003 - 2009). *Epidemiology Research International*, 1:1–7.

Straetemans, M., Bierrenbach, A. L., Nagelkerke, N., Glaziou, P., and van der Werf, M. J. (2010). The effect of tuberculosis on mortality in hiv positive people: A meta-analysis. *PLoS ONE*, 5(12):e15241.

Styblo, K. (1990). The global aspects of tuberculosis and hiv infection. *Bulletin of the International Union against Tuberculosis and Lung Disease*, 65:28–32.

Tanrikulu, A., Acemoglu, H., Palanci, Y., and Danli, C. (2007). Tuberculosis in turkey: high altitude and other socio-economic risk factors. *Public Health Journal of The Royal Institute of Public Health*, 122(122):613–9.

Taylor, B. M. and Diggle, P. J. (2013). Inla or mcmc? a tutorial and comparative

evaluation for spatial prediction in log-gaussian cox processes. *Journal of Statistical Computation and Simulation*, pages 1–19.

Taylor, H. and Karlin, S. (1984). *An Introduction to Stochastic Modeling*. Academic Press.

Thorson, A., Long, N. H., and Larsson, L. O. (2007). Chest x-ray findings in relation to gender and symptoms: A study of patients with smear positive tuberculosis in vietnam. *Scandinavian Journal of Infectious Diseases*, 39(1):33–37.

Tomori, O. (2015). Will africa's future epidemic ride on forgotten lessons from the ebola epidemic? *BMC Medicine*, 13(1):116.

Trottier, H. and Philippe, P. (2001). Deterministic modeling of infectious diseases: Theory and methods. *The Internet Journal of Infectious Diseases.*, 1(2).

Umlauf, N., Adler, D., Kneib, T., and Lang, S.and Zeileis, A. (2012). Structured additive regression models: An r interface to bayesx (no. 2012-10). *Working Papers in Economics and Statistics.*

Vargas, M., Furuya, M., and Perez-Guzman, C. (2004). Effect of altitude on the frequency of pulmonary tuberculosis. *The international journal of tuberculosis and lung disease: the official Journal of the International Union against Tuberculosis and Lung Disease.*, 8(11):1321–4.

Waaler, H., Geser, A., and S., A. (1962). The use of mathematical models in the study of the epidemiology of tuberculosis. *Am J Public Health*, 52:1002–13.

Waaler, H. and Piot, M. (1969). The use of an epidemiologic 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*, 41:75–93.

Watson, H. W. and Galton, F. (1875). On the probability of the extinction of families. *Journal of the Anthropological Institute of Great Britain*, 4:138–144.

Watson, H. W. and Galton, F. (1968). The mathematical analysis of an epidemic with two kinds of susceptibles. *Journal of Biometrics*, 24-3:557–566.

WHO (2003). *Treatment of tuberculosis: guidelines for national programmes*. WHO, Geneva.

WHO (2004). *Interim policy on collaborative TB/HIV activities.*, volume WHO/HTM/HIV/2004.1. WHO, Geneva.

WHO (2005). *Addressing Poverty in TB Control. Options for National TB Control Programmes.* WHO, Geneva.

WHO (2006). *The STOP TB Strategy.* WHO, Geneva.

WHO (2007). *Global Tuberculosis Surveilance Report.* WHO, Geneva.

WHO (2012). *Global Tuberculosis Surveilance Report.* WHO, Geneva.

WHO (2013). *Global Tuberculosis Surveilance Report.* WHO, Geneva.

Yanev, G. and Tsokos, C. (1999). Decision-theoretic estimation of the offspring mean in mortal branching processes. *Communications in Statistics- Stochastic Models*, 50:889–902.

APPENDICES

We now present selected R programs used for simulations and real data analyses in this thesis.

A.1 Simulation of Deterministic Compartmental Model Codes

```
#############################################################################
# Deterministic Compartmental Models (DCM) SIR R code
#############################################################################
rm(list=ls()) #remove all the variables from the workspace
#Packages Required
require(EpiModel)
#############################################################################
# Specifying epidemic parameters for Deterministic Compartmental Models.
#############################################################################
# Average number of transmissible acts per person per unit time
act_rate1<-1
act_rate2<-3
act_rate3<-5
act_rate4<-10
act_rate5<-15
#############################################################################
#di.rate
#Death or exit rate for infected
di.rate=0.03
#############################################################################
#ds.rate
#Death or exit rate for suspectible
#Population infected with TB=10
#Population susceptible to TB infection= (1/3*3000)
ds_rate<-(10)/(1/3*1000)
#############################################################################
#dr.rate
#Death or exit rate for recovered
#Population infected with TB=10
#Population susceptible to TB infection= (1/3*3000)=1000
dr_rate<-(10)/(1/3*3000)
```

```
#############################################################################
#bi_rate
#Birth and entry rate
#New infections per year=2.6
#Population suspectible to TB infection=(1/3*3000)
bi_rate<-(2.6)/(1/3*3000)
#############################################################################
# Setting the epidemic parameters for deterministic  compartmental models
param<-param.dcm(inf.prob=0.26,act.rate=10,rec.rate=0.8,
                 b.rate=bi_rate,ds.rate=ds_rate,di.rate=0.03,
                 dr.rate=dr_rate)
#Setting the intial conditions for deterministic compartmental models
init<-init.dcm(s.num=1000,i.num=3,r.num=0)
#Setting the controls for deterministic compartmental models simulated
with dcm
control<-control.dcm(type="SIR",nsteps=63)
#Solving the determinstic compartmental epidemic models
sim<-dcm(param,init,control)
#Saving the simulated stochastic individual contact models results as
an R object
save(sim,file="dcm_85_05.RData")
#Reload R saved bject
load("dcm_85_10.RData")
#Returning sim as a data frame
data<-as.data.frame(sim)
#Prints the data as a csv file
write.csv(data,file="sim model output.csv",row.names=F)
```

A.2 Stochastic Individual Contact Models SIR R Code

```
##########################################################################
# Stochastic Individual Contact Models (ICM) SIR R Code
##########################################################################
rm(list=ls()) #remove all the variables from the workspace
#Packages Required
require(EpiModel)
##########################################################################
# Specifying epidemic parameters for stochastic individual contact models.
##########################################################################
# Average number of transmissible acts per person per unit time
act_rate1<-1
act_rate2<-3
act_rate3<-5
act_rate4<-10
act_rate5<-15
##########################################################################
#di.rate
#Death or exit rate for infected
di.rate=0.03
##########################################################################
#ds.rate
#Death or exit rate for suspectible
#Population infected with TB=10
#Population susceptible to TB infection= (1/3*3000)
ds_rate<-(10)/(1/3*1000)
##########################################################################
#dr.rate
#Death or exit rate for recovered
#Population infected with TB=10
#Population susceptible to TB infection= (1/3*3000)=1000
dr_rate<-(10)/(1/3*3000)
##########################################################################
#bi_rate
#Birth and entry rate
#New infections per year=2.6
#Population suspectible to TB infection=(1/3*3000)
```

```
bi_rate<-(2.6)/(1/3*3000)
##############################################################################
# Setting the epidemic paramters for stochastic individual contact models
param<-param.icm(inf.prob=0.26,act.rate=10,rec.rate=0.8,
                 b.rate=bi_rate,ds.rate=ds_rate,di.rate=0.03,
                 dr.rate=dr_rate)
#Setting the intial conditions for stochastic
individual contact models simulated with icm
init<-init.icm(s.num=1000,i.num=3,r.num=0)
#Setting the controls for stochastic
individual contact models simulated with ICM
control<-control.icm(type="SIR",nsteps=63,nsims=5000)
#Simulating stochastic individual contact models
sim<-icm(param,init,control)
#Saving the simulated stochastic
individual contact models results as an R object
save(sim,file="ICM_85_05.RData")
#Reload R saved bject
load("ICM_85_10.RData")
#Returning sim as a data frame
data<-as.data.frame(sim)
#Prints the data as a csv file
write.csv(data,file="sim model output.csv",row.names=F)
```

A.3 Observed Deterministic Compartmental Models SIR R Code

```
 ##########################################################################
# Observed Deterministic Compartmental Models (DCM) SIR R code
##########################################################################
rm(list=ls()) #remove all the variables from the workspace
#Packages Required
require(EpiModel)
##########################################################################
# Specifying epidemic parameters for Deterministic Compartmental Models.
##########################################################################
# Average number of transmissible acts per person per unit time
act_rate1<-1
act_rate2<-3
act_rate3<-5
act_rate4<-10
act_rate5<-15
##########################################################################
#di.rate #Death or exit rate for infected
di.rate=0.03
##########################################################################
#ds.rate #Death or exit rate for suspectible #Population infected with
TB=142026
#Population susceptible to TB infection= (1/3*44350000)
ds_rate<-(142026)/(1/3*44350000)
##########################################################################
#dr.rate #Death or exit rate for recovered
#Population infected with TB=142046
#Population susceptible to TB infection= (1/3*44350000)
dr_rate<-(142026)/(1/3*44350000)
##########################################################################
#bi_rate #Birth and entry rate #New infections per year=38000
#Population suspectible to TB infection=(1/3*4435000)
bi_rate<-(38000)/(1/3*44350000)
##########################################################################
# Setting the epidemic parameters for deterministic  compartmental models
param<-param.dcm(inf.prob=0.26,act.rate=10,rec.rate=0.8,                  b.rate=
#Setting the intial conditions for deterministic compartmental models
```

```
init<-init.dcm(s.num=1000,i.num=3,r.num=0)
#Setting the controls for deterministic compartmental
models simulated with dcm
control<-control.dcm(type="SIR",nsteps=63)
#Solving the determinstic compartmental epidemic models
dcm<-dcm(param,init,control)
#Saving the simulated stochastic individual contact
models results as an R object
save(dcm,file="dcm_85_05.RData")
#Reload R saved bject
load("dcm_85_10.RData")
#Returning sim as a data frame
data<-as.data.frame(dcm)
#Prints the data as a csv file
write.csv(data,file="dcm model output.csv",row.names=F)
```

A.4 Observed Stochastic Individual Contact Models SIR R Code

```
#############################################################################
# Observed Stochastic Individual Contact Models (ICM) SIR R code
#############################################################################
rm(list=ls()) #remove all the variables from the workspace
#Packages Required
require(EpiModel)
#############################################################################
# Specifying epidemic parameters for stochastic individual contact models.
#############################################################################
# Average number of transmissible acts per person per unit time
act_rate1<-1
act_rate2<-3
act_rate3<-5
act_rate4<-10
act_rate5<-15
#############################################################################
#di.rate #Death or exit rate for infected
di.rate=0.03
#############################################################################
#ds.rate #Death or exit rate for suspectible
#Population infected with TB=142026
#Population susceptible to TB infection= (1/3*44350000)
#ds_rate<-(142026)/(1/3*44350000)
#############################################################################
#dr.rate #Death or exit rate for recovered
#Population infected with TB=142026
#Population susceptible to TB infection= (1/3*44350000)
dr_rate<-(142026)/(1/3*44350000)
#############################################################################
#bi_rate
#Birth and entry rate
#New infections per year=38000
#Population suspectible to TB infection=(1/3*44350000)
bi_rate<-(38000)/(1/3*44350000)
#############################################################################
# Setting the epidemic paramters for stochastic individual contact models
```

```
param<-param.icm(inf.prob=0.26,act.rate=10,rec.rate=0.8,
#b.rate=bi_rate,ds.rate=ds_rate,di.rate=0.03, dr.rate=dr_rate)
#Setting the intial conditions for stochastic
individual contact models simulated with icm
init<-init.icm(s.num=14766667,i.num=38000,r.num=0)
#Setting the controls for stochastic
individual contact models simulated with ICM
control<-control.icm(type="SIR",nsteps=63,nsims=20)
#Simulating stochastic individual contact models
sim<-icm(param,init,control)
#Saving the simulated stochastic
individual contact models results as an R object
save(sim,file="ICM_85_05.RData")
#Reload R saved bject
load("ICM_85_10.RData")
#Returning sim as a data frame
data<-as.data.frame(sim)
#Prints the data as a csv file
write.csv(data,file="sim model output.csv",row.names=F)
```

A.5 SAE INLA R Code for TB Treatment Outcomes in Kenya

```
# SAE INLA R CODE For TB Treatment outcomes in Kenya
rm(list=ls())
# Packages required
library("ctv")
library("sp")
library(maptools)
library(rgdal)
library(spdep)
require(INLA)
setwd("F:/DCM_140415/DCM and ICM plots_040515 and SAE/SAE RESULTS_040515")
ken_data<-read.csv("F:/DCM_140415/DCM and ICM plots_040515
and SAE/SAE RESULTS_040515/Kenya Treatment Outcome_1005.csv")
head(ken_data) names(ken_data)
#GLM CODE ~ Helps in choosing the
covariates to include in the spatial model
#Only include significant covariates
glm_model<-glm(X.Favourable_outcome~malesprop+HIVpos+malesprop+
Avg_Weight+Avg_BMI+Avg_Age+nutprops,
family=poisson(link="log"),offset=log(pop),data=ken_data) summary(glm_model)
# we need to add a new column as an index cannot be
# used twice, i.e the index "SLID" cannot be used
# for two f-function, the structured and unstructured
# random effect. ken_data<-cbind(ken_data, CID.unstruc= ken_data$CID)
ken_data<-cbind(ken_data, poptb= (ken_data$Population)/3) head(ken_data)
ken<- readShapePoly("kenCounty.shp") plot(ken)
adj_ken<-poly2nb(ken)
#Creates adjacency for ken adj_ken
#### plot the neighbourood structure plot(ken, border=gray(.5))
plot(adj_ken, coordinates(ken), add=TRUE)
nb2INLA("ken.graph",adj_ken) #INLA graph file #spdep command
#spatially unstructured
and spatially unstructured and correlated time as random effects
formula<-X.Favourable_outcome~f(CID,model="besag",
graph="ken.graph",adjust.for.con.comp = FALSE)+
f(CID.unstruc,model="iid")+f(Time,model="rw1")
model<-inla(formula,family="poisson",E=pop,
```

```
data=ken_data,control.compute=list(dic=TRUE,cpo=TRUE)
,verbose=TRUE)
save(model, file="model_Favourable1005.RData")
load("model_Favourable1005.RData")
summary(model)
exp(model$summary.fixed)
#calculating and mapping the risk
fitted<-model$summary.linear.predictor$mean
RR<-exp(fitted)
std_dev<-model$summary.fitted.values$sd
ken_data$RR<-RR
ken_data$std_dev<-std_dev
write.table(ken_data,
file="Kenya Treatment Outcome_1305.csv.csv",row.names=F,sep=",")
plot(model)
```

A.6 SAE INLA R Code for All Smear Positive TB Cases in Kenya

```
# SAE INLA R Code for All Smear Positive TB Cases in Kenya
rm(list=ls())
# Packages required
library("ctv")
library("sp")
library(maptools)
library(rgdal)
library(spdep)
require(INLA)
setwd("F:/DCM_140415/DCM and ICM plots_040515 and SAE/SAE RESULTS_040515")
ken_data<- read.csv("F:/DCM_140415/DCM and ICM plots_040515
and SAE/SAE RESULTS_040515/
Kenya TB Positive cases only_1005.csv") head(ken_data) names(ken_data)
#GLM CODE ~ Helps in choosing the covariates to include in the spatial model
#Only include significant covariates
glm_model<-glm(CNRPTB.~Malesprops+HIVprops+dot_byProps+
Avg_Weight+Avg_BMI+Avg_Age,
family=poisson(link="log"),
offset=log(pop),data=ken_data) summary(glm_model)
# we need to add a new column as an index cannot be
# used twice, i.e the index "SLID" cannot be used
# for two f-function, the structured and unstructured
# random effect. ken_data<-cbind(ken_data, CID.unstruc= ken_data$CID)
ken_data<-cbind(ken_data, poptb= (ken_data$Population)/3) head(ken_data)
ken<- readShapePoly("kenCounty.shp") plot(ken)
adj_ken<-poly2nb(ken) #Creates adjacency for ken adj_ken
#### plot the neighbourood structure plot(ken, border=gray(.5))
plot(adj_ken, coordinates(ken), add=TRUE)
nb2INLA("ken.graph",adj_ken) #INLA graph file #spdep command
#spatially unstructured and spatially unstructured
and correlated time as random effects
formula<-CNRPTB.~Malesprops+HIVprops+Avg_Age+
f(CID,model="besag",graph="ken.graph",
adjust.for.con.comp = FALSE)+
f(CID.unstruc,model="iid")+f(Time,model="rw1")
model<-inla(formula,family="poisson",E=pop,
```

```
data=ken_data,control.compute=list(dic=TRUE,cpo=TRUE)
,verbose=TRUE)
save(model, file="model_poisson Positive TB Cases1005.RData")
load("model_poisson Positive TB Cases1005.RData")
summary(model)
exp(model$summary.fixed)
#calculating and mapping the risk
fitted<-model$summary.linear.predictor$mean
RR<-exp(fitted)
std_dev<-model$summary.fitted.values$sd
ken_data$RR<-RR
ken_data$std_dev<-std_dev
write.table(ken_data,file="Kenya TB Positive cases only_1305.csv.csv",
row.names=F,sep=",")
plot(model)
```

A.7 SAE INLA R Code for All TB Cases in Kenya

```
# SAE INLA R CODE For All TB cases in Kenya
rm(list=ls())
# Packages required
library("ctv")
library("sp")
library(maptools)
library(rgdal)
library(spdep)
require(INLA)
setwd("F:/DCM_140415/DCM and ICM plots_040515 and SAE/SAE RESULTS_040515")
ken_data<- read.csv("F:/DCM_140415/DCM and ICM plots_040515
and SAE/SAE RESULTS_040515/Kenya TB_AllcasesWithout_DOTBY_1005.csv")
head(ken_data) names(ken_data)
#GLM CODE ~ Helps in choosing the covariates
to include in the spatial model
#Only include significant covariates
glm_model<-glm(TBPOSK~malesProp+HIVprops+
Dots_byprops+Avg_weight+Avg_BMI+Avg_Age,
family=poisson(link="log"),offset=log(pop),data=ken_data)
summary(glm_model)
# we need to add a new column as an index cannot be
# used twice, i.e the index "SLID" cannot be used
# for two f-function, the structured and unstructured
# random effect. ken_data<-cbind(ken_data, CID.unstruc= ken_data$CID)
ken_data<-cbind(ken_data, poptb= (ken_data$Population)/3) head(ken_data)
ken<- readShapePoly("kenCounty.shp") plot(ken)
adj_ken<-poly2nb(ken)
#Creates adjacency for kenya adj_ken
#### plot the neighbourood structure plot(ken, border=gray(.5))
plot(adj_ken, coordinates(ken), add=TRUE)
nb2INLA("ken.graph",adj_ken) #INLA graph file
#spdep command
#spatially unstructured and spatially unstructured
and correlated time as random effects
formula<-TBAllCases~malesProp+HIVprops+Avg_weight+
Avg_Age+DOTS_BY+f(CID,model="besag"
```

```
,graph="ken.graph",adjust.for.con.comp = FALSE)+
f(CID.unstruc,model="iid")+f(Time,model="rw1")
model<-inla(formula,family="poisson",E=pop,
data=ken_data,control.compute=list(dic=TRUE,cpo=TRUE,
verbose=TRUE)
save(model, file="model_poisson All TB Cases.RData")
load("model_poisson All TB Cases.RData")
summary(model)
names(ken_data)
exp(model$summary.fixed)
#calculating and mapping the risk
fitted<-model$summary.linear.predictor$mean
RR<-exp(fitted)
std_dev<-model$summary.fitted.values$sd
ken_data$RR<-RR
ken_data$std_dev<-std_dev
write.table(ken_data,file="Kenya TB_Allcases_1305.csv.csv"
,row.names=F,sep=",")
plot(model)
```

A.8 Branching Process Code

```
# Program spuRs/resources/scripts/bp.r
# branching process simulation
bp <- function(gen, rv.sim, ...) {
  # population of a branching process from generation 0 to gen
  # rv.sim(n, ...) simulates n rv's from the offspring distribution
  # Z[i] is population at generation i-1; Z[1] = 1
  Z <- rep(0, gen+1)
  Z[1] <- 1
  for (i in 1:gen) {
    if (Z[i] > 0) {
      Z[i+1] <- sum(rv.sim(Z[i], ...))
    }
  }
 return(Z)
}
bp.plot <- function(gen, rv.sim, ..., reps = 1, logplot = TRUE) {
  # simulates and plots the population of a branching process
  # from generation 0 to gen; rv.sim(n, ...) simulates n rv's
  # from the offspring distribution
  # the plot is repeated reps times
  # if logplot = TRUE then the population is plotted on a log scale
  # Z[i,j] is population at generation j-1 in the i-th repeat
  Z <- matrix(0, nrow = reps, ncol = gen+1)
  for (i in 1:reps) {
    Z[i,] <- bp(gen, rv.sim, ...)
  }
  if (logplot) {
    Z <- log(Z)
  }
  plot(c(0, gen), c(0, max(Z)), type = "n", xlab = "generation",
       ylab = if (logplot) "log population" else "population")
  for (i in 1:reps) {
    lines(0:gen, Z[i,])
  }
  return(invisible(Z))
}
```

```
bp.plot(63, rbinom, size=3, prob=.26, reps=1000, logplot = FALSE)
##############################################################################
##Second plot
##############################################################################
bp.sim <- function(gen, rv.sim, ...) {
  # population of a branching process at generation gen
  # rv.sim(n, ...) simulates n rv's from the offspring distribution
  Z <- 1
  for (i in 1:gen) {
    if (Z > 0) {
      Z <- sum(rv.sim(Z, ...))
    }
  }
  return(Z)
}
# set parameter values
gen <- 63
size <- 3
prob <- seq(0.3, 0.6, by = 0.01)
n.reps <- 1000# sample size for estimating E Z
# estimate E Z for each value of prob
mu <- rep(0, length(prob))
Z.mean <- rep(0, length(prob))
for (i in 1:length(prob)) {
  Z.sum <- 0
  for (k in 1:n.reps) {
    Z.sum <- Z.sum + bp.sim(gen, rbinom, size, prob[i])
  }
  mu[i] <- size*prob[i]
  Z.mean[i] <- Z.sum/n.reps
}
# plot estimates
# note that values of log(0) (= -infinity) are not plotted
plot(mu, log(Z.mean), type = "o",      xlab = "E family size"
, ylab = paste("log pop at gen", gen))
```

A.9 Monthly Reproduction Numbers (R0) Code

```
rm(list=ls())
library(R0) library(epitools)
setwd("F:/Branching Process/
Branching Process Simulation_010615/R0")
data<-read.csv("F:/Branching Process/
Branching Process Simulation_010615/R0/HK_MonthlyPTB_110715_1214.csv")
data$TIME<-as.Date(data$Time,format="%d %B %Y")
head(data);tail(data);str(data)
data2<-data$PTB
# check incid in input
check.incid(data2,date.first.obs = "2012-01-01",time.step = 31)
# create generation time : gamma distribution,
with mean 2.6 time units and standard deviation 1 time unit
GT.TB<-generation.time("gamma", c(15,1))
res.R<-estimate.R(data2,GT=GT.TB,begin="2012-01-01",end="2014-12-21",
date.first.obs = "2012-01-01",time.step=31,
methods=c("EG","ML","SB","TD"))
# applies methods EG, ML, SB, TD to the dataset plot(res.R)
# diplays results
res.R
# displays fit to the epidemic curve plotfit(res.R)
#sensitivity analysis according to choice
of time window for exponential growth
sensitivity.analysis(data2, GT.TB, begin=1:18,
end=19:36, est.method="EG", sa.type="time")
#sensitivity analysis according to generation time
# The use of the exact same call as
for the internal sensitivity analysis function
# sa.type = "GT"
# Here we will test GT with means of 1 to 15,
each time with SD constant (1)
# GT and SD can be either fixed value or vectors of values
sensitivity.analysis(data2, GT.type="gamma",
GT.mean=seq(1,36,1), GT.sd.range=1, begin=1, end=36,
est.method="EG", sa.type="GT")
# Actual value in simulations may differ, as they are adapted according
```

```
to the distribution type
tmp<-sensitivity.analysis(sa.type="GT", incid=data2,
GT.type="gamma", GT.mean=seq(1,36,1),
GT.sd.range=1, begin=1, end=36, est.method="EG")
## Results are stored in a matrix,
each line dedicated to a (mean,sd)
couple plot(x=tmp[,"GT.Mean"], xlab="mean GT (Years)",
y=tmp[,"R"], ylim=c(0,1.5), ylab="R0 (95% CI)",
type="p", pch=19, col="black", main="Sensitivity of R0 to mean GT")
arrows(x0=as.numeric(tmp[,"GT.Mean"]),
y0=as.numeric(tmp[,"CI.lower"]),
y1=as.numeric(tmp[,"CI.upper"]),
angle=90, code=3, col="black", length=0.05)
mGT<-generation.time("gamma", c(1,1))
sen<-sensitivity.analysis(sa.type="time", data2,
GT=mGT, begin=1:18, end=19:36,est.method="EG")
# ...
# Warning message: # If 'begin' and 'end' overlap, c
ases where begin >= end are skipped.
# These cases often return Rsquared = 1 and are thus ignored.
## A list with different estimates of reproduction ratio,
exponential growth rate and 95%CI
## With different pairs of begin and end dates in
form of data frame is returned.
## If method is "EG", results will
include growth rate and deviance R-squared measure
plot(sen, what=c("criterion","heatmap"))
## Returns complete data.frame of best R0 value for each time period
## (allows for quick visualization)
## The "best.fit" is the time period over which the estimate is the more robust
# $best.fit
# Time.period Begin.dates End.dates R Growth.rate Rsquared CI.lower. CI.upper.
```

A.10 Yearly Reproduction Numbers (R0) Code

```
rm(list=ls())
library(R0) library(epitools)
setwd("F:/Branching Process/Branching Process Simulation_010615/R0")
data<-read.csv("F:/Branching Process/
Branching Process Simulation_010615/R0/TB Cases 1987_2014.csv")
#data$TIME<-as.Date(data$Time,format="%d %B %Y")
head(data);tail(data);str(data)
data2<-data$Smear.Positive.Pulmonary.TB
# check incid in input
check.incid(data2,date.first.obs = "1987-12-31",time.step = 365)
# create generation time : gamma distribution,
with mean 2.6 time units and
standard deviation 1 time unit GT.TB<-generation.time("gamma", c(1,1))
#Applying the methods
res.R<-estimate.R(data2,GT=GT.TB,begin="1987-12-31",end="2014-12-24",
date.first.obs = "1987-12-31",time.step=365,methods=c("EG","ML","SB","TD"))
# applies methods EG, ML, SB, TD to the dataset
plot(res.R)
# diplays results
res.R
# displays fit to the epidemic curve plotfit(res.R)
#sensitivity analysis according to choice of time
window for exponential growth
sensitivity.analysis(data2, GT.TB, begin=1:14,end=15:28,
est.method="EG", sa.type="time")
#sensitivity analysis according to generation time
# The use of the exact same call as
for the internal sensitivity analysis function
# sa.type = "GT"
# Here we will test GT with means of 1 to 15,
each time with SD constant (1)
# GT and SD can be either fixed value or vectors of values
sensitivity.analysis(data2,
GT.type="gamma",GT.mean=seq(1,28,1), GT.sd.range=1, begin=1,
end=28,est.method="EG", sa.type="GT")
# Actual value in simulations may differ,
```

```
as they are adapted according to the distribution type
tmp<-sensitivity.analysis
(sa.type="GT", incid=data2, GT.type="gamma", GT.mean=seq(1,28,1),
GT.sd.range=1, begin=1, end=28, est.method="EG")
## Results are stored in a matrix, each line dedicated to a (mean,sd)
couple plot(x=tmp[,"GT.Mean"], xlab="mean GT (Years)",
y=tmp[,"R"], ylim=c(1.2, 5), ylab="R0 (95% CI)",
type="p", pch=19, col="black", main="Sensitivity of R0 to mean GT")
arrows(x0=as.numeric(tmp[,"GT.Mean"]), y0=as.numeric(tmp[,"CI.lower"]),
y1=as.numeric(tmp[,"CI.upper"]), angle=90, code=3, col="black", length=0.05)
mGT<-generation.time("gamma", c(1,1))
sen=sensitivity.analysis(sa.type="time", data2, GT=mGT,
begin=1:14, end=15:28,est.method="EG")
# ...
# Warning message:
# If 'begin' and 'end' overlap, cases where begin >= end are skipped.
# These cases often return Rsquared = 1 and are thus ignored.
## A list with different estimates of reproduction ratio,
exponential growth rate and 95%CI
## With different pairs of begin and end dates in
form of data frame is returned.
## If method is "EG", results will include
growth rate and deviance R-squared measure
plot(sen, what=c("criterion","heatmap"))
## Returns complete data.frame of
best R0 value for each time period
## (allows for quick visualization)
## The "best.fit" is the time period over
which the estimate is the more robust
# $best.fit # Time.period Begin.dates
End.dates R Growth.rate Rsquared CI.lower. CI.upper.
```