

**MULTIAPPROACH
COMPUTATIONAL MODELLING OF
TUBERCULOSIS. UNDERSTANDING
ITS EPIDEMIOLOGICAL DYNAMICS
FOR IMPROVING ITS CONTROL IN
NIGERIA**

**Universitat Politècnica de Catalunya
Doctoral Programme in Computational and Applied
Physics**

Nura Mohammad Rabiuh Ahmad

Co-directors: Daniel López Codina
Clara Prats Soler

Department of Physics
Universitat Politècnica de Catalunya BarcelonaTech

Dissertation submitted for Doctorate degree by
Universitat Politècnica de Catalunya

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. This dissertation is my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text and Acknowledgements.

Nura Mohammad Rabiuh Ahmad

June 2020

Acknowledgements

All praise be to Allah S.W.A that spares my life and directed me on this journey, chooses the right and the best instructors for me. I would like to thank my mother Aminatu Yunus for being there for me even when the going gets hard, she is my source of strength. I would also like to thank my father for making sure that we got the righteous up bringing that we deserve.

To my wife, my soulmate and my best friend, for being by my side through out this journey, for the constant support, understanding and compassion, resilient and prayers even when I was far from your reach I was never far from your heart. You are the best wife anyone could have hope for. To my little angels, my children, you are the source of my happiness.

My sincere gratitude goes to the two best people I know, Dr. Daniel Lopez and Dra. Clara Prats, you took me in like a son, you have supported, coached, trained and guided me throughout this journey, I feel like the luckiest person to have come across amazing people like you. I owe everything to the two of you, thank you. My sincere appreciation to the entire BIOCOM-SC research group, for your support and generosity through the journey of this project, thank you so much. I am also grateful for the great opportunity I had to work with people from Barcelona Super Computing Center, such as Dra. Cristina Montañola, I am deeply grateful for allowing me to use your resources and providing me a lot of knowledge and inspirations during my stay at the center, your contribution to this project is deeply appreciated.

I would also like to thank Dr. Mustapha Musa, for joining us in this scientific adventure and showing us hospitality during our visit to his center in Gombe State Nigeria.

I feel really fortunate to have all these amazing scientists by my side, and I hope we can keep our scientific collaborations for future projects.

My sincere appreciation goes to the Centre for Development Cooperation of the UPC and Caldes Solidària for their interest in our project and their financial support during the course of this project.

Last, but not the least, I would like to thank my friends, academic and non-academic for scientific and non-scientific conversations, for mutual support and help in all the aspects of life. Without them, the walk would be much more arduous.

Abstract

Tuberculosis (TB) is an infectious disease that is considered to be the biggest killer of mankind in the history of infectious diseases. It is responsible for more than 1 billion deaths in the last 200 years. Nowadays, there are still more than 10 million new TB cases every year and it causes more than 1.5 million deaths annually, according to World Health Organization estimates. Nigeria, with a persistent incidence of about 219 cases / 10^5 inhabitants for the year 2019, is among the 8 countries that accounted for two-thirds of the new TB cases in 2018. The control of the disease in this country is coordinated by the National Tuberculosis and Leprosy Control Program (NTBLCP). Despite efforts, that are mainly focused in the provision of free DOTS (Directly Observed Treatment, short-course) to persons with active TB that goes to the hospital, the estimated prevalence is around $330/10^5$ population, which is approximately the same as the estimated TB prevalence in 1990 ($323/100,000$).

Epidemiological models can be used for a more precise diagnosis of TB situation in certain territories, as well as to help on the design and evaluation of control actions. In this project, two modeling approaches have been used to these ends. First, a top-down approach at the country level (Nigeria) by means of the design, testing and fitting of several SEIR-type models, aimed to provide a global picture of the situation and to quantify some of the most relevant parameters. Second, a bottom-up approach to a smaller area (Gombe state, in the north-east of the country) by means of the design, testing and fitting of an agent-based model, aimed to unravel a particular context and to help on the design and quantification of control actions. In addition, fieldwork was also carried out in order to look for the particular socio-economic factors that are responsible for the epidemiological situation of TB spread; the statistical analysis of the data obtained is the third approach of this project.

Several SEIR-type models were built in an attempt to progressively characterize and understand TB dynamics in a different context, these models were fitted to both some selected High burden and low Burden countries around the world and, in particular, to the epidemiological situation in Nigeria. The splitting of the infectious (sick) population in two subpopulations, those that are diagnosed (and treated) and those that are not, confirmed a dramatic low notification rate that varies from 16 to 20% during the analyzed years (2000-2010). This factor revealed to be the bottleneck for the control of the disease in this

country. Model's predictions showed no relevant effects of control actions without a previous increase in the notification rate. Fieldwork was designed in coordination with NTBLCP local authorities with the aim of analyzing socio-economic factors that condition such notification rate in Gombe state. It consisted of an initial gathering of data from NTBLCP Gombe state branch, Gombe State Hospitals Management Board, and various hospitals in Gombe state, followed by a set of 52 in-depth interviews with TB patients from different health centers. Obtained data and interviews' outcomes were statistically analyzed using inferential statistics and Anova analysis of mean, with the help of machine learning techniques. Results were devastating: none of the patients interviewed had any knowledge of TB symptoms and 90% had no knowledge of TB transmission mechanisms after talking to the health workers. Mean patients' delay was 9.6 weeks with a standard deviation of 4.8; only 10% of interviewed patients went to the doctor within the first month of feeling sick, 30% within 1-2 months, 20% within 2-3 months, and 40% after feeling sick for more than 3 months. The epidemiological information obtained from the top-down approach and the results derived from the fieldwork were used for adapting an agent-based model (ABM) of TB spreading in the context of Gombe state. The resulting ABM was successfully fitted to the evolution of estimated prevalence and diagnosed cases from 2007 to 2016. Then, it was used to test different interventions aimed to increase the notification rate, decrease the diagnosis delay, and increase the population's awareness regarding TB transmissible. The multi-approach methodology used in this project revealed to be a robust way of tackling a real problem. It was capable of (1) providing a global and detailed picture of TB situation in a certain area, (2) relating model's parameters and outcomes with a real socioeconomic context, and (3) generating a useful tool for helping on the design and evaluation of TB control actions.

Resumen

La tuberculosis (TB) es la enfermedad infecciosa que más muertes ha causado en la historia de las enfermedades infecciosas, conociéndose como *the big killer*. Es responsable de más de mil millones de muertes en los últimos 200 años. Hoy en día, todavía hay más de 10 millones de casos nuevos de TB cada año y causa más de 1.5 millones de muertes anuales, según las estimaciones de la Organización Mundial de la Salud (OMS). Nigeria, con una incidencia persistente de aproximadamente 219 casos / 10^5 habitantes para el año 2019, se encuentra entre los 8 países que representaron conjuntamente dos tercios de los nuevos casos de tuberculosis en 2018. El control de la enfermedad en este país está coordinado por el Programa Nacional de Control de Tuberculosis y Lepra (NTBLCP). A pesar de los esfuerzos, que se centran principalmente en la provisión de DOTS gratuitos (tratamiento observado directamente, curso corto) a personas con TB activa que acuden al hospital, la prevalencia estimada es de alrededor de 330 / 10^5 habitantes, que es aproximadamente igual a la prevalencia estimada de TB en 1990 (323 / 100,000).

Los modelos epidemiológicos se pueden utilizar para un diagnóstico más preciso de la situación de TB en ciertos territorios, así como para ayudar en el diseño y evaluación de acciones de control. En este proyecto, se han utilizado dos enfoques de modelización para estos fines. Primero, un enfoque de arriba hacia abajo (*top-down*) a nivel de país (Nigeria) mediante el diseño, prueba y ajuste de varios modelos tipo SEIR, con el objetivo de proporcionar una imagen global de la situación y cuantificar algunos de los parámetros más relevantes. En segundo lugar, un enfoque de abajo a arriba (*bottom-up*) en una área más pequeña (estado de Gombe, en el noreste del país) mediante el diseño, prueba y ajuste de un modelo basado en agentes, con el objetivo de desentrañar un contexto particular y ayudar en el diseño y cuantificación de las acciones de control. Además, también se realizó un trabajo de campo para investigar los factores socioeconómicos particulares que son responsables de la situación epidemiológica de la propagación de la tuberculosis. El análisis estadístico de los datos obtenidos es el tercer enfoque de este proyecto.

Se construyeron varios modelos de tipo SEIR en un intento de caracterizar y comprender progresivamente la dinámica de la TB en diferentes contextos. Estos modelos se ajustaron a algunos países seleccionados de alta carga y baja carga en todo el mundo, y en particular a la situación epidemiológica en Nigeria. La división de la población infecciosa (enferma) en dos subpoblaciones, la de personas que se diagnostican (y se tratan) y la de las que no, confirman una tasa de notificación baja dramática que varía entre el 16 y el 20 % durante los años analizados (2000-2010). Este factor reveló ser el cuello de botella para el control de la enfermedad en este país. Las predicciones del modelo no mostraron efectos relevantes de las acciones de control sin un aumento previo en la tasa de notificación.

El trabajo de campo se diseñó en coordinación con las autoridades locales de NTBLCP con el objetivo de analizar los factores socioeconómicos que condicionan dicha tasa de notificación en el estado de Gombe. Consistió en una recopilación inicial de datos de la sucursal del estado de NTBLCP Gombe, la Junta de Administración de Hospitales del Estado de Gombe y varios hospitales en el estado de Gombe, seguida de un conjunto de 52 entrevistas en profundidad con pacientes con tuberculosis de diferentes centros de salud. Los datos obtenidos y los resultados de las entrevistas se analizaron estadísticamente utilizando estadísticas inferenciales y análisis de medias de Anova, con la ayuda de técnicas de aprendizaje automático. Los resultados fueron devastadores: ninguno de los pacientes entrevistados conocía los síntomas de la tuberculosis y el 90 % no conocía los mecanismos de transmisión de la tuberculosis después de hablar con el personal sanitario. El retraso medio de los pacientes en acudir al centro médico fue de 9.6 semanas con una desviación estándar de 4.8; solo el 10 % de los pacientes entrevistados acudió al médico dentro del primer mes de sentirse enfermo, el 30 % dentro de 1-2 meses, el 20 % dentro de 2-3 meses y el 40 % después de sentirse enfermo durante más de 3 meses.

La información epidemiológica obtenida del enfoque *bottom-up* y los resultados derivados del trabajo de campo se utilizaron para adaptar un modelo basado en agentes (ABM) de propagación de TB en el contexto del estado de Gombe. El ABM resultante se ajustó con éxito a la evolución de la prevalencia estimada y los casos diagnosticados de 2007 a 2016. Luego, se usó para probar diferentes intervenciones destinadas a aumentar la tasa de notificación, disminuir el retraso del diagnóstico y aumentar la conciencia de la población sobre la TB transmisibles.

La metodología de enfoque múltiple utilizada en este proyecto reveló ser una forma sólida de abordar un problema real. Fue capaz de (1) proporcionar una imagen global y detallada de la situación de la TB en un área determinada, (2) relacionar los parámetros y resultados del modelo con un contexto socioeconómico real, y (3) generar una herramienta útil para ayudar en el diseño y la evaluación de acciones de control de TB.

Resum

La tuberculosi (TB) és la malaltia infecciosa que ha causat més morts a la història de la humanitat. És responsable de més de mil milions de morts en els darrers 200 anys. Actualment, encara hi ha més de 10 milions de nous casos de tuberculosi cada any i provoca més d'1.5 milions de morts anuals, segons les estimacions de l'Organització Mundial de la Salut. Nigèria, amb una incidència persistent d'uns 219 casos / 10⁵ habitants durant l'any 2019, es troba entre els 8 països que van representar dos terços dels nous casos de tuberculosi el 2018. El control de la malaltia d'aquest país està coordinat pel Programa Nacional de Control de la Tuberculosi i la Lepra (NTBLCP). Malgrat els esforços, que es centren principalment en la prestació de DOTS gratuïts (Tractament observat directament, de curta durada) a persones amb tuberculosi activa que van a l'hospital, la prevalença estimada és d'uns 330 per cada 100,000 habitants, que és aproximadament la mateixa prevalença de tuberculosi que s'estimava l'any 1990 (323 / 100,000). Es poden utilitzar models epidemiològics per a un diagnòstic més precís de la situació de tuberculosi en determinats territoris, així com per ajudar en el disseny i avaluació de les accions de control de l'epidèmia.

Inicialment es van desenvolupar diversos models de tipus SEIR per intentar caracteritzar i comprendre progressivament la dinàmica de la tuberculosi en contextos diferents, aquests models s'adaptaven tant a països del món seleccionats amb una alta incidència com a països amb baixa incidència. Aquesta ha estat la primera aproximació per poder estudiar la situació de la tuberculosi a Nigèria, què és l'objectiu del treball.

En aquest projecte, s'han utilitzat dues estratègies de modelització per fer-ho. En primer lloc, una estratègia *top-down* a nivell de país (Nigèria) mitjançant el disseny, proves i adequació de diversos models de tipus SEIR, amb l'objectiu de proporcionar una imatge global de la situació i quantificar alguns dels paràmetres més rellevants. En segon lloc, una estratègia *bottom-up* estudiant una àrea més petita, l'Estat de Gombe, al nord-est del país, mitjançant el disseny, la prova i l'adequació d'un model basat en agents (ABM). El model ABM té per finalitat entendre en detall la situat particular de l'Estat de Gombe i ajudar en el disseny i quantificació de possibles accions de control. A més, també es va realitzar un treball de camp per buscar els factors socioeconòmics particulars responsables de la situació epidemiològica de la tuberculosi; l'anàlisi estadística de les dades obtingudes és la tercera estratègia de treball d'aquest projecte.

La divisió de la població infecciosa (malalta) en dues subpoblacions, les que són diagnosticades (i tractades) i les que no ho són, van confirmar una taxa de notificació dramàticament baixa que varia del 16 al 20 % durant els anys analitzats (2000-2010). Aquest factor es va revelar com el coll d'ampolla per al control de la malaltia en aquest país. Les prediccions del model no mostraven efectes rellevants de les accions de control sense un augment previ de la

taxa de notificació. El treball de camp es va dissenyar en coordinació amb les autoritats locals NTBLCP amb l'objectiu d'analitzar els factors socioeconòmics que condicionen aquesta taxa de notificació a l'estat de Gombe. Va consistir en una recollida inicial de dades de la sucursal estatal NTBLCP a Gombe, la Junta de Gestió d'Hospitals de l'Estat de Gombe i diversos hospitals de l'Estat de Gombe, seguida d'un conjunt de 52 entrevistes en profunditat amb pacients amb TB de diferents centres de salut. Les dades obtingudes i els resultats de les entrevistes es van analitzar estadísticament mitjançant estadístiques inferencials i Anova de mitjana, amb l'ajut de tècniques d'aprenentatge automàtic. Els resultats van ser devastadors: cap dels pacients entrevistats no coneixia els símptomes de la tuberculosi i el 90 % no coneixia els mecanismes de transmissió de la tuberculosi després de parlar amb el personal sanitari. El retard mitjà dels pacients va ser de 9,6 setmanes amb una desviació estàndard de 4,8; només el 10 % dels pacients entrevistats van acudir al metge durant el primer mes de sentir-se malalt, el 30 % entre els 1-2 mesos, el 20 % entre els 2-3 mesos i el 40 % després de sentir-se malalt durant més de 3 mesos.

La informació epidemiològica obtinguda amb l'estratègia *top-down* i els resultats derivats del treball de camp es van utilitzar per adaptar un model basat en agents (ABM) de propagació de la TB en el context de l'Estat de Gombe. L'ABM resultant es va ajustar amb èxit a l'evolució de la prevalença estimada i els casos diagnosticats del 2007 al 2016. A continuació, es va utilitzar per provar diferents intervencions destinades a augmentar la taxa de notificació, disminuir el retard de diagnòstic i augmentar la consciència de la població respecte a la tuberculosi transmissible. La metodologia utilitzant diferents estratègies utilitzada en aquest projecte s'ha mostrat com una forma robusta d'afrontar el problema real. Ha fet possible (1) proporcionar una imatge global i detallada de la situació de la tuberculosi en un determinat àmbit, (2) relacionar els paràmetres i els resultats del model amb un context socioeconòmic real i (3) generar una eina útil per ajudar en el disseny i l'avaluació d'accions de control de la tuberculosi.

Table of contents

List of figures	xix
List of tables	xxiii
1 Introduction	1
1.1 Tuberculosis	1
1.2 Epidemiology of tuberculosis	5
1.2.1 Basic concepts	5
1.2.2 Tuberculosis in the world	6
1.2.3 Tuberculosis in Nigeria	8
1.3 Global fight against TB	10
1.3.1 Sustainable Development Goals (SDGs)	10
1.3.2 WHO End TB Strategy 2035	13
1.3.3 The National TB and Leprosy Control Programme in Nigeria	15
1.4 Models in epidemiology	16
1.4.1 Compartmental models	18
1.4.2 Agent Based Models (ABMs)	25
1.4.3 Models in tuberculosis	25
1.5 Aim, approach and outline of the thesis	27
1.5.1 Specific objectives	27
1.5.2 Outline	28
2 Understanding tuberculosis dynamics in different contexts by a sequential increase in complexity of SEIR models	31
2.1 Introduction	31
2.2 Models description	32
2.2.1 Models formulation	32
2.2.2 Parameter estimation	44

2.3	Results	44
2.3.1	Setting the best models for HBC and LBC: Susceptible population as a limiting factor	44
2.3.2	Models complexity and parameter derivation, using Argentina as a case study	46
2.3.3	Analysis of the latent infected population, using Niger as a case study	49
2.3.4	Analysis of reinfection using Laos Republic as a case study	52
2.4	Discussion	53
3	Modelling tuberculosis dynamics in Nigeria for the analysis of its country-level situation	55
3.1	Introduction	55
3.1.1	Facts and hypotheses	55
3.2	Model structure	56
3.2.1	Model formulation	58
3.2.2	Model validation and calibration	60
3.3	Results	61
3.3.1	Hypothesis 1 & 2: Case detection and effective treatment	62
3.3.2	Hypothesis 3: Latent infected population	68
3.3.3	Hypothesis 4: Poor notification and mortality in sick population	69
3.4	Discussion	71
4	Analyzing socioeconomic factors that facilitate low notification rate in Nigeria. Case study: Gombe State	75
4.1	Introduction	75
4.2	Material and methods	76
4.2.1	Review of relevant literature and databases	76
4.2.2	Non-documentary data collection	76
4.2.3	Methods and instruments of data analysis	76
4.2.4	Socio-demographic data of Gombe state	77
4.3	Results	77
4.3.1	Results from the analysis of TB data from NTBLCP	77
4.3.2	Analysis of data from field's work	81
4.3.3	Inferential Statistical Analysis of interviews	84
4.3.4	Estimate of some of the most relevant parameters extracted from the interview	88

4.3.5	Significance and important relationships between some of the most relevant parameters extracted from the interviews	94
4.3.6	Multiple regression analysis	98
4.4	Discussion	99
5	Using Agent-based simulations to test interventions in Gombe	101
5.1	Introduction	101
5.2	ODD description of the model	102
5.2.1	Overview	102
5.2.2	Entities, state variables, and scales	102
5.2.3	Process overview and scheduling	103
5.2.4	Design concept	105
5.2.5	Details	107
5.3	Model implementation and results	111
5.3.1	C implementation of the model	111
5.3.2	Model validation	113
5.3.3	Experiments testing different types of intervention in Gombe state .	115
5.4	Discussion	120
6	Conclusions and further work	123
6.1	Specific conclusions	123
6.2	General conclusions	126
6.3	Perspectives and further work	128
	References	129
	Appendix A Mathematical analysis	137
	Appendix B Table of interviews from fieldwork	143

List of figures

1.1	Global TB incidence 2017	2
1.2	TB transmission	3
1.3	<i>Mycobacterium</i> infection circle in Human	4
1.4	Global TB cases 2017	7
1.5	Sustainable Development Goals	12
1.6	Expected Global trend of TB	15
1.7	Modeling diagram	17
1.8	SI model	19
1.9	SI model	19
1.10	SIR model	20
1.11	SI model	21
1.12	SIS model	22
1.13	SI model	23
1.14	SEIR model	23
1.15	SEIR model	24
2.1	Overview of the several proposed compartment models with different levels of complexity	33
2.2	The SEI model	34
2.3	The SEI2 model	36
2.4	The SE8IR model	37
2.5	The SE8IR-reinf model	40
2.6	The SE8I2R model	43
2.7	The effect of susceptible population size on the dynamics of TB demonstrated in both low and high burden countries	45
2.8	Comparison of <i>-low</i> models' fittings for epidemiological data in Argentina .	47
2.9	Analysis of TB dynamics in Argentina with SEI2 model showing the effect of diagnosis time delay	48

2.10	Distribution of exposed population in Argentina according to the SE8IR model	49
2.11	SE8IR model showing the effect of LTBI population to the dynamics of TB in the population of Niger	50
2.12	SE8IR model showing the effect of LTBI immigrant population to the dynamics of TB in the population of Niger	51
2.13	SE8IR-reinf model showing the effect of reinfection to the dynamics of TB in the population of Laos	52
3.1	A compartment model for TB transmission in Nigeria	57
3.2	Model simulation calibrated against tuberculosis prevalence data in Nigeria	62
3.3	Hypothesis 1: Baseline simulation and effect of the notification rate	64
3.4	Hypothesis 1: Evolution of tuberculosis prevalence in Nigeria after 10 years of simulation with different notification rate	65
3.5	Hypothesis 2: Effective treatment in the population with good surveillance and poor surveillance.	66
3.6	Hypothesis 1: Effective strategies in a gradual increase in TB notification in the case of poor tuberculosis surveillance	67
3.7	Hypothesis 3: Changes in latent tuberculosis with respect to an increase in notification rate or in effective treatment	68
3.8	Hypothesis 3: Latent infected population after 10 years of simulation of the model with different notification levels	69
3.9	Hypothesis 4: The figure shows the decrease in mortality due to TB in the sick population for a period of 10 years with different notification levels	70
3.10	Hypothesis 4: The result of 10-year simulation of the model using different notification rates	71
4.1	The map of Gombe state	78
4.2	Gombe state population	79
4.3	Gombe state gender and age group of the population	80
4.4	The mean reported TB case by gender and age group	81
4.5	The mean reported TB incidence by age group	82
4.6	HIV/TB Co-infection	83
4.7	The mean TB incidence/ 10^5	84
4.8	The result from field work	85
4.9	Anova analysis of Gombe state district surface area	86
4.10	Anova analysis of population density	87
4.11	Normal probability distribution test of age of patient interviewed	88

4.12	Number of dependent per patient	89
4.13	Normal distribution test for number of patients	91
4.14	Anova analysis on Patient Delay (PD)	92
4.15	Anova analysis on diagnosis delay at the hospital	93
4.16	One way Anova analysis on diagnosis delay before hospital and surface area	95
4.17	Interaction Plots	96
4.18	Multiple regression tree	98
5.1	State diagram of the model	103
5.2	General flow diagram of the model showing all processes and sub-processes	104
5.3	Details of the flow diagram of the model	105
5.4	Model validation Prevalence rate	113
5.5	Model validation of incidence rate	114
5.6	Model validation diagnosed patient	115
5.7	Test for increase in notification	116
5.8	Test for education and diagnosis delay	118
5.9	Test for 3 ways intervention strategy in TB control	119
B.1	Summary of parameters in SEI8R and SE8I2R models	144
B.2	Summary of parameters in SEI8R and SE8I2R models	145

List of tables

1.1	TB high burden countries as listed by WHO	9
2.1	Summary and description of parameters in SEI and SEI2 models	35
2.2	Summary of parameters in SEI8R and SEI2R models	39
3.1	Summary description of parameters of the model fitted to Nigeria data (2000-2015)	61
3.2	Summary of the mean prevalence of data and model simulations, as well as R^2 of the fitting	65
4.1	Summary result of analysis of TB incidence and surface area	86
4.2	Summary result of analysis of TB incidence and population density	86
4.3	Summary result of Anova analysis of population density	87
4.4	Summary result of analysis of TB incidence and population density	88
4.5	Age distribution Normality test of patients interviewed	89
4.6	Estimated income distribution of patient	90
4.7	Ratio of patients to hospital facilities in the districts of Gombe	91
4.8	Diagnosis delay before going to hospital of patient interviewed in the districts of Gombe, in weeks.	93
4.9	Diagnosis delay at the hospital of patient interviewed in the districts of Gombe, in days.	94
4.10	Diagnosis delay before hospital in weeks of patient interviewed in the districts of Gombe	94
4.11	Pearson correlation coefficient for model variables (P-values are given under the respective correlation values in red)	96
4.12	Test for difference in diagnosis delay at hospital between areas with high reported incidence and areas low reported incidence in Gombe state	97
4.13	Test for difference in diagnosis delay before hospital between areas with high incidence and areas low incidence in Gombe stated	97

5.1	Official data of Gombe state (Nigeria)	108
5.2	Continuation of table 5.1	109
5.3	Initial conditions of state variables	114
5.4	Comparison between interventions	120

Chapter 1

Introduction

1.1 Tuberculosis

Tuberculosis (TB) is an infectious disease caused by *Mycobacterium tuberculosis* (*Mtb*). It usually affects lungs, where it can cause typical TB lesions that can be detected through X-ray during diagnosis, but it can also affect other organs in the body [18]. TB has co-evolved with humans since paleolithic era [21], and it is considered the biggest killer of mankind in history of infectious diseases. It is the responsible for more than 1 billion deaths in the last 200 years [14]. Despite the fact that TB can be cured by a regimen of drugs, it still causes more than 1.5 million death annually, which represents more than HIV and diabetes combined. Its estimated global spread in 2017 is shown in Figure 1.1

The bacillus is transmitted by the inhalation of infected aerosols generated by active TB (ATB) patients [78], as shown in Figure 1.2. The inhalation of the bacilli will usually lead to the trigger of an immune response that can have one of the three different clinical outcomes: (1) complete clearance of the pathogen, (2) latent tuberculosis infection (LTBI) or (3) progression to primary active disease [19, 15].

Infected aerosols must be deposited on the pulmonary alveoli to be able to generate infection, as seen in Figure 1.3. In fact, this is one of keys of the success of *Mycobacterium tuberculosis*: its ability to grow inside the alveolar macrophages (AM) [19].

Some certain protective factors in the lungs can prevent the infective capacity of the *Mtb*. The quality of the aerosol is the first factor, as not all infected patient are able to generate sufficient quantity [18]. The second factor is the quality of the alveolar surfactant, which can also destroy the lipophilic wall of the *Mtb* just as may be destroyed by the AM when it is phagocytised [18]. Nevertheless, it is difficult to evaluate the infectiousness of the aerosol because it is not clear to what extent these factors can affect it. What is clear is that there must be close contact between the patient with active TB and susceptible persons before

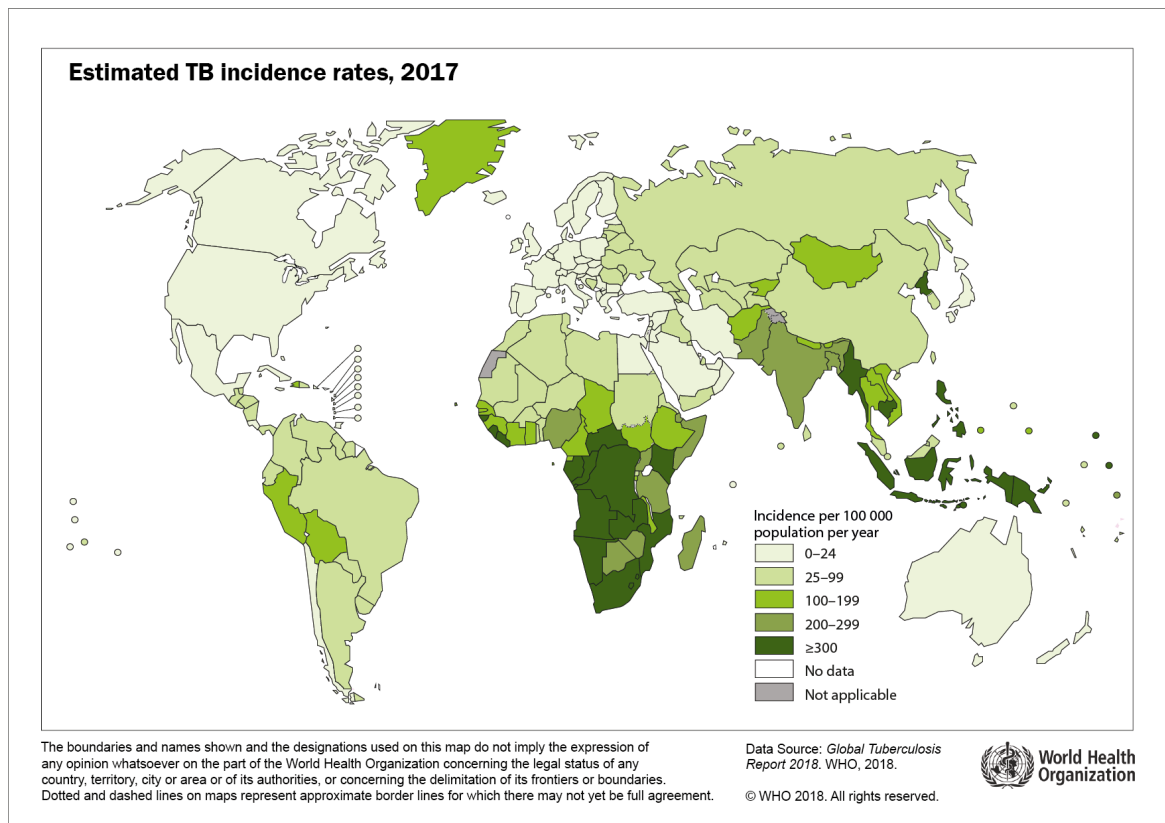


Fig. 1.1 Global view of estimated TB incidence per 100,000 inhabitants, in 2017
[98]

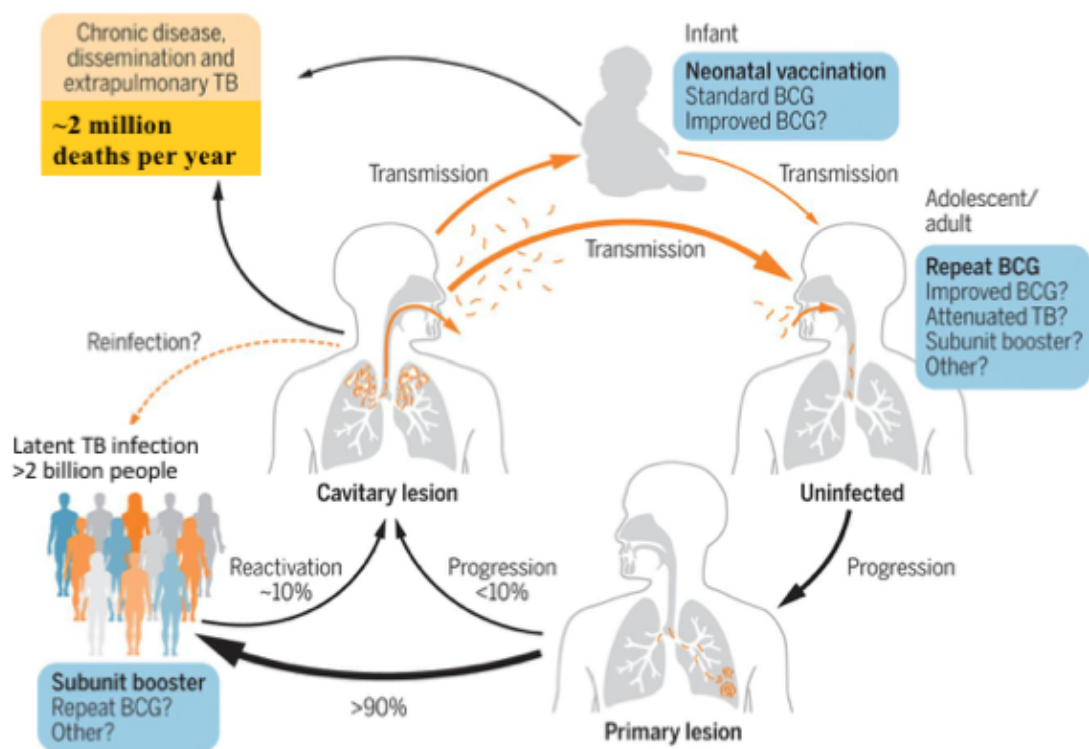


Fig. 1.2 TB transmission circle in the population [78]

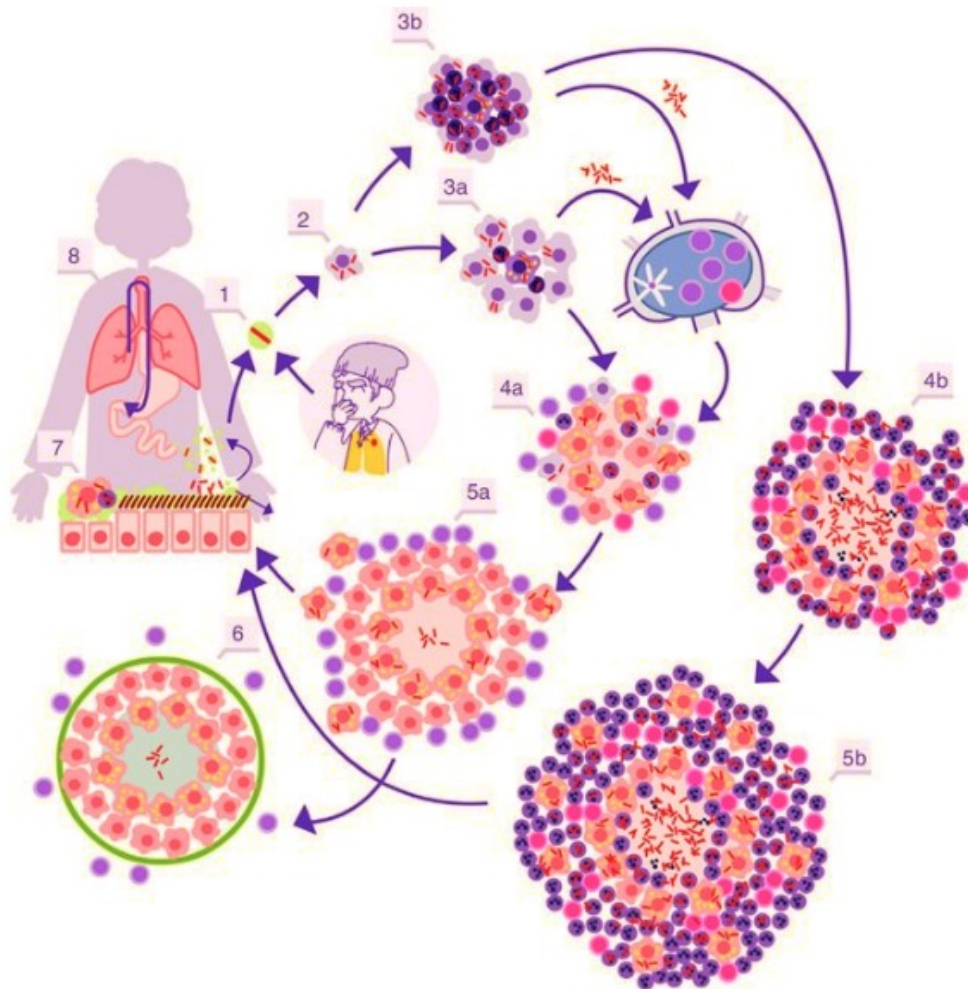


Fig. 1.3 Infectious cycle of *M. tuberculosis*. 1. Bacillus entry into the pulmonary alveolus through a drop of aerosol. 2. Phagocytosis by an Alveolar Macrophage (AM) and subsequent multiplication inside. 3. Destruction of the AM, local dissemination of *M. tuberculosis*, phagocytosis by other AMs and generation of a local inflammatory response dominated by monocytes (3 a) or polymorphonuclear cells (PMN) (3 b), thanks to which the bacilli can be drained into the regional lymph node, where Th1 or Th17 lymphocytes proliferate it. 4. Lymphocytes are attracted by the inflammatory response of the lesions and activate infected MAs or attract more PMN, depending on whether the immune response opts for a Th1 (4 a) or Th17 (4 b) response, respectively. In the first case there is a control of the bacillary population and there is a drainage of numb bacilli through the foamy macrophages (5 a), until it is controlled by encapsulating the lesion (6). In the second, the lesions are growing no more thanks to the entry of PMN and the extracellular bacillary growth in the NET, generating new peripheral lesions. In this case, the bacillary concentration is much higher, and hence the drainage is much more important, either through the alveolar fluid or at the systemic level by granuloma neovascularization (5 b). At the pulmonary level, the alveolar fluid bacilli (7) tend to be drained into the gastrointestinal tract (8), although they can be part of new aerosols, generating new lesions (1) [20].

new infection occurs [18]. It must be taken into account that, ability of each individual to develop ATB depend on many factors, including the bacillary load (which also depend on the intensity of contact with the case), i.e., more than six hours a day for a period which depends on the diagnosis delay time (60 to 90 days in countries with good healthcare system) [18, 84].

LTBI occurs when the host's immune response manages the initial containment of the *Mtb* by developing and encapsulating granulomas as seen in Figure 1.3. More often than not, the bacilli remain physically contained and immunologically constrained by these encapsulated granulomas throughout the lifetime of the hosts [77, 37]. During this process, an endogenous reinfection can produce new infection spotlights that will presumably undergo the same control dynamics [19]. However, even years after infection, especially in the case of immunocompromised hosts (like HIV patients), the control process can fail and the host can develop an ATB. On average, about 10% of the LTBI people develop an ATB during their life [87]. A latent-infected host can be re-infected several times, thereby increasing the load of *Mtb* in its body and hence increasing the chance of progressing to ATB. According to the World Health Organization (WHO), there was an estimated number of 2 to 3 billion people with LTBI in 2015 [70], thus at risk of developing an active disease.

Tuberculosis is still a major global health concern and one of the leading causes of death [16]. As reported by WHO, there were an estimated 10.7 million new ATB cases and 1.7 million TB-related deaths worldwide in 2017 [102]. Even though most TB cases occur in resource-limited countries, it is still a threat to higher-resource countries. This is due to the nature of the disease's strong interaction with HIV dynamics and also the recently world-wide emergence of drug-resistant TB. The main manifestation and the only infectious form of TB is the pulmonary form, hence worthy of study. Tuberculosis has killed more than any other infectious disease in the world, it was estimated that it took more than 1,000,000,000 lives over the cause of 200 years [76]. It is a very silent disease that has eluded mankind for a very long time. It affects people of all social status.

1.2 Epidemiology of tuberculosis

1.2.1 Basic concepts

The word *epidemiology* is an amalgamation of Greek word *epi* meaning on or upon, *demos* meaning people and *logos* meaning study. Put together, epidemiology implies study of what befalls a population of people [49]. Many definitions have been proposed by various scholars, but the following definition capture the general underline purpose.

Epidemiology is the study of the distribution and determinants of health-related states or events in specified populations, and the application of this study to the control of health problems [17]. Epidemiological information is used to plan and evaluate strategies to prevent illness and as a guide to the management of patients in whom disease has already developed. Epidemiology of a disease is an integral part of understanding pattern and behavior of the disease. Like in every discipline, some of the most common variables used in epidemiological studies includes the following.

Prevalence, sometimes referred to as prevalence rate, is the proportion of persons in a population who have a particular disease or attribute at a specified point in time or over a specified period of time [22]. Prevalence differs from incidence in that prevalence includes all cases, both new and preexisting, in the population at the specified time, whereas incidence is limited to new cases only [22]. Sometimes we used what is called a point prevalence.

Point prevalence refers to the prevalence measured at a particular point in time. It is the proportion of persons with a particular disease or attribute on a particular date. Incidence is the most common terminology in defining epidemiology of a disease [22].

Incidence refers to the occurrence of new cases of disease or injury in a population over a specified period of time. Although some epidemiologists use incidence to mean the number of new cases in a community, others use incidence to mean the number of new cases per unit of population. It is usually link to an attack rate [22].

An attack rate is the proportion of the population that develops illness during an outbreak, An alternative and more accurate phrase for attack rate is incidence proportion. to account for number of persons that died from a given out-brake, we normally talk about case fertility rate. A case-fatality rate is the proportion of persons with the disease who die from it [22].

1.2.2 Tuberculosis in the world

Latest on tuberculosis

Worldwide, TB is one of the top 10 causes of death and the leading cause from a single infectious agent (above HIV/AIDS) [102]. Millions of people continue to fall sick with TB each year. In 2017, TB caused an estimated 1.7 million deaths (range, 1.2–1.4 million) among HIV-negative people and there were an additional 300,000 deaths from TB (range, 266,000–335,000) [70] among HIV-positive people. Globally, the best estimate is that 10.0 million people (range, 9.0–11.1 million) developed TB disease in 2017: 5.8 million men, 3.2 million women and 1.0 million children [102]. There were cases in all countries and age groups, but overall 90% were adults (aged ≥ 15 years), 9% were people living with HIV (72% in Africa) and two thirds were in eight countries: India (27%), China (9%), Indonesia

(8%), the Philippines (6%), Pakistan (5%), Nigeria (4%), Bangladesh (4%) and South Africa (3%) [70]. These and 22 other countries in WHO's list of 30 high TB burden countries accounted for 87% of the world's cases. Only 6% of global cases were in the WHO European Region (3%) and WHO Region of the Americas (3%) [68].

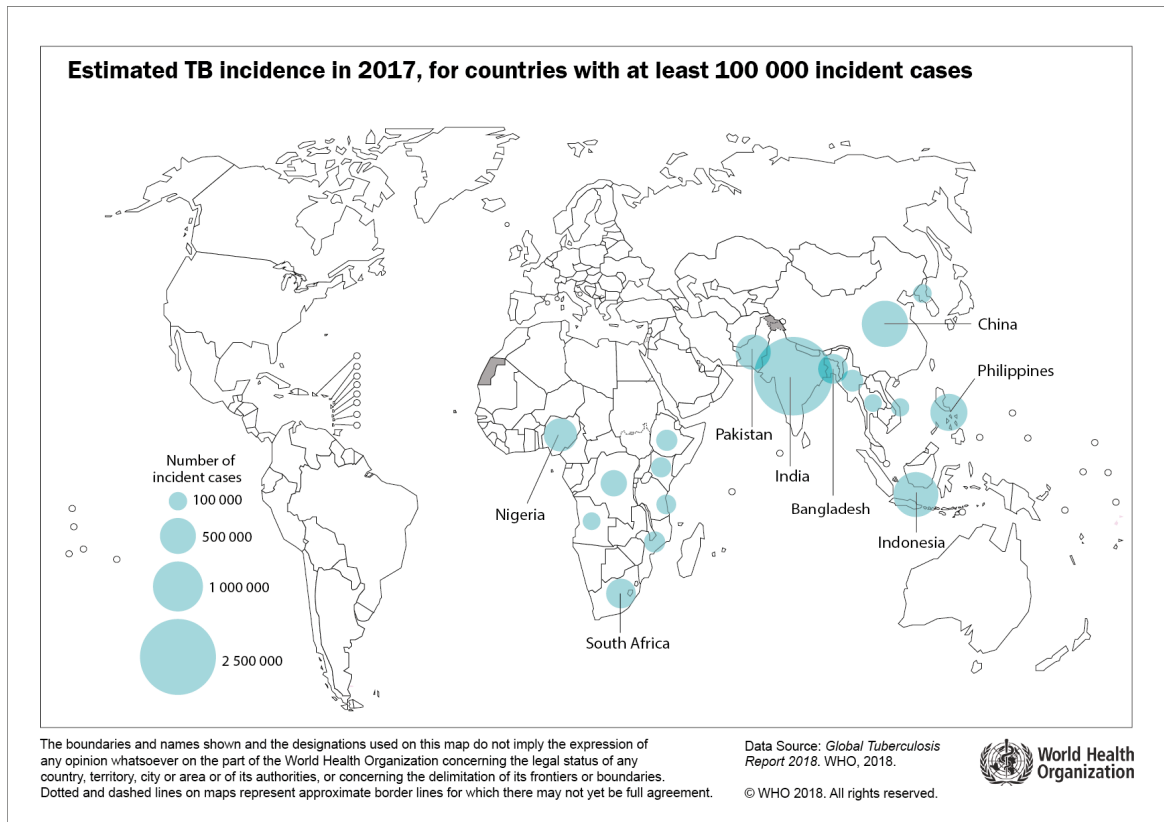


Fig. 1.4 Global view of TB incidence for countries with at least 100,000 cases [98]

The severity of national epidemics varies widely among countries. In 2017, there were fewer than 10 new cases per 100,000 population in most high-income countries, 150–400 in most of the 30 high TB burden countries, and above 500 in a few countries including Mozambique, the Philippines and South Africa [102]. Drug-resistant TB continues to be a public health crisis. The best estimate is that, worldwide in 2017, 558,000 people (range, 483,000–639,000) developed TB that was resistant to rifampicin (RR-TB), the most effective first line drug, and of these, 82% had multidrug-resistant TB (MDR-TB) [68]. Three countries accounted for almost half of the world's cases of MDR/RR-TB: India (24%), China (13%) and the Russian Federation (10%) [68]. Globally, 3.5% of new TB cases and 18% of previously treated cases had MDR/RR-TB. The highest proportions (>50% in previously treated cases) are in countries of the former Soviet Union. Among cases of MDR-TB in 2017,

8.5% (95% confidence interval, 6.2–11%) were estimated to have extensively drug-resistant TB (XDR-TB). About 1.7 billion people, 23% of the world's population, are estimated to have a latent TB infection, and are thus at risk of developing active TB disease during their lifetime [14]. The Sustainable Development Goals (SDG) and End TB Strategy targets set for 2030 cannot be met without intensified research and development [7]. Technological breakthroughs are needed, so that the annual decline in the global TB incidence rate can be accelerated to an average of 17% per year by 2025. Priorities include a vaccine to lower the risk of infection, a vaccine or new drug treatment to cut the risk of TB disease in the 1.7 billion people already latently infected, rapid diagnostics for use at the point of care and simpler, shorter drug regimens for treating TB disease

High burden countries

Here we present 30 countries with their corresponding estimated TB burden per 100000 population in 2017 that are regarded as high TB burden countries [11].

1.2.3 Tuberculosis in Nigeria

Nigeria had an estimated population of more than 200 million in 2019, with approximately more than 54% living in poverty [61]. The country is divided into 36 states plus the Federal Capital Territory (FCT). States are subdivided into 774 local government areas (LGAs) and geographically clustered into six geopolitical zones. The health system is structured along primary, secondary and tertiary levels roughly corresponding to the LGA, state and federal levels. Overcrowding and period in contact with patients are key factors of TB infection. Paradoxically, TB is often seen as a XIX century disease, mainly associated with the poor living conditions in crowded cities. Today, TB is reaching the highest number of cases in absolute numbers in the history [70] precisely because of the massive urbanization of populations. This fact is especially painful in Nigeria, a country that is undergoing rapid urbanization with a rapidly growing population. At the current growth rate of about 2.8% to 3.5% a year, it is estimated that Nigeria's urban population will double in the next two decades [8]. As a consequence, TB is endemic in Nigeria.

Despite current global control efforts to reduce TB, Nigeria's incidence is refusing to show any significant decline [70]. On March 14th, 2017, a wide circulated paper announced that Nigeria is rank 4th in TB infection worldwide [80]. This was followed by the press released by the national TB programme coordinator. The statistics show that over 80% of TB cases in Nigeria are still undetected while the disease claims millions of lives over the years in the country [5, 70].

Table 1.1 TB high burden countries as listed by WHO, together with absolute ($\times 10^3$) and relative (per 10^5 inhabitants) incidences in 2018 [99]. In bold, absolute incidences above 400,000 cases; in italic, relative incidences greater than $500/10^5$ inhabitants.

Countries	Absolute incidence ($\times 10^3$)	Relative incidence / 10^5	Countries	Absolute incidence ($\times 10^3$)	Relative incidence / 10^5
Angola	109	355	Russian Federation	79	64
Bangladesh	357	221	<i>South Africa</i>	301	<i>520</i>
Brazil	95	45	Thailand	106	153
China	866	61	Tanzania	142	253
<i>Democratic People Rep. of Korea</i>	131	<i>513</i>	Viet Nam	174	182
Democratic Rep. of Congo	270	321	Cambodia	49	302
Ethiopia	165	151	<i>Central Africa Rep.</i>	25	<i>540</i>
India	2,690	199	Congo	20	375
Indonesia	845	316	<i>Lesotho</i>	13	<i>611</i>
Kenya	150	292	Liberia	15	308
<i>Mozambique</i>	162	<i>551</i>	<i>Namibia</i>	13	<i>524</i>
Myanmar	181	338	Papua New Guinea	37	432
Nigeria	429	219	Sierra Leone	23	298
Pakistan	562	265	Zambia	60	346
<i>Philippines</i>	591	<i>554</i>	Zimbabwe	30	210

It is reasonable to say that, in spite of effort being made by all stakeholders battling TB in Nigeria, the epidemic is still very persistent. The current estimate of TB prevalence in Nigeria by WHO is around 330/100,000 population, which is approximately the same as the estimated TB prevalence in 1990 (323/100,000). The same situation is observed in the number of deaths due to the disease (70/100,000 to 65/100,000). The reason for this steady state of the TB epidemic in the country is still unclear. Many associate it with poverty and the consistent growth in population due to a very high fertility rate of the women in the country (the population grew from 95,62 million people in 1990 to 206,20 in 2018), others presume that there is a need for a better strategy in order to fight the disease towards a significant decline in the near future [27].

Nigeria is a setting where several health-care options (medical pluralism) including orthodox medicine (public, private, or drugstores), traditional medicine, spiritual healers etc., operate freely [27]; the public health facilities within which the TB control programme

operates is "distanced" from the people and are often not the first choice during health seeking decisions. In fact, high TB prevalence in Nigeria has a lot to do with policies of the TB control program. Several studies have shown that TB prevalence has a lot to do with policies made in TB control [45, 50, 55, 28]. Nigeria adopted the passive case finding approach as recommended by WHO, based on a study conducted in India [57, 13]. successful TB control only happens when social and cultural factors are taken into consideration [83]. TB is a major public health problem in Nigeria and it was declared a national emergency in 2006 [73]. Following the Abuja Declaration in 2001, Directly Observed Therapy Short course (DOTS) activities have been scaled up across Nigeria. In spite of the documented effectiveness of DOTS in the Nigerian context Nigeria has one of largest burden of TB cases in the world [44]. Despite expressions of political will to control TB and a clearly articulated national TB policy, neither the set target for CDR nor cure rates has been achieved nationally. TB continues to account for high rates of morbidity and mortality within Nigeria in spite of reports of 99% geographic coverage using the DOTS strategy by 2008 [73, 66]. This situation has been further compounded by the high HIV prevalence of 4.1% in Nigeria. Nigeria accounts for about 10% of the total global burden of HIV with estimates of 3.4 million people living with HIV by the end of 2011 [73, 64].

1.3 Global fight against TB

1.3.1 Sustainable Development Goals (SDGs)

The Sustainable Development Goals (SDGs) were born at the United Nations Conference on Sustainable Development in Rio de Janeiro in 2012. In order to meet up with the urgent universal needs that include both political, economic and environmental, some universal goals were introduced during the conference. The Millennium Development Goals (MDGs), which started a global effort in 2000 to tackle the indignity of poverty, were replaced by the SDGs that year.

The MDGs lasted for 15 years where it drove progress in several important areas, such as reducing extreme poverty, access to water and sanitation, reduction in child mortality and maternal health among others. Some of the global movement that were kicked started by the MDGs includes but are not limited to free primary education, inspired countries to invest in the future generations, very huge stride in combating HIV/AIDS and other treatable diseases such as malaria and tuberculosis.

The SDGs were born out of commitment to finish what was stated and to tackle even more pressing challenges facing our world today. A total of 17 interconnected goals were set

1.5. The goals were interconnected in such a way that success in one will affect success in the rest. Annihilating climate change threat will impact how to manage our natural resources extensively, achieving gender equality, better health, eradicating poverty will foster and facilitate peace and inclusive society, reduce inequality and integrated economies. The SDGs could be seen as the greatest chance to improve the future generation. The uniqueness of the SDGs includes the reaffirmation of the international commitment to end poverty, permanently, everywhere, making sure no one is left behind. Most important, involved in making or building a more sustainable, safer, more prosperous planet for all humanity.

The SDGs came into effect in January 2016, and they will continue to guide United Nations Development Programme (UNDP) policy and funding until 2030. As the lead United Nations (UN) development agency, UNDP is uniquely placed to help implement the Goals through their work in some 170 countries and territories.

Each of the aforementioned goals carries sub goals and sub categories. In particular, the goals within the 3rd objective (*Ensure healthy lives and promote well-being for all at all ages*) is a mandate that WHO describe as:

1. By 2030, reduce the global maternal mortality ratio to less than 70 per 100,000 live births.
2. By 2030, end preventable deaths of newborns and children under 5 years of age, with all countries aiming to reduce neonatal mortality to as low as 12 per 1000 live births and under-5 mortality to at least as low as 25 per 1,000 live births.
3. By 2030, end the epidemics of AIDS, tuberculosis, malaria and neglected tropical diseases and combat hepatitis, water-borne diseases and other communicable diseases.
4. By 2030, reduce by one third premature mortality from non-communicable diseases through prevention, treatment and promote mental health and well-being.
5. Strengthen the prevention and treatment of substance abuse, including narcotic drug abuse and harmful use of alcohol.
6. By 2020, halve the number of global deaths and injuries from road traffic accidents.
7. By 2030, ensure universal access to sexual and reproductive health-care services, including for family planning, information and education, and the integration of reproductive health into national strategies and programmes



Fig. 1.5 Sustainable Development Goals as they came into effect in January of 2016
[98]

8. Achieve universal health coverage, including financial risk protection, access to quality essential health-care services and access to safe, effective, quality and affordable essential medicines and vaccines for all.
9. By 2030, substantially reduce the number of deaths and illnesses from hazardous chemicals and air, water and soil pollution and contamination.

Our vested interest in this research work is related to 3.3, which is related to ending TB epidemics.

1.3.2 WHO End TB Strategy 2035

The movement of the United Nations from the 2015 Millennium Development Goals to the sustainable Development Goals puts WHO at a point of inflexion.

To complement the transition from the MGDs to SDGs, the world community has launched a dramatic accelerated fight against TB and thus most affected by it: the poorest, most vulnerable, socially marginalized and inequitably reserved. TB has been a threat to health security, a challenge hitch to the human development, a whip on the human population. The World Health Organization's new and holistic strategy approved by the World Health Assembly of 194 Member States in 2014 places patients and communities at the heart of the response. Some of the progress made include but not limited to

1. 43 million lives saved between 2000 and 2014 through effective TB diagnosis and treatment
2. 47% decline in TB mortality rate and 42% decline in TB prevalence rate since 1990
3. HIV-related TB deaths down by 32% in the last decade
4. Fragile progress in MDR-TB Treatment for MDRTB has increased with almost all cases detected in 2014 started treatment

And some of the major challenges includes but not limited to

1. US\$ 1.4 billion funding gap per year for implementation of existing TB interventions. An additional gap of US\$ 1.3 billion exists for research
2. 3.6 million people with TB are missed by health systems every year and therefore may not get adequate care they need
3. TB/HIV response needs acceleration Anti-retroviral treatment, treatment of latent TB infection and other key interventions still need further scale-up

4. MDR-TB remains a public health crisis Only one in four MDR-TB cases detected and one in two cases cured

End TB Vision

The vision to end TB aim at everyone with TB should have access to the innovative tools and services they need for rapid diagnosis, treatment and care. This is a matter of social justice, fundamental to the goal of universal health coverage. Considering the growing prevalence of drug-resistant tuberculosis, ensuring high quality and complete care will also benefit global health security.

End TB Strategy and Targets

The End TB strategy aims to end the global TB epidemics, with targets to reduce TB deaths by 95% and to cut new cases by 90% between 2015 and 2035 (shown in Fig 1.1), and to ensure that no family is burdened with catastrophic expenses due to TB. It sets interim milestones for 2020, 2025, and 2030. There are 3 main pillars for the End TB Strategy, namely

1. Integrated, patient centred care and prevention
2. Bold policies and supportive systems
3. Intensified research and innovation

The targets for the set strategy are

1. Reduction in number of TB deaths by 35% in 2020, 75% in 2025, 90% in 2030, 95% in 2035 compared with 2015
2. Reduction in TB incidence rate by 20% in 2020, 50% in 2025, 80% in 2030, 90% in 2035 compared with 2015
3. TB family facing catastrophic cost incidence rate by 0% in 2020, 0% in 2025, 0% in 2030, 0% in 2035 compared with 2015.

The May 2014 resolution calls on governments to adapt and implement strategy with high-level commitment and financing. It reinforces a focus within the strategy on serving populations highly vulnerable to infection and poor health care access, such as migrants. It also highlight the need to engage partners within the health sector and beyond, such as in the fields of social protection, labour, immigration and justice. The WHO was requested

by the resolution Secretariat to help Member States adapt and operationalize the strategy, monitor implementation and evaluate milestone progress towards the 2035 targets, noting the importance of tackling the problem of multidrug-resistant TB and promoting collaboration across international borders.

In order to reinforce Pillar 3, the WHO in 2018 publish a Guidelines for country-level TB modelling. The document aims to provide concrete, pragmatic guidance for how TB modelling and related technical assistance is undertaken to support country decision-making of which we based our research on [69].

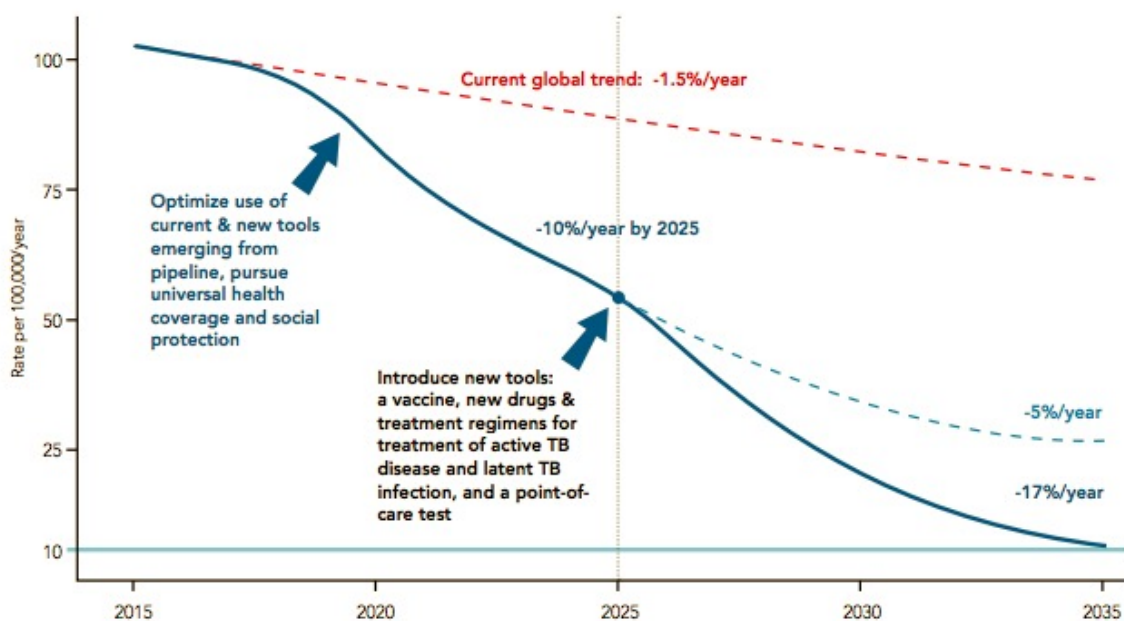


Fig. 1.6 Desired decline in global TB incidence rates to reach the 2035 targets [69]

1.3.3 The National TB and Leprosy Control Programme in Nigeria

The control of the disease in Nigeria is coordinated by the *National Tuberculosis and Leprosy Control Program (NTBLCP)*, in line with the *End TB Partnership* initiatives whose ultimate target is to eliminate TB as a public health problem by the year 2050 (meaning reaching less than 1 case per million population) [27].

NTBLCP of Nigeria was officially launched in 1991 with a mandate to coordinate TB and Leprosy Control activities in the Nigeria and reduce the burden of both diseases. The NTBLCP is structured along the three tiers of government *i.e.* Federal, State and Local Government Areas. Its activities are supported by development partners. Prominent among these are WHO, Global Fund to fight AIDS, TB and Malaria (GFATM), International

Union Against Tuberculosis and Lung Diseases (IUATLD) and United States Agency for International Development (USAID). There is a National TB training centre in Zaria which is responsible for development of human resource, training guidelines and operational research relating to TB [73]. The basic strategy of NTBLCP remains the provision of free DOTS to all persons with ATB. NTBLCP efforts have yielded results as to contained TB burden by doubling of smear positive case detection rate (CDR) from 16.3% to 31.1% over the years [73]. However, these indicators of NTBLCP good performance appear to be overshadowed by the exponential increase in total number of new cases of TB in the country.

1.4 Models in epidemiology

Even with long history of epidemiology and all, the study of diseases using mathematical techniques begins about 350 years ago. The work of John Graunt (1620-1674), is somewhat unanimously regarded as the first statistical study of infectious disease. His book of 1663 of Natural and political observations describes the Bill of Mortality which was directly related with methods of public health statistics. After a century or so, a mathematician called Daniel Bernoulli used method of mathematical analysis to analyze mortality from smallpox. The work was published in 1766 and was considered the first epidemiological model [49]. The author argued that inoculation with live virus obtained from a mild case of smallpox would reduce the death rate and thereby increase the population, even if the inoculation itself might occasionally be fatal. A contemporary reformulation of Bernoulli's approach in terms of differential equations is given in [49]. Between 1873 to 1894, P.D. En'ko developed a transmission model based on some specific assumptions without resorting to previously developed theory. He compared synthetic epidemics, the number of which are calculated using the model actual outbreaks of measles and scarlet fever in two boarding school in St. Petersburg [9]. In 1902, Ronald Ross received the Nobel Prize in Physiology or Medicine for his work on the life cycle of the malaria parasite, he explained the complete life cycle in human with the inclusion of mosquito as a vector and the Plasmodium parasite. The work was extended by Macdonald, the formulated mosquito-born transmission is sometimes referred to as the Ross-Macdonald model [30].

Threshold theorem was developed by Kermack and McKendrick between 1927 and 1939. It states that introduction of infectious individual in a susceptible population could cause an epidemic only if the density of the susceptible exceeds a certain critical value or threshold.

The description of a system using mathematical tools, techniques or language is termed as mathematical models. The process of developing the model is termed as modelling. Generally, modelling approaches are considered of two types, the top-down approach and

the bottom-up approach. An example of the top-down approach is a compartmental model, while Agent Based Models are good example of the bottom-up approach.

Models could be classified in multiple ways: linear or nonlinear, static or dynamic, discrete or continuous, deterministic or stochastic, among others. Furthermore, a wide variety of modelling methodologies exist, ranging from decision analytic models to compartment models, network models, Agent-Based models, and many other types and hybrids along the spectrum of possibilities. Nevertheless, the principle of parsimony specifies that a model should not be more complex than needed [35].

Modelling process includes several steps and phases. Figure 1.7 shows a possible modelling process diagram including one of the most critical point, the formulation of the question to be answered. This will determine the kind of model and approach to be used, as well as the final modelling cycle that will follow. In general, models may be used for (1) understanding purposes, (2) predictive purposes and/or (3) controlling purposes.

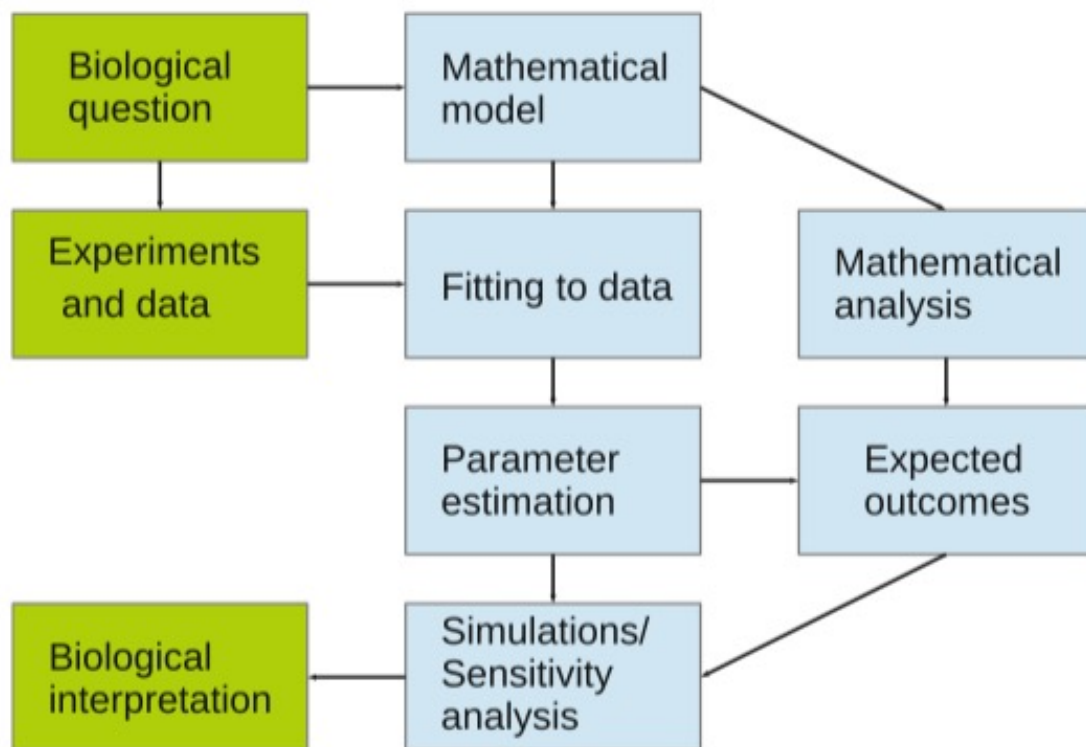


Fig. 1.7 Modeling diagram
[49]

Usually, modeling process (described in Fig. 1.7) involves the use of schematic diagram, which requires translation of a biological scenario into a mathematical problem. The modeling process typically begins with a clear description of the process based on the scientific

understanding of the system. The translation into mathematical equations should be made with the specific goal or biological question in mind. Then, the verbal description of the system is encoded in mathematical equations. The model should incorporate only those features that are relevant to the specific goal or biological question in mind. Once the model is formulated, it can be investigated with a number of mathematical tools. Finally, the results must be interpreted in the light of the biological scenario considered and potentially seek the answer of the biological question that was set forth at the beginning. At the very least we must address these questions: What did we learn about the real world from the model? Is our model's message supported by the information about the system? [49]. Our interest in this research lies with the compartmental models and the Agent Based models. Thus, we briefly describe each of the aforementioned models.

1.4.1 Compartmental models

In compartmental models, individuals in the population are divided into groups (or compartments) and the model tracks the infection process for these individuals collectively. This kind of model can be deterministic or stochastic. The compartmental models are used to study different scenarios in disease transmission in order to understand its dynamics and assess the effects of control strategies. We begin by introducing some of the most classical compartmental models used in epidemiology and which system of ordinary differential equation is used to describe the dynamics in each compartment. This type of model is inspired by the work of Kermack and McKendrick [39], where the author proposed for the first time in 1927 the simplest compartmental models involving susceptible population, infectious and the recovered individuals (SIR models).

SI model

SI model is considered the simplest of all the compartmental models, where S represents the susceptible sub-population and I stands for the infectious collective. The schematic diagram of the model is given in Figure 1.8.

This model is described by equations 1.1 to 1.2.

$$\frac{dS}{dt} = -\beta SI, \quad (1.1)$$

$$\frac{dI}{dt} = \beta SI \quad (1.2)$$

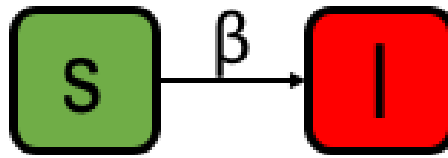


Fig. 1.8 SI model flow diagram (S: susceptible; I: infectious; β : infection rate)

where βI is referred to as force of infection or the incidence rate, i.e., the rate at which the susceptible population becomes infected per unit of time.

This model is a simple mathematical model that can describe the behavior of some epidemics efficiently. Figure 1.9 shows the graphical solution of the SI model, where all the susceptible population becomes sick after a period of time t .

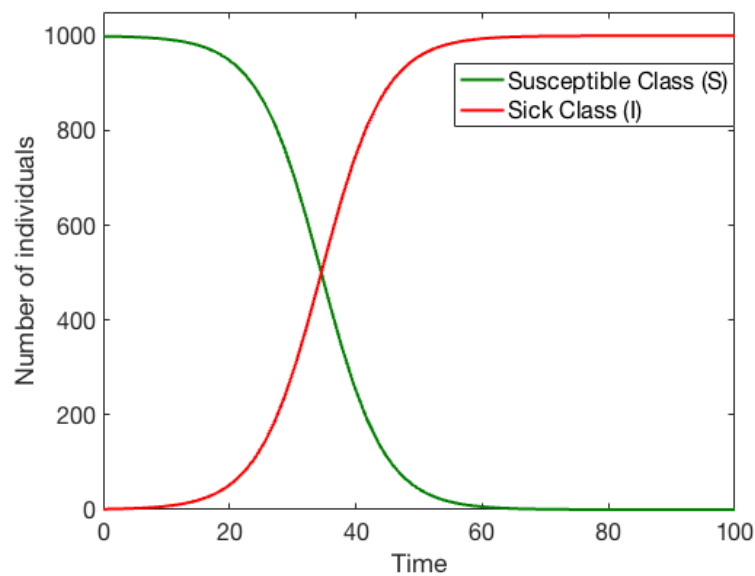


Fig. 1.9 SI model dynamics (S: susceptible; I: infectious)

SIR model

One of the most extended ways of introducing an epidemic model is through the work of Kermack and McKendrick, which is one of the first epidemiological models, introduced in 1927 [38].

In this classic way of modeling, the spread of the disease in a population is considered by splitting the population into 3 non-intersecting compartments, namely the healthy but

are vulnerable to contract the disease which are called *Susceptible* and denoted by S , the ones that have already contracted the disease and are infectious sick denoted by I , and finally the ones that have recovered and can not contract the disease again called the *Recovered or Removed*, denoted by R .

The number of individuals in each of these classes changes with time, that is, $S(t)$, $I(t)$, and $R(t)$ are functions of time t . The total population size N is the sum of the sizes of these three classes:

$$N(t) = S(t) + I(t) + R(t) \quad [9] \quad (1.3)$$

In order to simplify reality, the model considers 2 assumptions: (1) the sick individuals are infectious, and (2) the total population size is constant. Equations 1.4 to 1.6 describe the SIR model schematic diagram in Figure 1.10.

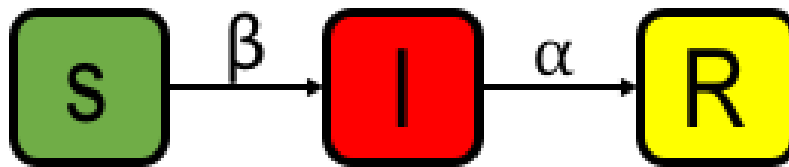


Fig. 1.10 SIR model diagram (S: susceptible; I: infectious sick; R: recovered or removed)

$$\frac{dS}{dt} = -\beta SI, \quad (1.4)$$

$$\frac{dI}{dt} = \beta SI - \alpha I, \quad (1.5)$$

$$\frac{dR}{dt} = \alpha I \quad (1.6)$$

Equations 1.4 to 1.6 were derived by considering how the classes change over time. As susceptible individual enters into contact with an infectious individual, that individual could become infected with a certain probability and migrate to the sick class at per capita rate β . The susceptible decrease in a unit of time by all individual that becomes sick at that time. Individuals leaving the sick compartment by recovering goes to the removed compartment at constant per capita probability per unit of time α , called the recovery rate.

Figure 1.11 shows a simple solution of the SIR model. If we let $\lambda(t) = \beta I$, then the number of individuals who become infected per unit of time is equal to $\lambda(t)$. The function

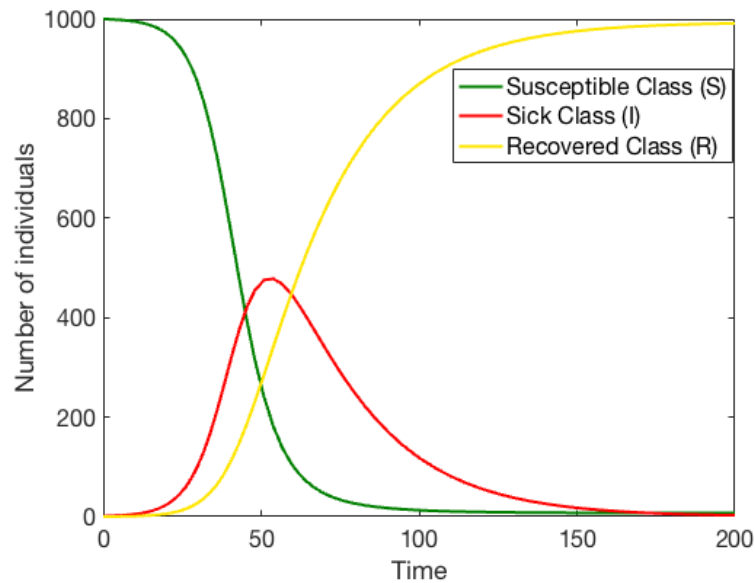


Fig. 1.11 SIR model dynamics (S: susceptible; I: infectious sick; R: recovered or removed)

$\lambda(t)$ is called the force of infection. The coefficient β is the constant of proportionality called the transmission rate constant. The number of infectious individuals in the population, $I(t)$, is called the prevalence of the disease. For the model to be well defined, we equipped equation 1.4 to 1.6 with $S(0)$, $I(0)$, and $R(0)$ as initial conditions, also as a word of cautious, we need to be careful with the units of the quantities involved when forming a model. Independent of the initial conditions.

$$\lim S(t) = S_{\infty} \quad (1.7)$$

This is because $S(t)$ is monotone non increasing and bounded from below by 0. The recovered individual has a monotone non-decreasing behavior and bounded from above by $N(t)$. Therefore:

$$\lim R(t) = R_{\infty} \quad (1.8)$$

On the other hand, the sick population may have monotone non increasing behavior, or may increase to some threshold and then decrease to zero as in Figure 1.11. Therefore, from equation 1.5, a necessary and sufficient condition for an initial increase in the number of sick is $\frac{\beta S(0)}{\alpha} > 1$. The limit S_{∞} and R_{∞} of $S(t)$ and $R(t)$ can easily be evaluated and the quantity S_{∞} is called the final size of the epidemics. For the epidemics to die out, if

$$\lim I(t) = I_\infty \quad (1.9)$$

Then I_∞ must be zero. From equation 1.4.

$$\int_0^\infty S'(t) dt = -\beta \int_0^\infty S(t)I(t)dt \quad (1.10)$$

$$S_0 - S_\infty = \beta \int_0^\infty S(t)I(t)dt \quad (1.11)$$

$$S_0 - S_\infty \geq \beta S_\infty \int_0^\infty I(t)dt \quad (1.12)$$

Equation 1.12 shows that $I(t)$ is integrable on $[0, \infty)$ and hence the $\lim I(t) = I_\infty = 0$

Another important property of epidemics that can be estimated from the SIR model is the maximum number of sick reached in the epidemic. This signifies the severity of the epidemic, estimating it will allow us to know when a number of infections in a population will begin to decline given a newly occurring infectious disease.

SIS model

The assumption for permanent immunity after recovery is not realistic for some types of disease such as influenza. If we assume a simple scenario where recovered individual becomes susceptible immediately, then we come up with another simple model called the SIS model as shown in Figure 1.12. The model is described by equation 1.13 to 1.14 and Figure 1.13 gives a typical graphical solution of the model.

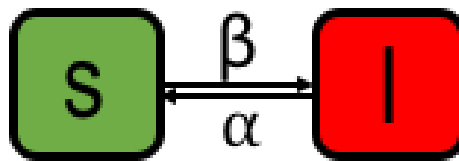


Fig. 1.12 SIS model diagram (S: susceptible; I: infectious)

$$\frac{dS}{dt} = -\beta SI + \alpha I, \quad (1.13)$$

$$\frac{dI}{dt} = \beta SI - \alpha I \quad (1.14)$$

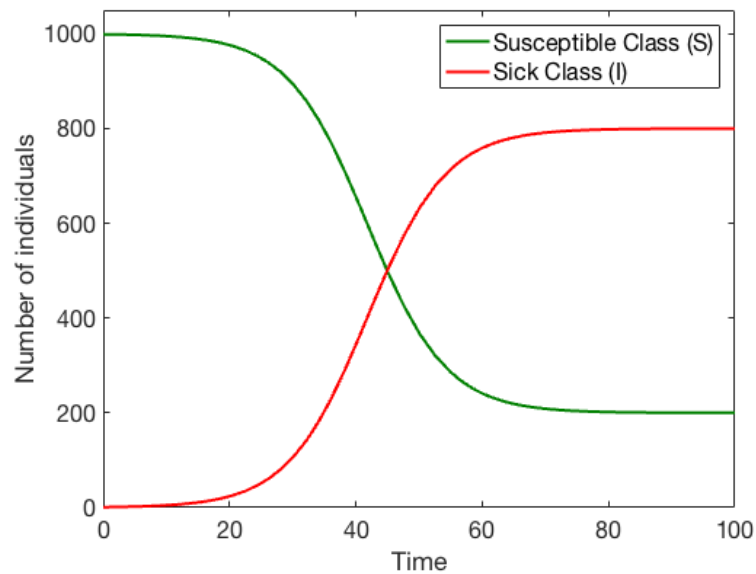


Fig. 1.13 SI model dynamics (S: susceptible; I: infectious)

SEIR model

Other types of diseases like tuberculosis require a latent period of incubation, thus the SIR model was extended by including an additional class called the exposed (E) class. Individuals that are infected but not yet infectious are assigned to this class. We take $\frac{1}{\alpha}$ to be the mean time spent in the exposed compartment. The flow chart of the model is given in Figure 1.14.

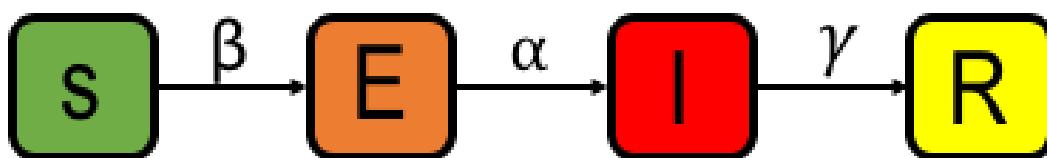


Fig. 1.14 SEIR model diagram (S: susceptible; E: exposed; I: infectious; R: recovered or removed)

Transmission events happen via contact of S and E with constant rate (β). The SEIR model can be expressed mathematically through equation 1.15 to 1.18

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

$$\frac{dS}{dt} = -\beta S \frac{1}{N}, \quad (1.15)$$

$$\frac{dE}{dt} = \beta S \frac{1}{N} + \varepsilon E, \quad (1.16)$$

$$\frac{dI}{dt} = \varepsilon E - \gamma I \quad (1.17)$$

$$\frac{dR}{dt} = \gamma I. \quad (1.18)$$

Here N also describes the total host population size and R_0 which is a threshold frequency for an epidemic is given by:

$$R_0 = \frac{\beta}{\gamma} \quad (1.19)$$

A typical solution of the model is given by 1.15.

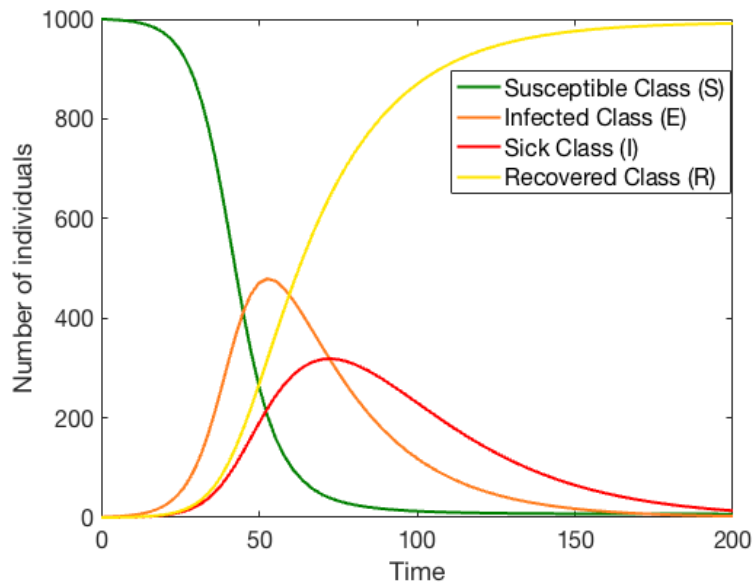


Fig. 1.15 SEIR model dynamics (S: susceptible; E: exposed; I: infectious; R: recovered or removed)

In some special cases, when the infectious cycle requires it, exposed class can include infectious individuals without clinical symptoms. Then, the infection equations must be modified to account for this possibility. In fact, the models described in this Section are the simplest ones, and are usually taken as the starting point to develop more complex models that account for the specificity of the studied disease.

1.4.2 Agent Based Models (ABMs)

Historically, the complexity of scientific models was often limited by mathematical tractability: when differential calculus was the only approach we had for modeling, we had to keep models simple enough to “solve” mathematically and so, unfortunately, we were often limited to modeling quite simple problems. With current computer capabilities, the limitation of mathematical tractability is removed, so we can start addressing problems that require models that are less simplified and include more characteristics and heterogeneity of the real systems. Agent Based Models (ABMs) are less simplified in one specific and important way: they represent a system’s individual components and their behaviors. Instead of describing a system only with variables representing the state of the whole system, we model its individual agents. ABMs are thus models where individuals or agents are described as unique and autonomous entities that usually interact with each other and their environment locally. A unique behavior of each agent is usually described by a simple set rules, and the interaction between the respective agents usually influence their behaviours. Modelling the agent individually allows for the observation of the full effect of system as a whole through the diversity that existed between the respective agent [48].

Agents can be organisms, humans, businesses, institutions, and any other entity that pursues a certain goal. Being unique implies that agents are usually different one from each other in such characteristics as size, location or resource reserves and history, among others. Interacting locally means that agents usually do not interact with all other agents but only with their neighbors-in geographic space or in some other kind of “space” such as a network. Being autonomous implies that agents act independently of each other and pursue their own objectives. They adjust their behavior to the current states of themselves, of other agents, and of their environment. Using ABMs allows us address problems that concern emergence: system dynamics that arise from how the system’s individual components interact with and respond to each other and their environment. Hence, with ABMs we can study questions of how a system’s behavior arises from, and is linked to, the characteristics and behaviors of its individual components [81].

1.4.3 Models in tuberculosis

Several mathematical models have been used to estimate long-term dynamics of TB to help to assess the development of strategies of TB control. The literature on compartment models to describe complex systems is extensive [40, 24, 11]. Most of the models found in literature are of SEIR type. Many authors use rigorous mathematical methods to show relationship and

important properties from the model [74], while few uses data from agencies such as WHO, local and national TB agencies to validate their studies.

Many different studies have been able to draw important conclusions on TB dynamics by means of models and other methods [60, 56, 89, 58, 75, 31, 51, 41, 10, 46, 47]. Okuonghae and Ikhimwin [62], for example, developed a model which classified the population by their TB awareness level, a key factor which could affect the case detection rate. The model highlighted the importance of the backward bifurcation phenomena when latently infected individuals are exogenously re-infected. Another factor which could contribute to better adjust TB models is the HIV dynamics, especially in the sub-Saharan African region [42]. HIV patients have a higher risk to become infected and also to progress to active disease once infected than non-HIV infected people. According to WHO [71], nearly all HIV-positive people with active TB will die.

By means of mathematical models, Wallis [91] was able to identify individuals with an innate resistance to *Mtb*. He concluded that understanding the mechanisms of resistance may lead to therapeutic strategies to counter immune evasion by *Mtb*. Moreover, models were also used by Wallis [91] to assess the affinity of LTBI reactivation when patients are administered drugs with TNF blockers. He was able to determine the specific TNF drugs that are more efficient at reactivating the LTBI and thus predict the risk of relapse in persons undergoing treatment. Similarly, another research project by Moualeu-Ngangue et al. [53] worked with a model simulating the global TB dynamics. The study showed that TB spreading crucially depends on the basic reproduction number (number of new cases one case generates during its sick period, which is estimated to be between 10-15 [88, 91]). The research was done in Cameroon where the role of TB diagnosis, treatment, TB awareness level and traditional medicine on the dynamics of TB were assessed. Song et al. [85] investigated the epidemiological time scales of TB and they evaluated the risk of infection from both close contacts (clusters) and casual contacts (random). They concluded that the risk of infection depends on the source of infection as well as on different environmental characteristics.

Lastly, Guzzeta *et al* [34] presented three different ways to model TB dynamics: (1) an ODE model with no age structure and constant population size, (2) an age-structured, stochastic version of the ODE model and (3) a socio-demographic Agent-Based Model (ABM). The models were fitted to epidemiological data from Arkansas, USA. The author compared the three different models and concluded that different modeling techniques have their advantages and drawbacks and they should be chosen carefully. For example, an ODE model is best suited to describe the evolution of prevalence, incidence, and mortality. On the other hand, an ABM, which gives insight on the individual dynamics, would be best to estimate the fraction of reactivated cases or to fit age-specific incidence of active TB.

Although there are many studies into TB epidemic, cause, spread and suggestive measures for therapy and control, none of the aforementioned studies uses mathematical model to evaluate the actual epidemic of TB in Nigeria using data obtained from the local TB control program in the community. This thesis proposes a mathematical model that will be used to evaluate TB burden in this country. We aim to point out effective strategies that could be used to effectively reduce TB burden and death due to TB in Nigeria.

1.5 Aim, approach and outline of the thesis

The aim of this thesis is to increase our understanding of TB epidemiology in high burden contexts and propose improvements on its control strategies, with the final objective of contribution in reducing TB epidemic in these settings.

In order to make progress to this end, we propose the combination of different approaches that provide a global perspective of TB in these countries. We try to take into consideration the main biological factors that govern TB spread together with social, economical, cultural and political determinants. Although biological factors can be more or less common everywhere, the other factors may vary greatly from one country to another and even in some countries, from region to region. Therefore, we propose to progressively get closer to a real territory to finally analyse social patterns that can affect the spread of the disease, cultural behaviours that may facilitate the misdiagnosis or the political regulations that determine the health facilities structures, operations and limitations, among others. In that sense, this thesis will start by studying different country-level dynamics and will continue by analysing the situation of Nigeria as a whole to finally focus on Gombe state. The methodological approaches to be used in this process will also vary and be adapted to the scales and factors consecutively tackled. The final picture will arise from a combination of compartmental top-down models, statistical models and agent-based models together with field work carried out in a collaboration with the NTBLCP Gombe state.

1.5.1 Specific objectives

The specific objectives to reach the main goal are detailed below. They have been grouped into research, methodological and public health objectives.

Research objectives

- R1 To understand the main features of TB dynamics in low burden countries and high burden countries.

- R2 To determine the cause of high incidence and its persistence in Nigeria.
- R3 To analyse the behavioural and socio-economic factors resulted to TB persistence in Nigeria using Gombe as a case study.
- R4 To propose interventions to improve TB control in Gombe (Nigeria).

Methodological objectives

- M1 To develop several compartmental models with increasing complexity adapted to study different context and address specific questions.
- M2 To adapt a general Agent-Based Model for the study of the specific situation in Gombe

Public health objective

- PH To look for possible in near future, real and feasible interventions to improve TB control in Gombe (Nigeria)

1.5.2 Outline

The thesis consists of the following chapters. In chapter 1 we have briefly introduced the main epidemiological features of tuberculosis disease and its present status around the world. We also have shown the importance of mathematical models in epidemiology by introducing a number of research work around this field for a very long period of time. In chapter 2, we developed a number of complex SEIR models, starting with the simple approach and then continued to introduced new features of the disease in to the subsequent models until all features (population as a limiting factor of the susceptible population, diagnosis delay time, reinfection process etc) of the disease were explored. We test the proposed models in two different context, the low burden and high burden context. Data from over 30 different countries are used to validate the proposed models. We also establish the stability analysis of the model using second generation matrix method. In chapter 3, we use one of the compartmental models from chapter 2 with a slight modification to assess the TB situation in Nigeria. Notification of patient, number of latent infected individual, death due to TB and diagnosis delay time among others are investigated. The model is validated using demographic data from Nigeria. Control strategies such as efficacy of treatment over increased in notification are tested, major problems behind the persistent high prevalence of TB in Nigeria are investigated and identified. The research included in this Chapter has been

published in Complexity (WILEY HINDAWI), the journal has Impact Factor of 2.59 and Rank as Q1 under Mathematics and Interdisciplinary Applications by JCR in 2018.

In chapter 4, we collect data of TB cases from NTBLCP in Gombe state, which is one of the states in the north eastern region of Nigeria. We also describe and show the results of a field work where TB patient were interviewed. This is an effort to identify some of the major challenges that lead to a very low notification of TB cases in Nigeria, which is identified in chapter 3 as the major challenge that facilitates the persistence of TB prevalence in Nigeria. Factors affecting behavioral and socioeconomic aspect of TB in Gombe state are identified from the analysis of the data set collected. We carry out statistical analysis of the aforementioned data sets, and establish relationships between various factors affecting TB control in the state. Some of the result from this Chapter was present at 5th BSC Severo Ochoa Doctoral Symposium.

In chapter 5, we use results from chapter 4 to build an Agent Based Model for TB dynamics in Gombe. The inferential statistical analysis of the data set in chapter 4 with the help of machine learning tools allow for an estimation of the parameters used in building the simulator. We carry out virtual experiments and test 3 types of intervention strategies that can be adopted in order to reduce TB prevalence in Gombe state and Nigeria at large. The result from this Chapter was presented in the International TB Conference in Barcelona, and also at the Annual Departmental Physics Research meeting.

Finally, the sixth chapter consist of a general as well as specific conclusion about the research, with remarks concerning future work.

Chapter 2

Understanding tuberculosis dynamics in different contexts by a sequential increase in complexity of SEIR models

2.1 Introduction

In this chapter, mathematical models of increasing complexity are used to study the effect of susceptible population relative size and the role of the infected population on TB dynamics. Starting from a classical SEI model, two additional compartment are proposed. First, the sick class is split into infectious and non-infectious. Second, the infected population is divided according to the number of years since the initial infection. Virtual experiments are performed to evaluate if the increase in complexity (i) contributes to model accuracy in predicting TB dynamics, and (ii) provides better information about TB. The effect of different actions related to the epidemiological control is assessed by using 20 different countries data of high and low burden settings. Splitting the sick class helps to quantitatively estimate the effects on diagnosis time delay and allows obtaining the percentage of people who abandon treatment soon after diagnosis. Dividing the infected class enables studying the effect of the whole set of latent TB infection population, elucidating the effect of migration flows, and analyzing particular situations such as the reinfection effect.

2.2 Models description

TB epidemiology presents several characteristics that greatly contribute to its underlying complex dynamics. Compartmental models facilitate obtaining a good approximation to the problem. In the case of TB, they allow us to

1. Describe and understand the factors which determine the epidemiological behaviour of the disease in a certain context
2. Facilitate the communication of these factors to actors which are not specialist in mathematical modelling
3. Evaluate several control actions and their effects

A simple SEIR model may correctly describe the global tendency of TB in a certain context and evaluate the main epidemiological parameters like the force of infection. However, this kind of model does not allow to understand the mechanisms that are responsible for such tendency, which may differ drastically from one context to another. Therefore, several compartmental models with different degrees of complexity have been developed (Fig 2.1). On the one hand, simplest models are easier to parameterize, given a certain prevalence data set, and thus appropriate to assess global indicators. On the other hand, more complex models are capable of answering more complex questions regarding the causes for the global tendency and the possible solutions to reverse a bad trend.

2.2.1 Models formulation

The four basic models that have been built and tested are, from the simplest to the most complex, the mathematical analysis of the model is given in Appendix 1:

- The classic SEI model, with 3 compartments (susceptible, S ; exposed or latently infected, E ; and sick and infectious, I). This model allows to estimate global parameters like the force of infection or the size of LTBI population, among others.
- SEI2 model, with 4 compartments (susceptible S ; exposed or latently infected E ; sick and infectious I_1 ; and sick but not infectious I_2). This model allows to answer questions related with the diagnosis delay.
- SE8IR model, with 11 compartments (susceptible S ; exposed in their i -st year of infection, E_i , $i = 1..7$; exposed with more than seven years since infection $E_{>7}$; sick and infectious, I ; and recovered, R). The recovered class do not play a relevant

epidemiological role, but it is considered separately for allowing the identification of the population that has overcome an infectious cycle. This model is appropriate to answer questions related with the characteristics of the exposed class and their consequences on the TB dynamics.

- SE8I2R model, with 12 compartments (S ; E_1 to E_7 ; $E_{>7}$; I_1 ; I_2 and R).

Additionally, an extended model with the same compartments as SE8I has also been developed including a reinfection variable, SE8I-reinf, for the specific case of a context where reinfection is relevant. All of the basic features of the models are depicted in Fig 2.1 and briefly explained below. Finally, all models are simulated in both LBC and HBC countries. A simplification for LBC is also tested in some cases (SEI-low, SEI2-low, SE8IR-low and SE8I2R-low), considering susceptible population as a non limiting factor, as detailed below.

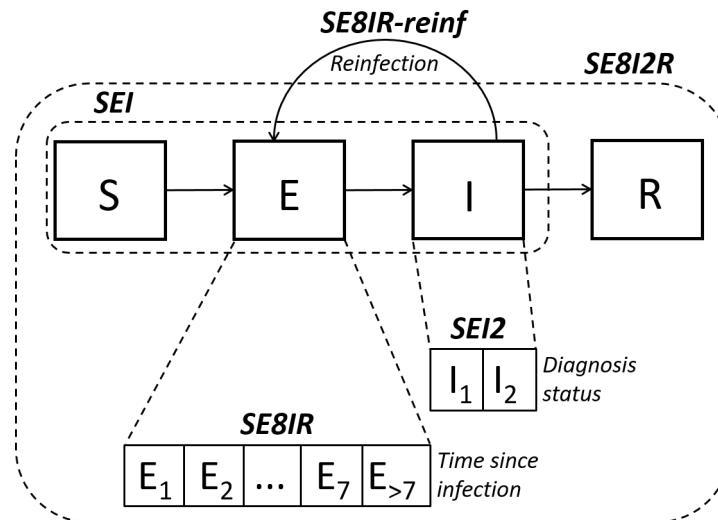


Fig. 2.1 Overview of the several proposed compartment models with different levels of complexity. The main compartments are susceptible (S), exposed or latently infected (E), sick (I) and recovered (R). Sub-compartments of more complex models differentiate between infectious (I_1) and non-infectious (I_2), and account for time since infection. Finally, reinfection of exposed individuals is explicitly considered in one of the cases, thus allowing the flow $E_{i>1}$ to E_1 .

SEI model

Fig 2.2 shows the structure and flows of the classical SEI model for TB dynamics. The new recruits inflow into the susceptible population box (S) at a rate Π , equivalent to the birth rate; this population either die at a rate μ_1 or get infected and move to the exposed compartment

(E) at a rate Λ , where $\Lambda = \frac{\alpha I}{N}$ is the force of infection, α is the contact rate and N is the total population. Since α indicates the number of people infected by every TB-sick, it is possible to assess the impact of this parameter on the disease dynamics.

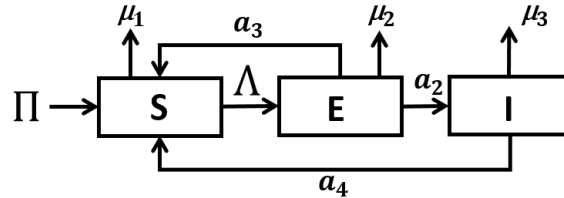


Fig. 2.2 The SEI model with S standing for susceptible, E , the exposed, and I the infectious.

Following a given probability which is generally understood, 10% of the latent infection individuals in E may develop an active disease and move to the sick class (I) within a period of 7 years, the probability function rate is given as a_2 . They can also recover and return to S at a rate a_3 , or die at a rate μ_2 . The sick population spend a given time in the I compartment until they recover and go back to S at a rate a_4 , or die at a rate μ_3 . The dynamics of this model are driven by equations 2.1 to 2.3. The summary and description of the parameters is given in Table 2.1.

$$\frac{dS}{dt} = \Pi + a_3E + a_4I - (\Lambda + \mu_1)S, \quad (2.1)$$

$$\frac{dE}{dt} = \Lambda S - (a_3 + a_2 + \mu_2)E, \quad (2.2)$$

$$\frac{dI}{dt} = a_2E - (a_4 + \mu_3)I. \quad (2.3)$$

SEI-low model

The model SEI is also considered in a setting where the susceptible population is approximated to be the general population (this is the case of low burden countries), that is $N \approx S$. In this setting, the equations for the model are also 2.1 to 2.3 but assuming that force of

Table 2.1 Summary and description of parameters in SEI and SEI2 models

SEI	Interpretation	SEI2	Interpretation
Π	Recruitment rate	Π	Recruitment rate
Λ	TB force of infection	Λ	TB force of infection
a_2	Probabilty rate from E to I class	a_2	Probability rate fro E to I_1 class
a_3	Progression rate from E to S class	a_3	Progression rate E to S class
a_4	Progression rate I to S class	a_4	Progression rate I_2 to S class
μ_1	Death rate in S class	μ_1	Death rate in S class
μ_2	Death rate in E class	μ_2	Death rate in E class
μ_3	Death rate I class	μ_3	Death rate in I_1 class
		a_5	Progression rate I_1 to I_2 class
		a_6	Relapse rate I_2 to I_1 class
		μ_4	Death rate in I_2 class

infection $\Lambda^* = \frac{\alpha I}{S}$.

$$\frac{dS}{dt} = \Pi + a_3E + a_4I - \Lambda^* - \mu_1S, \quad (2.4)$$

$$\frac{dE}{dt} = \Lambda^* - (a_3 + a_2 + \mu_2)E, \quad (2.5)$$

$$\frac{dI}{dt} = a_2E - (a_4 + \mu_3)I. \quad (2.6)$$

SEI2 models

Fig 2.3 shows the flowchart of the SEI2 model. In this new version of the model, the sick class is split into two sub-populations, I_1 and I_2 , where only the individuals in I_1 can infect the susceptible population. I_2 corresponds non infectious. Two new transitions appear. First, the time delay parameter a_5 between I_1 and I_2 . Second, relapse probability for individual in I_2 abandon treatment halfway at a given rate a_6 . This model can be described by equations 2.7 to 2.10. The description of parameters can be found in Table 2.1.

$$\frac{dS}{dt} = \Pi + a_3E + a_4I_2 - (\Lambda + \mu_1)S, \quad (2.7)$$

$$\frac{dE}{dt} = \Lambda S - (a_3 + a_2 + \mu_2)E, \quad (2.8)$$

$$\frac{dI_1}{dt} = a_2E - (a_5 + \mu_3)I_1 + a_6I_2, \quad (2.9)$$

$$\frac{dI_2}{dt} = a_5I_1 - (\mu_4 + a_6 + a_4)I_2. \quad (2.10)$$

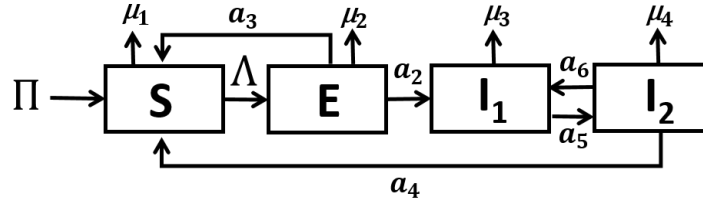


Fig. 2.3 The SEI2 model, with two sick compartments I_1 and I_2 in addition to the susceptible (S) and the exposed (E) compartment.

SEI2-low model

Similarly, the SEI2-low considers a non-limiting susceptible population ($N \approx S$) for the case of LBC. It is described by equations 2.11 to 2.14 with a force of infection $\Lambda^* = \frac{\alpha I}{S}$.

$$\frac{dS}{dt} = \Pi + a_3E + a_4I_2 - \Lambda^* - \mu_1S, \quad (2.11)$$

$$\frac{dE}{dt} = \Lambda^* - (a_3 + a_2 + \mu_2)E, \quad (2.12)$$

$$\frac{dI_1}{dt} = a_2E - (a_5 + \mu_3)I_1 + a_6I_2, \quad (2.13)$$

$$\frac{dI_2}{dt} = a_5I_1 - (\mu_4 + a_6 + a_4)I_2. \quad (2.14)$$

SE8IR model

As stated by WHO, an integral part of the population suffers from latent TB [70], [71], [53], [102]. We consider the probability of developing active TB in the range of 10 – 13% in the first seven years of infection and 5 – 6% the first two years [100] of infection. Therefore, the time since infection distribution among exposed people is very important in understanding TB dynamics. The SE8IR model (Fig 2.4) is an extended version of the SEI model with exposed compartment divided into 8 sub-compartment. We also introduced a an additional compartment for recovered individuals R .

The individuals will move through these eight compartments depending on the time since they became infected. Each infected will spend 12 month in each E_i box with probability $\frac{1}{N_{E_i}} \cdot p_i$ for developing active disease. The total probability for developing active disease of a box E_i at a time t is approximated to be $p_i(t)$. The flow of individual from E_i to E_{i+1} at time t is given as $\frac{i}{N_{E_i}} \cdot k_1$, the total flow is therefore given as $\sum_{j=i}^{N_{E_i}} (\frac{j}{N_{E_i}} \cdot k_1) = k_1$ Compartment

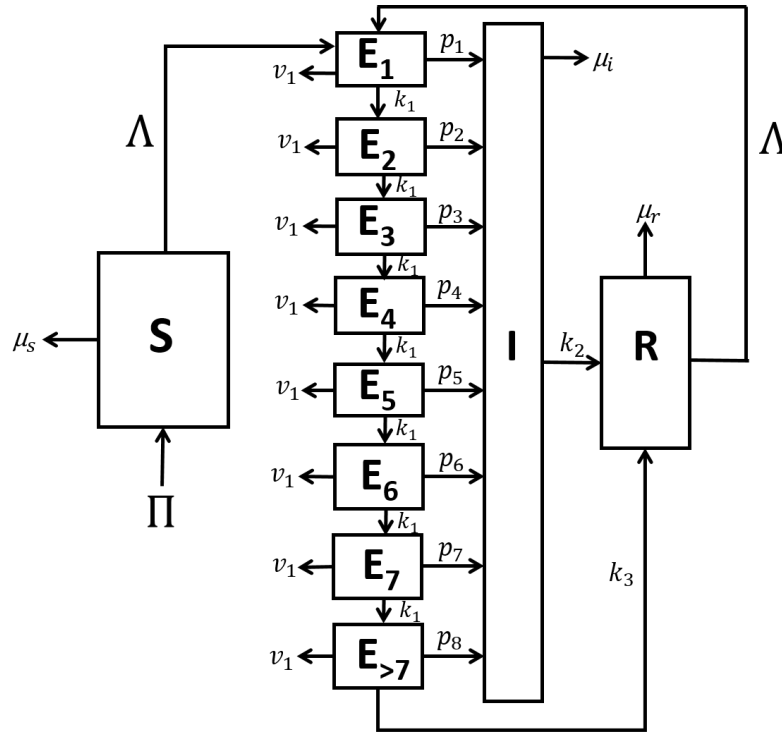


Fig. 2.4 The SE8IR model. In addition to the susceptible class (S) and the sick-infectious (I), the exposed population (E) is divided into eight sub-populations according to the time since infection with a specific probability of developing an active infection, and the recovered class (R).

$E_{>7}$ comprises all the people that have been infected for more than 7 years. The chance to develop an active disease decreases with time [26, 15, 77], therefore the probability of developing active disease in the different compartments (E_i) satisfies the relation $p_{i+1} \leq p_i$ ($p_i = ai^2 + bi + c$, $1 \leq i \leq 8$, $a = 0.001035$, $b = -0.0152$, $c = 0.06$).

The death rate is assigned to each of the E_i s compartment as v_1 . A fraction of individuals in $E_{>7}$ may return to the susceptible population at a rate k_3 . The sick population can either move into the recovered compartment at a rate k_2 or die at a rate μ_i . Equations 2.15 to 2.25 describe the dynamics of the SE8I model and Table B.2 describes the parameters. The model opens the possibility to study effect of migration especially in the LBC and also to evaluate the transitory behavior caused by evolution of the E_i distribution.

$$\frac{dS}{dt} = \Pi - (\Lambda + \mu_s)S, \quad (2.15)$$

$$\frac{dE_1}{dt} = \Lambda(S + R) - (v_1 + p_1 + k_1)E_1, \quad (2.16)$$

$$\frac{dE_2}{dt} = k_1E_1 - (v_1 + p_2 + k_1)E_2, \quad (2.17)$$

$$\frac{dE_3}{dt} = k_1E_2 - (v_1 + p_3 + k_1)E_3, \quad (2.18)$$

$$\frac{dE_4}{dt} = k_1E_3 - (v_1 + p_4 + k_1)E_4, \quad (2.19)$$

$$\frac{dE_5}{dt} = k_1E_4 - (v_1 + p_5 + k_1)E_5, \quad (2.20)$$

$$\frac{dE_6}{dt} = k_1E_5 - (v_1 + p_6 + k_1)E_6, \quad (2.21)$$

$$\frac{dE_7}{dt} = k_1E_6 - (v_1 + p_7 + k_1)E_7, \quad (2.22)$$

$$\frac{dE_{>7}}{dt} = k_1E_7 - (v_1 + p_8 + k_3)E_{>7}, \quad (2.23)$$

$$\frac{dI}{dt} = \sum_{i=1}^7 (p_i E_i) + p_8 E_{>7} - (\mu_i + k_2)I \quad (2.24)$$

$$\frac{dR}{dt} = k_3 E_{>7} + k_2 I - (\mu_r + \Lambda)R. \quad (2.25)$$

SE8IR-low model

Again, a simplification for LBC is considered ($N \approx S$), taking $\Lambda^* = \frac{\alpha I}{S}$ as force of infection in equations 2.26 to 2.36.

Table 2.2 Summary of parameters in SEI8R and SE8I2R models

SE8IR	Interpretation	SE8I2R	Interpretation
Π	Recruitment rate in S class	Π	Recruitment rate in S class
Λ	TB force of infection	Λ	TB force of infection
v_i	Death rate in E_i class	v_i	Death rate in E_i class
p_i	Probability function ($p_i = ai^2 + bi + c$) from E_i to I	p_i	Probability function ($p_i = ai^2 + bi + c$) from E_i to I_1
k_1	Progression rate from E_i to E_{i+1} class	k_1	Progression rate from E_i to E_{i+1} class
k_3	Progression rate from $E_{>7}$ to R class	k_3	Progression rate from $E_{>7}$ to R
k_2	Progression rate from I to R class	k_4	Progression rate from I_2 to R class
μ_i	Death rate in I class	μ_i	Death rate in I_1 class
μ_s	Death rate in S class	μ_s	Death rate in S class
μ_r	Death rate in R class	μ_r	Death rate in R class
		k_2	Progression rate from I_1 to I_2 class I_1 to I_2 class
		g	Relapse rate from I_2 to I_1 class

$$\frac{dS}{dt} = \Pi - \Lambda^* - \mu_s S, \quad (2.26)$$

$$\frac{dE_1}{dt} = \Lambda^* - (v_1 + p_1 + k_1)E_1, \quad (2.27)$$

$$\frac{dE_2}{dt} = k_1 E_1 - (v_1 + p_2 + k_1)E_2, \quad (2.28)$$

$$\frac{dE_3}{dt} = k_1 E_2 - (v_1 + p_3 + k_1)E_3, \quad (2.29)$$

$$\frac{dE_4}{dt} = k_1 E_3 - (v_1 + p_4 + k_1)E_4, \quad (2.30)$$

$$\frac{dE_5}{dt} = k_1 E_4 - (v_1 + p_5 + k_1)E_5, \quad (2.31)$$

$$\frac{dE_6}{dt} = k_1 E_5 - (v_1 + p_6 + k_1)E_6, \quad (2.32)$$

$$\frac{dE_7}{dt} = k_1 E_6 - (v_1 + p_7 + k_1)E_7, \quad (2.33)$$

$$\frac{dE_{>7}}{dt} = k_1 E_7 - (v_1 + p_8 + k_3)E_{>7}, \quad (2.34)$$

$$\frac{dI}{dt} = \sum_{i=1}^7 (p_i E_i) + p_8 E_{>7} - (\mu_i + k_2)I \quad (2.35)$$

$$\frac{dR}{dt} = k_3 E_{>7} + k_2 I - \mu_r R. \quad (2.36)$$

SE8IR-reinf model

The possibility of reinfection has been included in an extended version of the SE8IR model, and it is shown in Fig 2.5. Such force of reinfection is introduced as Λ_1 . The rest of the flows and rates are equivalent to those explained for the SE8IR model. Equations 2.37 to 2.46 show the dynamics of this model.

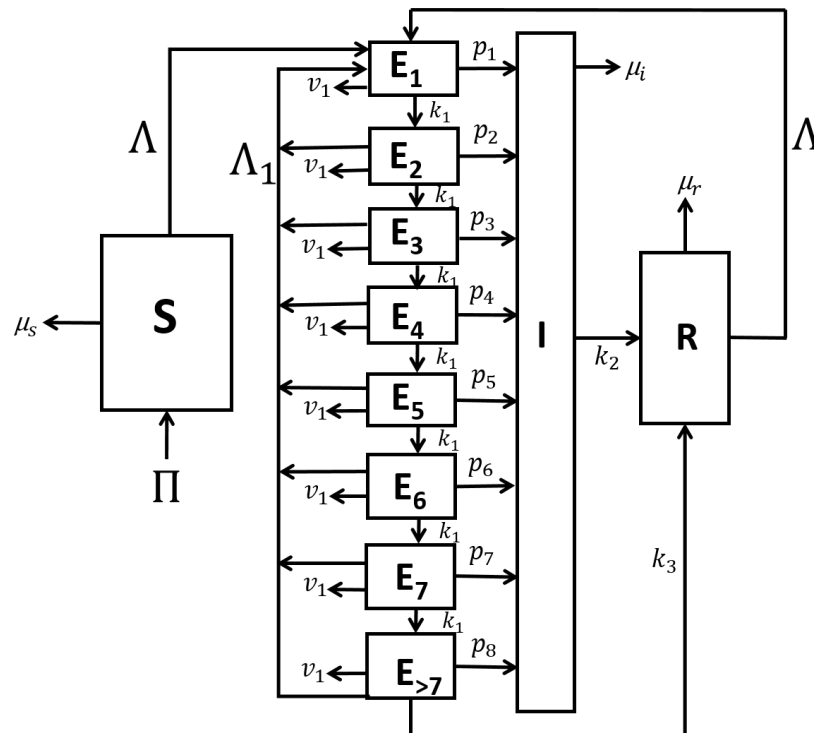


Fig. 2.5 The SE8IR-reinf model, where the feature of reinfection is integrated into the SE8IR model.

$$\frac{dS}{dt} = \Pi - (\Lambda + \mu_s)S, \quad (2.37)$$

$$\frac{dE_1}{dt} = \Lambda(S + R) + \Lambda_1 \left(\sum_{i=1}^7 E_i + E_{>7} \right) - (v_1 + p_1 + k)E_1, \quad (2.38)$$

$$\frac{dE_2}{dt} = k_1 E_1 - (v_1 + p_2 + k_1)E_2 - \Lambda_1 E_2, \quad (2.39)$$

$$\frac{dE_3}{dt} = k_1 E_2 - (v_1 + p_3 + k_1)E_3 - \Lambda_1 E_3, \quad (2.40)$$

$$\frac{dE_4}{dt} = k_1 E_3 - (v_1 + p_4 + k_1)E_4 - \Lambda_1 E_4, \quad (2.41)$$

$$\frac{dE_5}{dt} = k_1 E_4 - (v_1 + p_5 + k_1)E_5 - \Lambda_1 E_5, \quad (2.42)$$

$$\frac{dE_6}{dt} = k_1 E_5 - (v_1 + p_6 + k_1)E_6 - \Lambda_1 E_6, \quad (2.43)$$

$$\frac{dE_7}{dt} = k_1 E_6 - (v_1 + p_7 + k_1)E_7 - \Lambda_1 E_7, \quad (2.44)$$

$$\frac{dE_{>7}}{dt} = k_1 E_7 - (v_8 + p_8 + k_3)E_{>7} - \Lambda_1 E_{>7}, \quad (2.45)$$

$$\frac{dI}{dt} = \sum_{i=1}^7 (p_i E_i) + p_8 E_{>7} - (\mu_i + k_2)I \quad (2.46)$$

$$\frac{dR}{dt} = k_3 E_{>7} + k_1 I - (\mu_r + \Lambda)R. \quad (2.47)$$

SE8I2R model

The SE8I2R model (Fig 2.6) inherits the complexity introduced in SEI2 as well as in SE8IR. Therefore, the exposed population is divided into eight sub-populations according to the time since first infection and in addition, the sick population is divided into two sub-populations: the individuals who have not yet been diagnosed and are spreading the disease (I_1) and the diagnosed, under treatment and isolated individuals who stopped spreading the disease although they are still in the sick phase (I_2). The flows and rates are analogous to those explained above for SEI2 and SE8IR models and are formulated in equations 2.48 to 2.59. Table B.2 describes the parameters of the model.

$$\frac{dS}{dt} = \Pi - (\Lambda + \mu_s)S, \quad (2.48)$$

$$\frac{dE_1}{dt} = \Lambda(S + R) - (v_1 + p_1 + k_1)E_1, \quad (2.49)$$

$$\frac{dE_2}{dt} = k_1E_1 - (v_1 + p_2 + k_1)E_2, \quad (2.50)$$

$$\frac{dE_3}{dt} = k_1E_2 - (v_1 + p_3 + k_1)E_3, \quad (2.51)$$

$$\frac{dE_4}{dt} = k_1E_3 - (v_1 + p_4 + k_1)E_4, \quad (2.52)$$

$$\frac{dE_5}{dt} = k_1E_4 - (v_1 + p_5 + k_1)E_5, \quad (2.53)$$

$$\frac{dE_6}{dt} = k_1E_5 - (v_1 + p_6 + k_1)E_6, \quad (2.54)$$

$$\frac{dE_7}{dt} = k_1E_6 - (v_1 + p_7 + k_1)E_7, \quad (2.55)$$

$$\frac{dE_{>7}}{dt} = k_1E_7 - (v_1 + p_8 + k_3)E_{>7}, \quad (2.56)$$

$$\frac{dI}{dt} = \sum_{i=1}^7 (p_i E_i) + p_8 E_{>7} + gI_2 - (\mu_i + k_2)I_1 \quad (2.57)$$

$$\frac{dI_2}{dt} = k_2I_1 - (g + \mu_i)I_2. \quad (2.58)$$

$$\frac{dR}{dt} = k_3E_{>7} + k_4I_2 - (\mu_r + \Lambda)R. \quad (2.59)$$

SE8I2R-low model

The simplification for LBC, $N \approx S$, involves the same set of equations (2.60 to 2.71) but taking $\Lambda^* = \frac{\alpha I}{S}$ as force of infection.

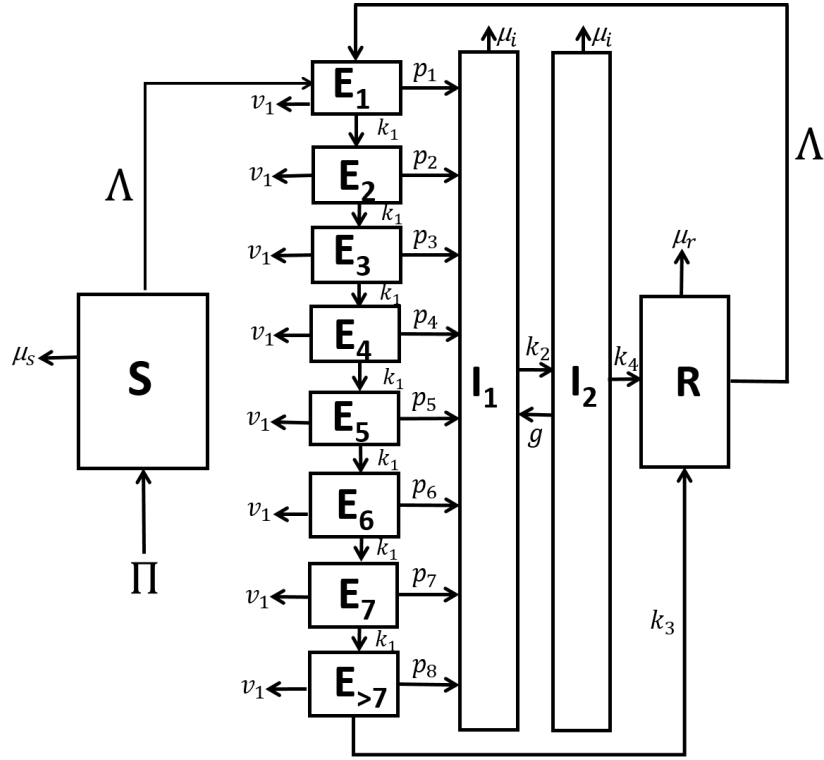


Fig. 2.6 The SE8I2R model. The difference in feature with the previous model is that the sick class also is subdivided in to I_1 and I_2 .

$$\frac{dS}{dt} = \Pi - \Lambda^* - \mu_s S, \quad (2.60)$$

$$\frac{dE_1}{dt} = \Lambda^* - (v_1 + p_1 + k_1)E_1, \quad (2.61)$$

$$\frac{dE_2}{dt} = k_1 E_1 - (v_1 + p_2 + k_1)E_2, \quad (2.62)$$

$$\frac{dE_3}{dt} = k_1 E_2 - (v_1 + p_3 + k_1)E_3, \quad (2.63)$$

$$\frac{dE_4}{dt} = k_1 E_3 - (v_1 + p_4 + k_1)E_4, \quad (2.64)$$

$$\frac{dE_5}{dt} = k_1 E_4 - (v_1 + p_5 + k_1)E_5, \quad (2.65)$$

$$\frac{dE_6}{dt} = k_1 E_5 - (v_1 + p_6 + k_1)E_6, \quad (2.66)$$

$$\frac{dE_7}{dt} = k_1 E_6 - (v_1 + p_7 + k_1)E_7, \quad (2.67)$$

$$\frac{dE_{>7}}{dt} = k_1 E_7 - (v_1 + p_8 + k_3)E_{>7}, \quad (2.68)$$

$$\frac{dI_1}{dt} = \sum_{i=1}^7 (p_i E_i) + p_8 E_{>7} + g I_2 - (\mu_i + k_2) I_1, \quad (2.69)$$

$$\frac{dI_2}{dt} = k_2 I_1 - (g + \mu_i) I_2, \quad (2.70)$$

$$\frac{dR}{dt} = k_3 E_{>7} + k_4 I_2 - \mu_r R. \quad (2.71)$$

2.2.2 Parameter estimation

Overall, before running the simulations, the parameters of the different models are established and analyzed. The differentials are solved in order to establish some correlation between the models. To estimate the different parameters, the fittings are carried out sequentially. Starting with the simplest model (SEI), simulations are fitted to the epidemiological data of a specific country. Subsequently, using the parameter estimates for the simplest model, the SE2I parameters are fitted. This process is repeated for SE8IR, SE8I2R and SE8IR-reinf. The need to fit a gradually increasing number of parameters in the more complex models is partially eased by using information from the previous simpler models fitted. This step-by-step procedure ensures consistency between the fittings of the different models. Moreover, it drastically simplifies the fitting of the most complex models. This methodology allows us to understand the role played by each parameter and deduce that simple models may not be enough to understand some specific behaviours.

2.3 Results

TB epidemiological data from 20 countries is analyzed and validate the models from Section 2.2. The selected countries are categorized into 10 LBC and 10 HBC according to their average TB prevalence (more details in Supplementary material). In this section, a description of how these models are fitted to 10 LBC and 10 HBC is given. We begin by exploring (i) the possibility of using the susceptible non-limiting simplification for LBC countries (i.e., the *-low* version of the different models), and the consequences of using this kind of simplification for HBC, (ii) why simple models are not enough in the study of complex dynamics of TB, and how the increase in models' complexity can reveal some important behavior TB dynamics. Furthermore, the role of the latent infected population and the role played by time since infection in the population on the disease's dynamics in a population are shown. Finally, the effect of reinfection, especially in a HBC, is investigated.

2.3.1 Setting the best models for HBC and LBC: Susceptible population as a limiting factor

A fundamental question regarding the role of population in TB dynamics is to what extent the population size affects the dynamics of TB for every epidemic situation(LBC and HBC). To address this subject matter, various virtual experiments are performed in the two settings with the two sets of models presented in section 2.2. We are interested in observing the epidemic trend when using the model with Λ and the model with Λ^* as their force of infection (i.e.,

taking susceptible population as a limiting factor or assuming that $S = N$). Firstly, based on the average prevalence as either low burden or high burden, the countries are classified. The detail of the countries fitted can be found in the Supplementary material. Finally, simulations with the two sets of models and fitting of these countries are conducted.

Comparison of models performance between low and high incidence burden setting

To illustrate our findings, two countries are taken: Spain as an example of LBC and Nigeria as an example of HBC. Fig 2.7 (A) shows the prevalence dynamics of model with Λ (SE8IR model) and model with Λ^* (SE8IR-low model) as their force of infection versus time in Spain, while Fig 2.7 (B) shows the prevalence dynamics vs time in Nigeria. It is important to note the Y axis in this figure to see the steady state achieved in Nigeria still corresponds to a high incidence. In the case of LBC, all the proposed models perform very well with a very high R^2 value, even after a long period of time. For instance, in the case of Spain, R^2 remains over 0.968 for all models. R^2 for the rest of countries can be consulted in the supplementary materials. In the case of HBC, however, the SE8IR-low model conducts poorly as time elapses. Another point of note is that the SE8IR-low model tends to underestimate the incidence of the disease in HBC. Therefore, as expected, the simplification $S = N$ could only be valid in the case of LBC.

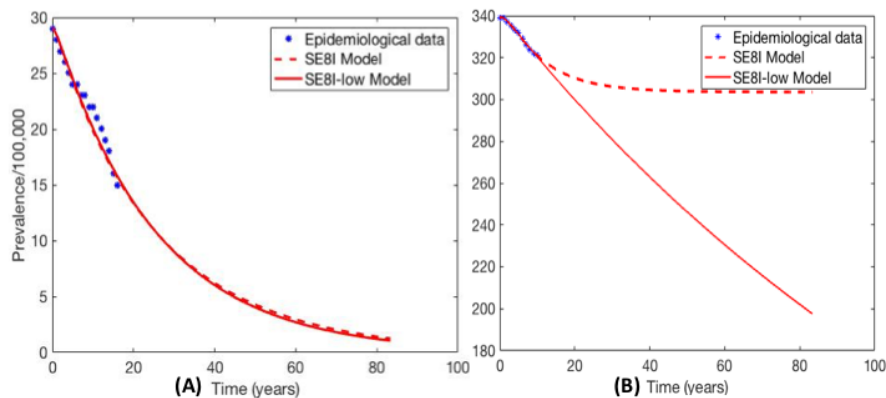


Fig. 2.7 The effect of susceptible population size on the dynamics of TB demonstrated in both low and high burden countries. (A): Result of the model simulation with susceptible population size as a limiting factor (SE8IR model) and without this limiting factor (SE8IR-low model) in an LBC (Spain). (B): Result of the simulation of the model with susceptible population size as a limiting factor (SE8IR model) and the model without this limiting factor (SE8IR-low model) in an HBC (Nigeria).

2.3.2 Models complexity and parameter derivation, using Argentina as a case study

Given the complex nature of TB, a sequence of parameter relations starting from the simplest model SEI to the most complex model SE8IR-reinf is carried out. For each country under study, the SEI model is fitted to the epidemiological data. Subsequently, the SEI2 is fitted to the same data by finding the relationship between the parameters of SEI and SEI2. In a similar way, more complex models are fitted to the epidemiological data, and the parameters of SE8IR, SE8I2R, SE8IR-reinf are also inferred successively. This step-by-step strategy makes it easier to predict and make good estimates of the parameters involved in the more complex models. Furthermore, it allows us to obtain a very good fitting for the countries studied. In the Supplementary material, tables report the best fitting parameter sets for all the models without and with the $S = N$ simplification (*-low*) have been added.

In order to discuss this process, a specific country setting is used and the kind of questions that can be answered by each of the models is also presented. Let us present Argentina, a LBC, to explain how the fitting is carried out. The importance of the models' complexity to understand TB dynamics is elucidated. First, Argentina's setting with the SEI-low model is analyzed. Second, the diagnosis delay in the model with SEI2-low is included. Third, the role of the latent infected individuals with the SE8IR-low model is explored. Finally, the SE8I2R-low model to combine both the diagnosis delay and the time since infection variables in the population.

Fig 2.8 presents the simulation results for Argentina with different models of various complexity levels and compares it with the epidemiological data during 20 years. In this case, the *-low* versions of the models are used because the simplification is feasible for an LBC setting. The experiment starts by fitting the prevalence of Argentina to the SEI-low model and solving the model numerically using Matlab software package ODE-45, the R_0 was analytically computed and the results are in the supplementary materials. The contact rate of the sick individuals in the I compartment for Argentina is then estimated using a formula derived from the R_0 . The solution vector of the sick class I is plotted as a function of time and the parameters of the model are then adjusted until the models are perfectly fitted to the prevalence of Argentina.

The SEI-low model provides a general understanding of the TB dynamics in the country without the need to increase complexity. For example, the disease-free equilibrium (DFE) and the conditions under which an equilibrium exists may be identified for the sick class to approach the DFE, the supremum of the effective force of infection Λ must be found such that $R_0 < 1$. This force of infection for each sick individual strictly depends on the contact rate, which is estimated to be 1.01 per month (12.12 new infected persons per year).

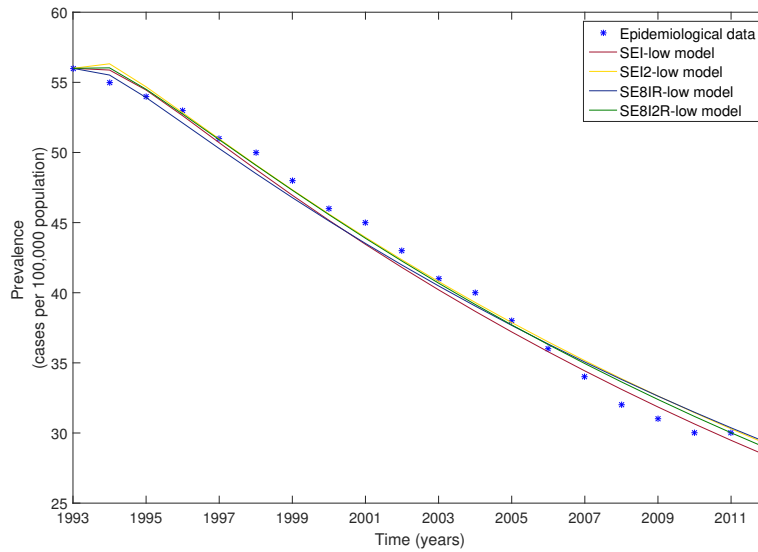


Fig. 2.8 Comparison of *-low* models' fittings for epidemiological data in Argentina. Results of the models simulations with different complexity levels without the population size as a limiting factor for Argentina. Models' parameters are sequentially optimized computationally to fit the simulation results. The models' results are contrasted with epidemiological data of Argentina (blue stars).

Among those infected, only 10-13% are expected to develop active disease. Therefore, each sick individual will produce less than one sick person in a period of six months, which can be considered the threshold of infecting new people by a sick individual before complete recovery.

Consecutively, epidemiological data from Argentina is fitted to the SEI2-low model, which includes an extra compartment taking into account the diagnosis delay time as seen in Fig 2.3. Introducing a diagnosis delay time (a_5) allows the model to assess the effect of an early treatment abandonment (probability of moving from receiving treatment I_2 back to infectious I_1). In Argentina, the diagnosis time delay is fixed at 2 months with a contact rate of 6.1/month, which reduces the number of newly infected TB cases. It is estimated that each case generates 12.2 new infected individuals before they are diagnosed. After dividing the sick I compartment from SEI-low model into a non-infectious I_1 and an infectious I_2 state in the SEI2-low model, the contact rate parameter must be increased to fit the epidemiological data from Argentina, since the time window to infect new individuals is now smaller. This model simulates more accurately the reality of TB dynamics, especially in countries where there is a good TB control (LBC), and facilitates to estimate the number of new infections generated by a non-diagnosed infectious TB sick. From the average 12.2 new infected

cases generated during the first 2 months throughout which the sick individuals remain non-diagnosed, 5% are expected to become sick within the first 2 years since infection and another 5-6% within the next 7 years since infection. Lastly, the different diagnosis delay times tested are ($a_5 = 60$ -days; $a_5 = 45$ -days; and $a_5 = 30$ -days), while all other parameters are kept constant (including contact rate). The results are shown in Fig 2.9. It appears that reducing the diagnosis delay time would lead to a better control of the disease. Implementing policies to reduce the diagnosis time delay would imply faster eradication of the disease.

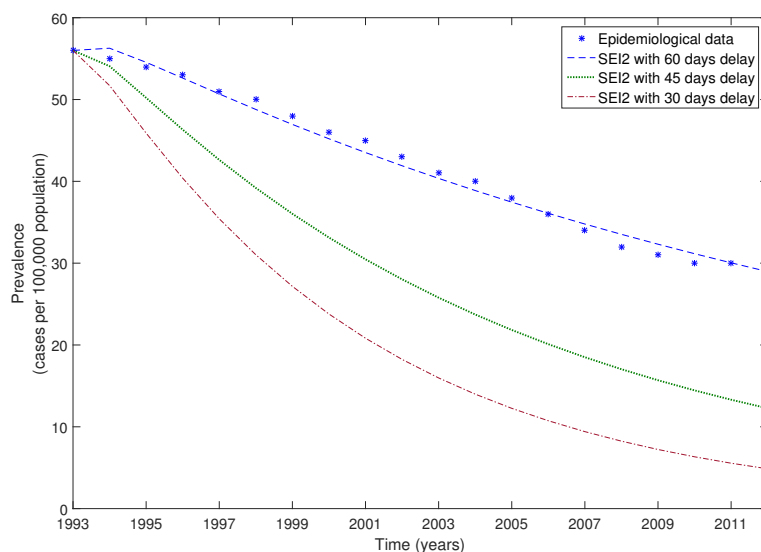


Fig. 2.9 Analysis of TB dynamics in Argentina with SEI2 model showing the effect of diagnosis time delay. The SEI2 model is simulated in the population of Argentina with different diagnosis time delay parameter (60, 45 and 30 days) and compared with the epidemiological data from Argentina while keeping all other parameters constant including the contact rate.

Some of the most intriguing questions in the study of TB are (i) how to determine the population percentage that are latently infected as they show no clinical symptoms, and (ii) what role does time since infection plays in developing active TB. The SE8IR-low is purposely designed to address these problems. Thus, a new way to give an actual approximation of the LTBI in a population is given here. A more detailed analysis of how the infected population can affect the dynamic of tuberculosis is also given in the next subsection.

In Fig 2.8 we present the result of SE8IR-low (blue curve) simulation in the population of Argentina. The LTBI population (E) from the SEI model is split into eight boxes, in an effort to distinguish between the time since infection in this population. The reason behind it is that developing active TB has a lot to do with how long an individual is infected [60, 79, 91].

The initial population of each box is created using the contact rate estimated from R_0 for Argentina, and the probability of developing active TB is set. The result of the simulation shows that about 20 to 28% of the Argentina's population were infected with latent TB during the simulation time span. Fig 2.10 shows our results. Note that about 25% of the LTBI were infected for a period of more than 7 years.

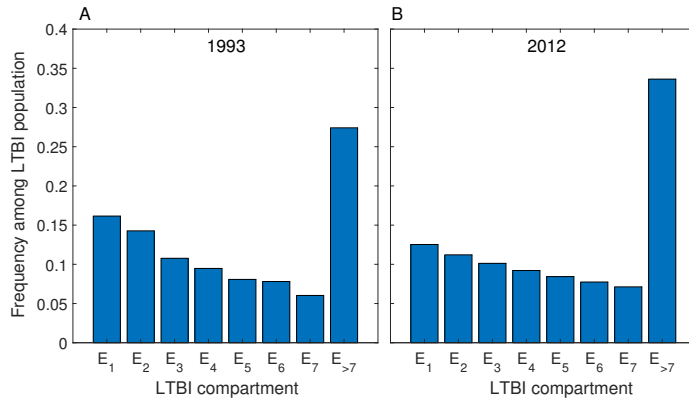


Fig. 2.10 Distribution of exposed population in Argentina according to the SE8IR model. A: Frequency distribution of time since infection among LTBI population using data from Argentina in 1993. B: Frequency distribution of time since infection among LTBI population using data from Argentina in 2012.

With SE8I2R-low model, the diagnosis time delay is combined with the time since infection. The fitting is carried out similarly to the case of SEI8R-low model and the result from this experiment is presented in Fig 2.8 (green curve). In this case, the contact rate for the SE8I2R-low is slightly larger than that of the SE8IR-low. This is partly due to the fact that the sick population I_1 has a smaller window of time to infect the general population. Hence, both the contact rate and the time spent in the sick box have a great impact on the nature and behaviour of TB dynamics. These virtual experiments show that the diagnosis time delay in Argentina is two months and the early abandonment rate for people receiving treatment is found to be 0.2%.

2.3.3 Analysis of the latent infected population, using Niger as a case study

Fig 2.11 shows the role played by the latent infected population in the case of TB dynamics in Niger Republic. The initial population of the latent infected population is estimated to be 33.6% of the population, using the contact rate (α) obtained by calculating the R_0 . The latent infected population is segregated based on the time period since their first infection.

The initial distribution of the LTBI population among E_i compartments and other parameters are adjusted until the model is fitted to the epidemiological data in Niger. The distribution of the latent infected population that fit the data best with an R^2 value of 0.99987 is given in Fig 2.11 (A), in blue. This distribution shows that at least 75.1% of the latent infected population were infected less than 7 years ago, with E_1 and E_2 accumulating more than 28.29%. 24.9% have been infected for more than 7 years.

In order to explore the effect of the LTBI population structure in terms of time since infection, the initial total amount of exposed individuals is redistributed in different ways among E_i compartments, as seen in Fig 2.11 (A). This redistribution is carried out by taking some members of $E_{>7}$ compartment and redistributing them uniformly into (E_1) and (E_2) (green distribution). The result is shown in Fig 2.11 (B) (green curve). A second redistribution is explored by relocating some E_i individuals into the $E_{>7}$ compartment. The result of increasing the population with more than 7 years of infection to about 70% of the latent infected population is also shown in Fig 2.11 (A) (red distribution), and it is presented in Fig 2.11 (B) (red curve).

Even though the total initial LTBI population (i.e., $\sum_{i=1}^7 E_i + E_{>7}$) and all other parameters involved are kept constant, there is clear evidence that the time since infection has a strong impact on the dynamics of TB, and it cannot be neglected when designing control strategies.

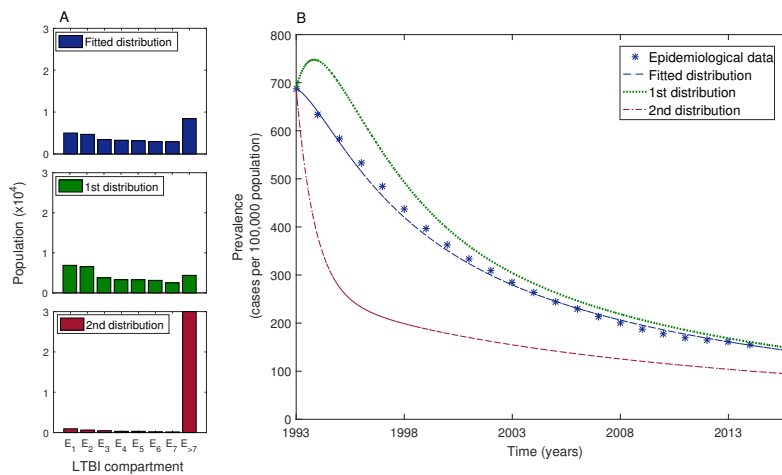


Fig. 2.11 SE8IR model showing the effect of LTBI population to the dynamics of TB in the population of Niger. A: Different initial distributions of the LTBI population used for the simulations. B: Results of the simulations with different initial distributions of a latent infected population, while keeping the total LTBI population constant. Blue: original data and fitted model. Green: relative increase in recent infections (population from $E_{>7}$ is redistributed into E_1 and E_2). Red: relative increase in old infections (population from E_i is redistributed into $E_{>7}$).

Another interesting question that these virtual experiments can address is the impact of migration from HBC on the dynamics of TB. To have an insight into this phenomenon, a similar experiment is carried out where 30,000 infected individuals are introduced with different times since infection. The result of this experiment is presented in Fig 2.12. In the first experiment, the individuals introduced are assumed to be in their eighth (or more) year of infection (box $E_{>7}$). This leads to an increase in the dynamics of the disease as it can be seen in the figure. In the second situation, entering individuals are assumed to be in their fourth, fifth and sixth-year of infection. The same number of individuals introduced in the first experiment is reintroduced into these compartments uniformly and the result is presented in Fig 2.12. Finally, the last simulation introduces all the incoming individuals into the first, second and third-year box. It can be concluded that immigration is also one of the factors which needs to be investigated and taken into consideration when designing TB control strategies in a given community. Despite not showing clinical symptoms, those that have a latent TB infection might still affect the TB dynamics in a community.

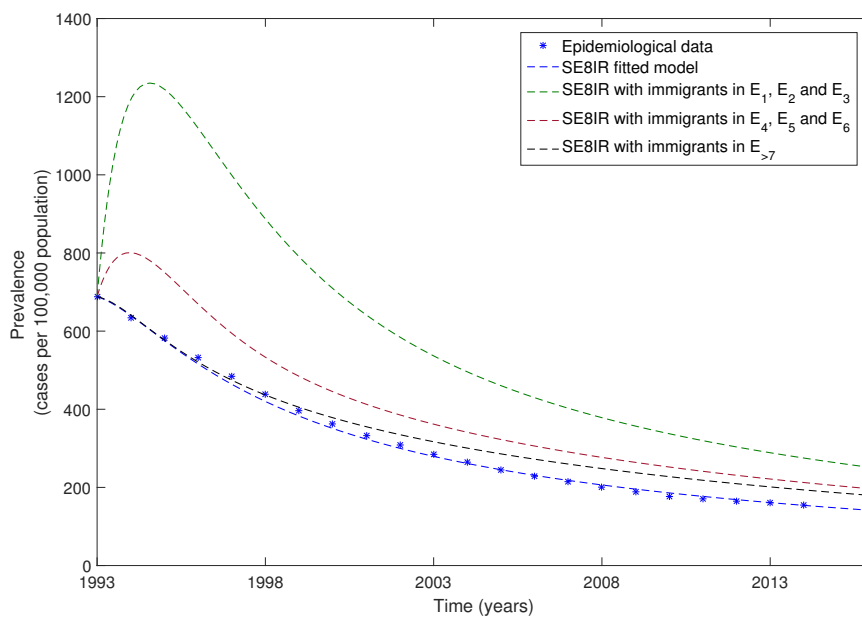


Fig. 2.12 SE8IR model showing the effect of LTBI immigrant population to the dynamics of TB in the population of Niger. The same number of individuals is repeatedly introduced into the latent infected population. In each simulation, the individuals are considered to have different times since infection. Green curve: individuals are uniformly introduced into the first, second and third-year since infection box. Red curve: individuals are introduced into the fourth, fifth, and sixth-year box. Black curve: all the individuals are introduced into the $E_{>7}$ box

2.3.4 Analysis of reinfection using Laos Republic as a case study

In this section we study a context with a very high prevalence, the Laos Republic, where reinfection is relevant. The global procedure of a gradual increase in the models complexity is also followed. By studying such a HBC case, we observed that the higher up the ladder of complexity, the more abrupt the changes are in the behaviour of the models.

The fitting of the models was carried out with the above-mentioned methodology. Let us concentrate on the results of the SE8IR model when fitted to the epidemiological data. The initial exposed population is set as plotted in Fig 2.13 (B), with an initial prevalence of 1500 per 100,000 inhabitants and a period of 69 months for infecting while remaining sick. The fitted contact rate (α) is 0.219. The result of this experiment is given in Fig 2.13 (C) (red dotted curve). Even though the model is correctly fitted to the epidemiological data, the obtained parameters do not seem reliable. Analyzing this situation, it can be observed that both the contact rate and the initial distribution of the exposed class are too small compared to the average prevalence. Moreover, the infection period of the sick class (6 years) appears to be too long as well. In fact, the average lifespan of sick people without treatment is significantly shorter than 6 years. Any attempt to either increase the contact rate, change the initial conditions of the exposed class or decrease the sick period results in an abrupt decrease on the prevalence, which turns out to be smaller than the epidemiological data. Therefore, the model probably is not capturing the main drivers of the TB dynamics in this context.

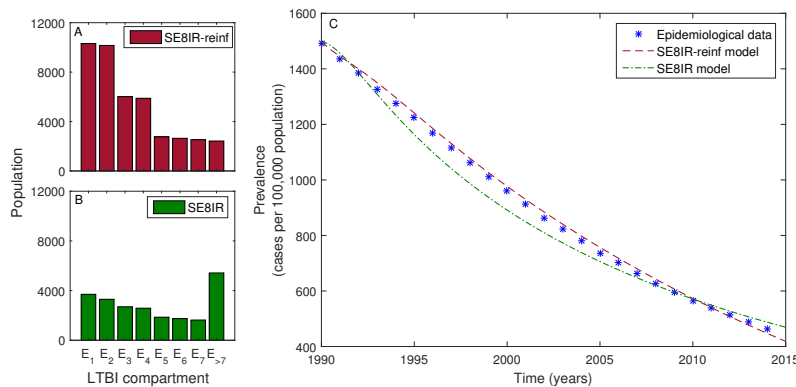


Fig. 2.13 SE8IR-reinf model showing the effect of reinfection to the dynamics of TB in the population of Laos. (A) Initial distribution of the LTBI population used to simulate SE8IR-reinf model in Laos. (B) Distribution of the LTBI population used to simulate SE8IR model in Laos. (C) The red curve shows SE8IR-reinf model fitted to the epidemiological data from Laos showing the effect of reinfection with realistic parameter values. The green curve shows SE8IR model fitted to epidemiological data from Laos with unrealistic parameters.

A possible modification of the model that could make sense in such context is the introduction of the reinfection process (SE8IR-reinf). The exposed or infected classes are assigned a force of reinfection given as $\Lambda_1 = 5 * (\Lambda)$ and using the initial distribution of the infected class shown in Fig 2.13 (A) with about 45% of the total population. The model's fitting to the data, with an initial prevalence of 1500, provides an average contact rate of 1.263/month and a period of 18 months for infecting while remaining sick, with all other parameters set similarly to the SE8IR model. These values appear closer to reality. The model is fitted with a R^2 of 0.999 and the R_0 is found to be 2.0028. The result of this experiment is given in Fig 2.13 (C). This experiment demonstrates how important the reinfection of the exposed class is, especially in countries with a higher prevalence. Furthermore, the initial conditions of the exposed class are also very important in understanding TB dynamics. In the case of HBC, not only does the initial distribution of the exposed class determine the shape of the curve, it also paves a way to the period of time for the sick class to stay active. Other modifications of the model could be tested, as for instance the inclusion of the long-term non-diagnosed class. Nevertheless, the simple consideration of a reinfection process has drastically changed the results. This is a sign of the importance of this process in HBC contexts.

2.4 Discussion

The peculiar behavior of *Mycobacterium tuberculosis* in human lungs is the cause of quite particular epidemiological dynamics compared to other infectious diseases. In this context, mathematical models help us raising questions which enable the understanding and analysis of control methods and the subsequent corrections of health interventions. In this research, we present a set of mathematical models of increasing complexity that allow studying different types of situations.

Departing from classical models (SEI) which are able to evaluate the number of infected people per each sick, new important additions are proposed. First, sick compartment (I) is split into two compartments: the infectious sick and the non-infectious sick who follow a pharmacological treatment. This division helps quantitatively evaluating the effects on diagnosis time delay, not only in terms of the number of people infected by every sick but also in the number of people infected by each unit of time. We study the case of Argentina to illustrate this issue. Furthermore, the splitting allows obtaining the percentage of people abandoning the treatment soon after diagnosis. Second, the set of infected people (E) is divided into eight compartments depending on the number of years since the initial infection. The probability of becoming sick is considered to present a parabolic decreasing behavior. In

this context, we present the case study of Niger. On one hand, this approach allows studying the epidemiological effect of the whole set of LTBI and elucidate the effect of migration flows. On the other hand, dividing LTBI enables the analysis of particular situations such as the reinfection effect, as in the case of Laos Republic.

The proposed methodology with models of increasing complexity provides a progressively ease in the parameterization process of models. In this sense, simpler models are convenient to attain a first approximation of those that are more complex. With this process, it is possible to obtain information on diverse epidemiological characteristics, from the infection speed to the distribution of people according to the time since infected. Moreover, this methodology allows carrying virtual experiments to evaluate the effects of different actions related to the epidemiological control, as shown by using 20 different countries data sets of high and low burden settings.

The proposed mathematical methodology of this work could be extended to study alternative situations. For instance, it would be feasible to assess the effect of having a low percentage of diagnosis by introducing another compartment with sick who do not receive treatment. Another example could be to examine the effect of drug-resistant strains by subdividing the compartment in drug-sensitive and drug-resistant. In this way, the mathematical strategy is able to accurately respond to every situation of a city or country.

Despite mathematical equations proposed in this research are not easily understandable by non-mathematical experts, the model dynamics with flows and compartments is simple enough to facilitate its comprehension by non-specialists. Therefore, models streamline the dialog among specialists in modelling and epidemiologists without mathematical training so virtual experiments can be carried out to help to gauge the effectiveness of control strategies.

Chapter 3

Modelling tuberculosis dynamics in Nigeria for the analysis of its country-level situation

3.1 Introduction

In this chapter, we employ the use of a mathematical compartmental model to evaluate TB burden in Nigeria. We use data obtained from NTBLCP and WHO to validate the model. We point out effective strategies that could be used to effectively reduce TB burden and death due to TB in this country at different levels. We investigate the latently infected population, and try to pinpoint the major reason for TB persistence in the population of Nigeria. Our goal is to develop a close enough to reality mathematical model of TB epidemics that can allow us to investigate effects of demographics and notification of TB epidemics in a population of Nigeria, as well as to estimate the actual burden of the disease including death toll and case fertility ratio in some of the regions in the country.

3.1.1 Facts and hypotheses

All newborn into the model susceptible class is uninfected and free from HIV. For the purpose of this research, we assume a constant birth rate Π and TB/HIV co-infection is not explicitly modeled. These assumptions are simplifications that are not relevant for the current purpose, which is the identification of significant damage caused by poor case detection on TB control program.

We propose to use the model in the exploration of different situations in Nigeria as a virtual experiments platform. The situations evaluated have been selected to start working on

four wide hypothesis, although the final confirmation of all of them would require further research including on-field projects and incorporating experts from other involved disciplines. The four starting hypothesis are the following:

1. TB prevalence still remains a major health challenge in Nigeria due to poor case detection.
2. Increase in effective treatment may not necessarily cause a significant decrease in the prevalence of TB in the country if active case finding is not implemented. In other words, increasing the notification rate is necessary for decreasing TB transmission in Nigeria.
3. The persistence and non-decreasing dynamics of TB in Nigeria are also related to a large number of latently infected population.
4. Poor notification rate has resulted in a large number (hundreds of thousands) of deaths of TB patients.

3.2 Model structure

Figure 3.1 represents the model diagram, which we briefly outline. Birth occurs at a constant rate Π into the susceptible class S with the assumption that all newborn are susceptible to *Mtb*. $E_1, E_2, \dots, E_7, E_{>7}$ represent the non-infectious population infected with tuberculosis without any clinical symptoms (LTBI) in their various years of infection; i.e., an individual who was infected recently (less than one year ago) will be assigned to E_1 , while an individual who was infected last year would be assigned to E_2 population. The last LTBI compartment, $E_{>7}$, consists of all latently infected population with more than 7 years of infection. I_1 and I_2 represent the population that is sick, i.e., they have an active disease with clinical symptoms and they can infect the general population; I_3 is the sick population that was diagnosed and are receiving effective chemoprophylaxis, thus unable to infect anyone. As shown in the diagram, only people in the I_2 compartment will be diagnosed and will move towards I_3 compartment, while the I_1 compartment accounts for the missed TB cases. Death rates in the model depend on disease status; they are fixed into μ_S for susceptible population, ν_1 for exposed classes, and μ_{I_1}, μ_{I_2} and μ_{I_3} for sick individuals of I_1, I_2 and I_3 , respectively. Based on the disparate time scale of natural death versus death due to TB disease, we assume that $\mu_S \leq \nu_1 \leq \mu_{I_3} \leq \mu_{I_2} \leq \mu_{I_1}$

Transmission of *Mtb* occurs following adequate contact between the sick infectious individuals (I_1 and I_2) and the susceptible population. We assume that the latent infected (E_i)

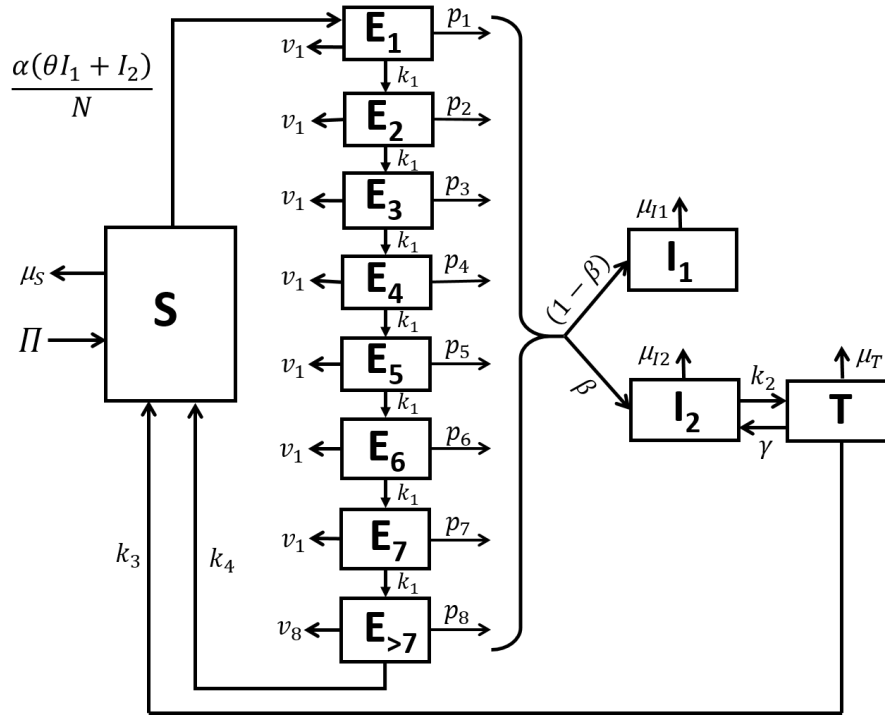


Fig. 3.1 The model: a compartment model for TB transmission in Nigeria. The model shows the dynamic flow of TB including susceptible population S ; the latently infected population are discerned based on the time since they were infected, therefore E_1 stands for people being latently infected for a period of one year; E_2 stands for people infected for 2 years, etc.; $E_{>7}$ stands for people that have been infected for more than 7 years; two types of non-diagnosed sick population are considered, I_1 and I_2 , and sick under treatment is shown as I_3 . Birth occurs at constant rate Π . Transmission of *Mtb* depends on $\frac{\alpha(\theta I_1 + I_2)}{N}$ and probability leading to active TB is given as $p_i = ai^2 + bi + c$, where i is the i^{th} year of infection. k_1 indicate succession from one infected compartment to the next. A fraction of the sick, β , will be detected and notified; k_2 is related to diagnosis time delay before starting treatment and γ indicates a fraction of relapse for sick under treatment. Recovery rate is given as k_3 while k_4 represents the fraction of the infected population that gets rid of the disease spontaneously. We account for all course of death, μ_s , v_i , μ_{I_2} , μ_{I_3} , and death due to active TB, μ_{I_1} .

are not infectious and thus not capable of transmitting the bacteria. We use the incidence expression $\frac{\alpha(\theta I_1 + I_2)}{N}$ to indicate successful transmission of *Mtb* due to nonlinear contact dynamics in large population [54] α is the transmission rate that represents possible interactions that may occur among the susceptible population and the sick infectious population, which is defined as the average contact per unit time. Infectivity of I_1 is assumed to be lower

($0 \leq \theta \leq 1$), as the undiagnosed long-term sick individuals probably reduce their activity due to their poor health conditions.

Newly infected individuals progress directly to the infected class E_1 and stay there for a period of one year, where the probability of developing active disease is p_1 . If they do not become sick, they progress to E_2 (at a *per capita* rate k_1) the subsequent year, where the probability of developing the active disease is p_2 . The progression through the latent infected compartment will continue yearly at a constant *per capita* rate k_1 . Finally, all latent infected patients with more than seven years of infection will remain in the compartment $E_{>7}$ for a very long period of time, where the probability of becoming sick (p_8) is very low and later would be integrated back into the susceptible society at a constant *per capita* rate of k_4 .

The probability of developing an active disease decreases across the latent infected compartment, p_i , is assumed to be an autonomous decreasing quadratic function, given as $p_i = ai^2 + bi + c$, where i is the i^{th} year of infection. The sum ($\sum p_i E_i$) represents the 10-13% of the latently infected population that will become sick in seven years or more, and $p_1 E_1 + p_2 E_2$ is about 5-6% of the latently infected population that will become sick in the first two years of infection.

We define T as the fraction of the sick infectious population that were notified and are receiving effective chemoprophylaxis, k_2 as the rate of effective *per capita* notification and k_3 as the rate of *per capita* of successful therapy completion. We assume that starting a treatment removes an individual from infectious class I_2 and place them into T . We also assume that individuals receiving chemoprophylaxis can abandon therapy, thus showing a relapse at a *per capita* rate γ .

3.2.1 Model formulation

We used a system of non-linear differential equations to model the dynamics of individuals within the population settings. Setting the $N(t) = S(t) + E_1(t) + E_2(t) + E_3(t) + E_4(t) + E_5(t) + E_6(t) + E_7(t) + E_{>7}(t) + I_1(t) + I_2(t) + T(t)$, and suppressing time dependence, t , for each variable, $N(t)$ represent the total population size at time t . The twelve model equations are:

$$\frac{dS}{dt} = \Pi + k_4 E_{>7} + k_3 T - \left(\frac{\alpha(\theta I_1 + I_2)}{N} + \mu_s \right) S, \quad (3.1)$$

$$\frac{dE_1}{dt} = \left(\frac{\alpha(\theta I_1 + I_2)}{N} \right) S - (\nu_1 + k_1) E_1 + p_1 E_1, \quad (3.2)$$

$$\frac{dE_2}{dt} = k_1 E_1 - (\nu_1 + k_1) E_2 + p_2 E_2, \quad (3.3)$$

$$\frac{dE_3}{dt} = k_1 E_2 - (\nu_1 + k_1) E_3 + p_3 E_3, \quad (3.4)$$

$$\frac{dE_4}{dt} = k_1 E_3 - (\nu_1 + k_1) E_4 + p_4 E_4, \quad (3.5)$$

$$\frac{dE_5}{dt} = k_1 E_4 - (\nu_1 + k_1) E_5 + p_5 E_5, \quad (3.6)$$

$$\frac{dE_6}{dt} = k_1 E_5 - (\nu_1 + k_1) E_6 + p_6 E_6, \quad (3.7)$$

$$\frac{dE_7}{dt} = k_1 E_6 - (\nu_1 + k_1) E_7 + p_7 E_7, \quad (3.8)$$

$$\frac{dE_{>7}}{dt} = k_1 E_7 - (\nu_1 + k_1) E_{>7} + p_8 E_{>7}, \quad (3.9)$$

$$\frac{dI_1}{dt} = \sum (1 - \beta)(p_i E_i) + (1 - \beta) p_8 E_{>7} - \mu_{I_1} I_1, \quad i = 1, 2, 3, \dots, 7. \quad (3.10)$$

$$\frac{dI_2}{dt} = \sum \beta(p_i E_i) + \beta p_8 E_{>7} + \gamma T - (\mu_{I_2} + k_2) I_2, \quad i = 1, 2, 3, \dots, 7. \quad (3.11)$$

$$\frac{dT}{dt} = k_2 I_2 - (\gamma + k_3 + \mu_T) T. \quad (3.12)$$

Equation 3.1 describe the rate of change of the susceptible population S . There is a gain into this population through constant birth rate Π . A loss in this population occurs as a result of infection with *Mtb* with transmission rate $\frac{\alpha(\theta I_1 + I_2)S}{N}$ and constant death rate μ_s . Equations 3.2 to 3.9 represent the rate of change of the latent infected population over time. The rate of change of E_1 increases as a result of *Mtb* infection that results in latent infection at a rate $\frac{\alpha(\theta I_1 + I_2)S}{N}$, and decrease by developing active TB at rate $p_1 E_1$, natural death ν_1 , and movement to the second year of latent infection class at rate $k_1 E_1$. $E_2(t), E_3(t), \dots, E_{>7}(t)$, increase by successive movement of latent infected individual from one class to the next and decrease by progressive movement to the next infected class, natural death ν_1 , and developing active TB at the rate $p_i E_i$ in each i 'th latent infected class. A fraction $(1 - \beta)$ of the *Mtb* infection which progress to active TB and were undetected decrease by death due to TB and other causes at the rate μ_{I_1} , and the detected fraction are increased by relapse at a rate γT and decrease by diagnosis at a rate $k_2 I_2$ and by death at a rate μ_{I_2} . The rate of

change of sick individuals under treatment are increased at a rate k_2I_2 , and reduced at a rate γT , recovery rate k_3T , and natural death at the rate μ_T .

3.2.2 Model validation and calibration

In order to determine the effects of various parameters on the dynamics of TB in Nigeria, equation 3.1 to 3.12 are integrated by a Runge–Kutta method of order 4, using a Matlab software (Ode45). We started by fitting our model to the 2000-2010 prevalence data on Nigeria from the official WHO reports and Nigerian epidemiological fact sheets [59]. The reason why this period was chosen is because of reliability of available data together with the fact that all parameters can be approximately considered constant for that period, what would not be applicable if the period was longer. We used the least squares curve fitting in Matlab, by specifying the lower and upper bounds of specific parameters to be estimated. The recruitment rate of susceptible was chosen and calculated such that population of the country remains constant during the simulation. The treatment efficacy $(1 - \gamma)$ is considered to be 80% with probability of relapse taking as 20%. The parameter values that gave the best fit are given in Table 3.1, and were obtained with a $R^2 = 0.9992$.

The initial conditions were chosen in accord with available data when possible (i.e., for total population, estimates of global LTBI population, population in sick compartments and the ratio I_1/I_2). The distribution of LTBI among time since infection compartments was also fitted in this process. We finally obtained these initial conditions (considering a population of 100,000):

$$\begin{aligned} (S(0), E_1(0), E_2(0), E_3(0), E_4(0), E_5(0), E_6(0), E_7(0), E_{>7}(0), I_1(0), I_2(0), T(0)) \quad (3.13) \\ = (58303, 6100, 5000, 4438, 4020, 3550, 3140, 2860, 12250, 284, 55, 0) \end{aligned}$$

This corresponds to an initial prevalence of 339 and a notification rate of 16%. Figure 3.2 show the epidemiological data together with the best fit of the model.

As a validation of the previous calibration, we took data from a wider period (1990-2015) and confronted it with the model. Keeping all the parameters constant but with a small change in the force of infection value, as well as in the initial conditions, the new R^2 was 0.9706. Therefore, we assumed that the calibration for 2000-2010 period as the baseline for our subsequent virtual experiments.

Table 3.1 Summary description of parameters of the model fitted to Nigeria data

Parameter	Parameter description	Values ($month^{-1}$)	Source
Π	Birth rate in S class	Constant	NA
μ_s	Death rate in S	0.012	calibrated
μ_{i_1}	Death rate in I_1	0.146	calibrated
μ_{i_2}	Death rate in I_2	0.146	calibrated
μ_T	Death rate in T	0.01	calibrated
ν_1	Death rate in LTBI	0.0108	calibrated
α	Contact rate of sick population	3.54	calibrated
k_1	Movement rate in LTBI population	$\frac{1}{12}$	calibrated
p_i	Probability of devolving active TB	$ai^2 + bi + c$	derived
β	Notification rate	0.16	[71]
θ	Reduced probability for sick in I_1	0.78	calibrated
k_3	Recovery rate	$\frac{1}{6}$	[71]
k_4	Inverse TB clearance rate	$\frac{1}{120}$	calibrated
γ	Relapse rate	0.20	[71]
ν_8	Death rate in $E_{>7}$	0.008	calibrated
k_2	Inverse of diagnosis delay time	$\frac{1}{3}$	[27]
a	Probability component for active TB	0.001035	calibrated
b	Probability component for active TB	-0.0152	calibrated
c	Probability component for active TB	0.06	calibrated
$(1 - \gamma)$	Effective TB treatment rate	0.80	[71]

3.3 Results

We begin by fitting our model to the current prevalence data on Nigeria from the WHO reports and epidemiological fact sheets [6]. We use the least squares curve fitting in Matlab, by specifying the lower and upper bounds of specific parameters to be estimated. The parameter values that produce the best fit are given in table 3.1. We begin by a plot showing the result of the model as function of time calibrated against the current TB prevalence scenario in Nigeria (Fig. 3.2).

Numerical simulations of the model allowed us to estimate some important parameters associated with TB in Nigeria. In addition, we were able to observe and quantify the effect of the infected population on the prevalence of TB in the country. It also allowed us to make a distinct connection between the notification parameter and the death toll on the sick population (infectious) population, the relationship between the death and the prevalence of the disease in the country. To illustrate these effects, we divide this section into 3 subsections.

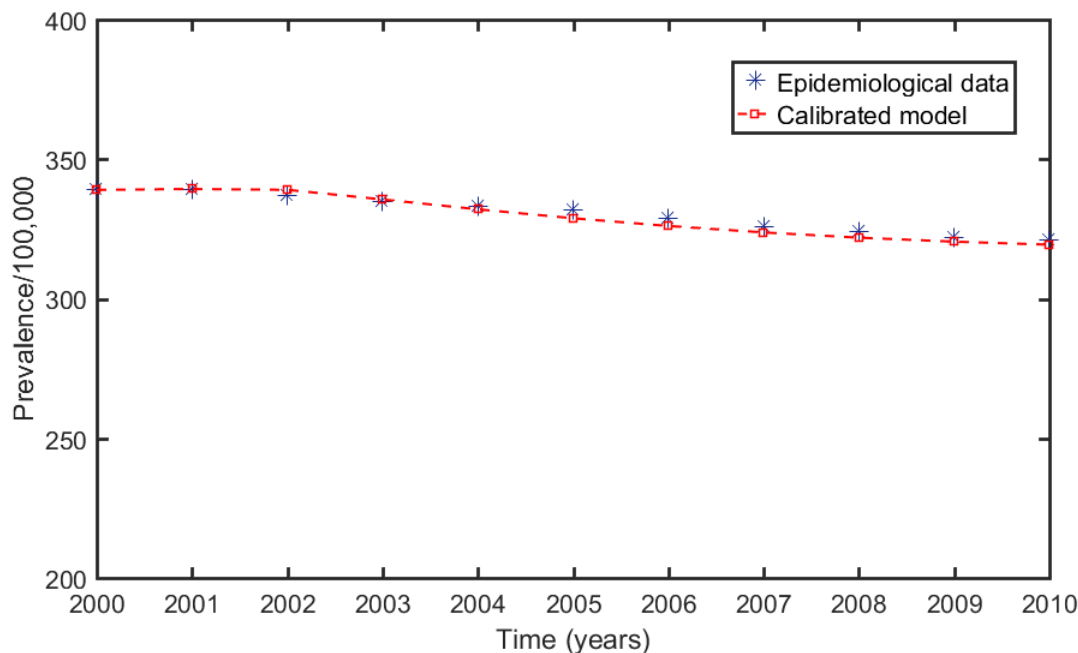


Fig. 3.2 Model simulation calibrated against tuberculosis prevalence data in Nigeria: The figure shows the tuberculosis prevalence curve of the model, calibrated with 2000-2010 data, compared with the data estimates from WHO in that period. The model prevalence curve is obtained by solving equations (3.1) to (3.12), using the initial conditions in equation (3.13) and parameter values in Table 3.1. The goodness-of-fit measure (R^2) for simulation trajectories was evaluated and found to be 0.9992.

Hypotheses 1 & 2 were used to explain the case detection and how effective treatment affects TB dynamics in Nigeria; we made a prediction of TB new cases and provide an alternative strategy for TB control. Hypothesis 3 was used to explain the relationship between LTBI and the dynamics of TB prevalence. Finally, by means of hypothesis 4 we quantified the death due to TB and estimate the death toll due to TB for the next decade in Nigeria, and we also estimate lives that could be saved with the alternative strategy proposed.

3.3.1 Hypothesis 1 & 2: Case detection and effective treatment

The model was successfully fitted into WHO data of TB prevalence in Nigeria. Then, we compared the fitted tendency with predictions regarding an increase in notification parameter. The parameter values obtained after fitting the model into data of Nigeria are briefly summarized and discussed in table 3.1. As shown in Fig. 3.2, the obtained Baseline simulation represents the epidemic TB situation in Nigeria successfully. Values for many

parameters were determined from vital statistics TB data from NTBLCP Gombe state, official TB data from the World Health Organization (WHO) and other recent literature mentioned in Table 3.1. When the values could not be estimated from data or literature, as in the case of the parameters associated with the contact rate, they were obtained with the model's fitting into epidemiological data ($R^2 = 0.9992$).

Recall, WHO estimates that about 30% of the world population is infected with LTBI [104]. However, the model revealed that about 34-37% (depending on the choice of contact rate for the sick population and some other parameters that were calibrated) of Nigeria's population is latently infected with TB. This population is distributed across the various infected classes ranging from E_1 to $E_{>7}$, and the time since infection in this population was also fitted.

We then used our model to investigate how the notification of new TB cases affects the dynamics of TB in Nigeria. We tested the hypothesis that the non-improvement in notification observed in the situation of Nigeria may partially explain why the prevalence of TB in this country is very persistent and not declining as compared to others, showing an incidence rate between 400 and 500/100,000 (more than 130 times the incidence rate in the USA) [43, 45, 23]. The notification rate of TB in Nigeria is about 16% [71]. Therefore, initial conditions for new TB cases were calculated by assigning 16% of new cases into I_2 (i.e., the compartment that gathers the sick population that goes to hospital and receive care at some point), and 84% were assigned to I_1 (i.e., the population that remain sick throughout their life cycle until death).

Fig. 3.3 shows the fitting of the model into the prevalence of TB in Nigeria, namely baseline, together with some simulations where an increase in the notification rate was explored. The mean prevalence for the epidemiological data is 330.64 with a standard deviation of 6.7, while the mean for the model is 330.39 with a standard deviation of 6.6 and a R^2 coefficient of 0.9992 (Table 3.2). The model is pretty consistent with the epidemiological data. The mean resultant prevalence for each simulated notification rate is recorded in Table 3.2 as well. These results show a clear decrease in the prevalence of TB when the notification of the new TB cases increases.

In the Baseline simulation, the model predicted 62,000 deaths due to TB in the year 1990, 72,000 deaths in the year 1995, and 118,000 in the year 2014. These results are within the range of annual TB deaths estimated by WHO [71]. Fig. 3.4 shows the predicted decline in prevalence after 10 years for each of the notification rates tested. In fact, a simply 10% relative increase in case notification of people with active TB reduces the TB prevalence by 15% when compared with the current situation. Predicted TB deaths also would decline by

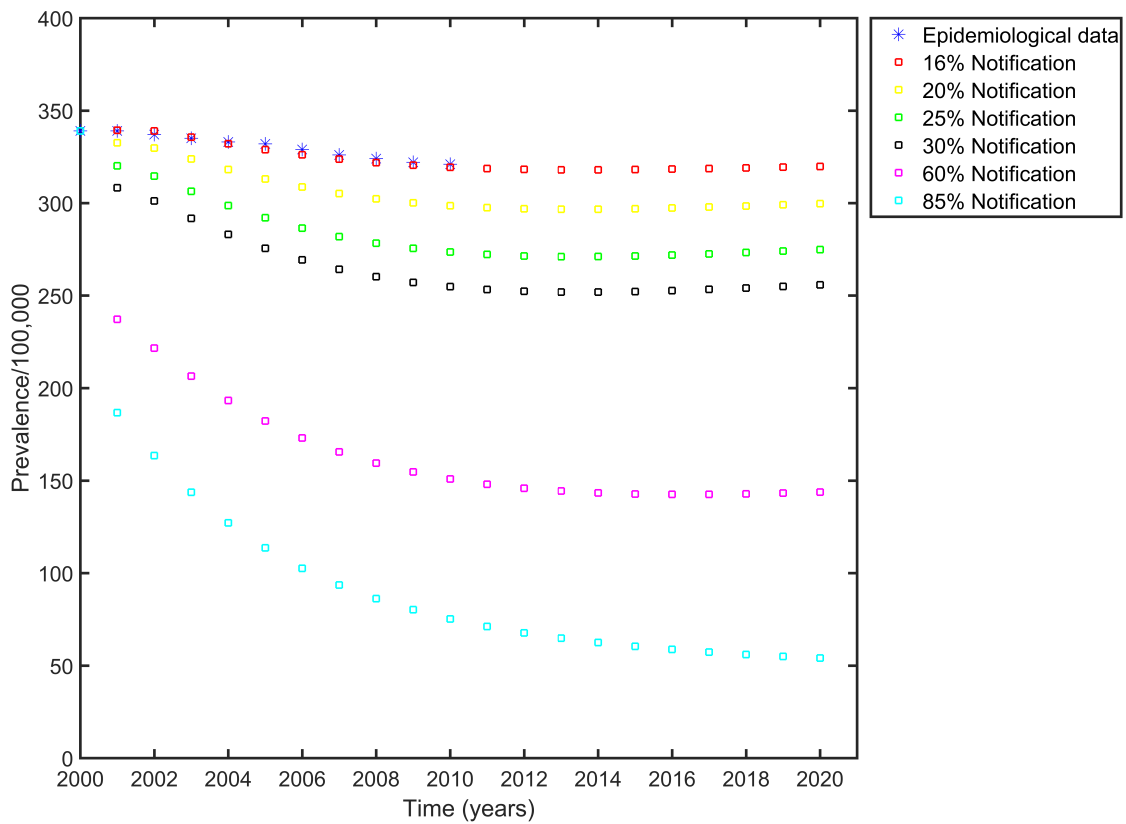


Fig. 3.3 Hypothesis 1: Baseline simulation and effect of the notification rate Model fitted to the prevalence of tuberculosis in Nigeria, the blue star represents the epidemiological data of TB prevalence while the red line-circle represents the model simulation. The fitting was done with 16% of the sick population being notified. The experiment was repeated with 20% (yellow curve), 25% (green curve), 30% (black curve), 60% (magenta curve), and 85% (cyan curve) notification respectively

more than 15% when compared with the Baseline simulation that fitted the epidemiological data.

A new simulation series was designed in order to explore the effect of an improvement of effective treatment, i.e., a decrease in γ . The results showed that a 10% improvement of effective treatment among active TB individual produces a decline of only 5% in the prevalence of TB and related deaths compared with the Baseline fitting, as shown in Fig. ?? . Given these results, the most effective intervention for reducing future TB cases and related deaths in Nigeria as predicted by the model is active case finding.

Table 3.2 Summary of the mean prevalence of data and model simulations, as well as R^2 of the fitting

Case detection rate	Mean prevalence	Standard deviation	References
Epidemiological data	330.64	6.7	[71]
Model 16% Notification	330.40	6.8, $R^2 = 0.9992$	Model
Model 20% Notification	311	NA	Model
Model 25% Notification	288	NA	Model
Model 30% Notification	266	NA	Model
Model 60% Notification	156	NA	Model
Model 85% Notification	105	NA	Model

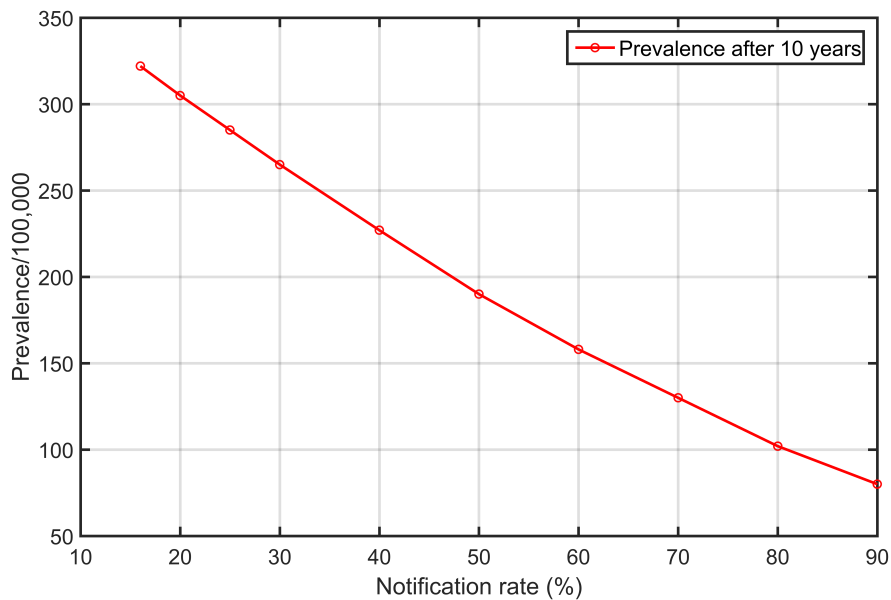


Fig. 3.4 Hypothesis 1: Evolution of tuberculosis prevalence in Nigeria after 10 years of simulation with different notification rate. Each point on the curve shows the prevalence of tuberculosis in Nigeria when simulated with corresponding notification value on the x-axis for 10 years.

A final virtual experiment was designed in order to explore the effect of an improvement of effective treatment in a context with a high case detection rate. In this simulation series, the detection rate was fixed at 80% and different treatment success rates were explored. The results showed that, in this case, the improvement in treatment effectiveness would provide a significant decrease in TB prevalence (Figure 3.5(b)). With a notification rate of 80%, the increase in effective treatment by 10% will lead to a decline in the prevalence of 23% and associated deaths by almost 30%. In the case of a poor case detection rate, increase in

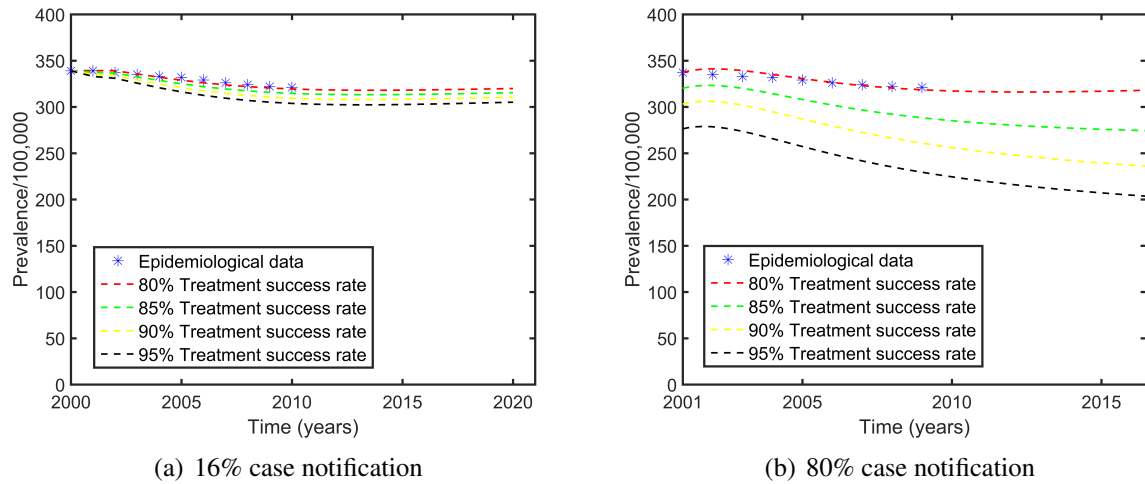


Fig. 3.5 Hypothesis 2: Effective treatment in the population with good surveillance and poor surveillance. (a) The figure depicts a relative increase in effective treatment from 80%, 85%, 90% to 95% in a population with poor case detection, where the notification rate is only 16%. (b) The figure shows a significant decrease in prevalence due to increase in effective treatment from 80%, 85%, 90% to 95%, in a population with a very good case detection, where the notification rate is 80%.

10% of effective treatment is associated with only 5% decrease in prevalence and related deaths. Therefore, an increase in the treatment effectiveness is only relevant if it is combined with an active case finding or in a scenario where the notification rate is significantly high. Nevertheless, with the growing problem multi-drug resistant TB, investment in treatment becomes more relevant and needs to be maintained.

TB prediction in Nigeria

Looking at the current TB situation in Nigeria, one can wonder what the situation will be if all parameters are left unchanged especially in terms of passive case finding. We envisioned this scenario and with the help of the Baseline simulation results, the following results are presented. Fig. 3.6(a) shows the predicted evolution of TB when the notification rate is gradually increased, while Fig. 3.6(b) shows the effect of a sudden increase in this parameter. The simulation starts with the Baseline that represents the current situation (16% notification, red-dashed line). Keeping the notification at this rate without any changes may keep the prevalence of TB in the country in between 350 to 320/100,000 even after 40 years from the present day. This is precisely the present situation of TB in Nigeria, where there has not been any significant changes or decline in TB prevalence for the past 25 years (323/100,000 in 1990 and 332/100,000 in 2015 [71]). Increasing the notification rate up to 20% in 10 years

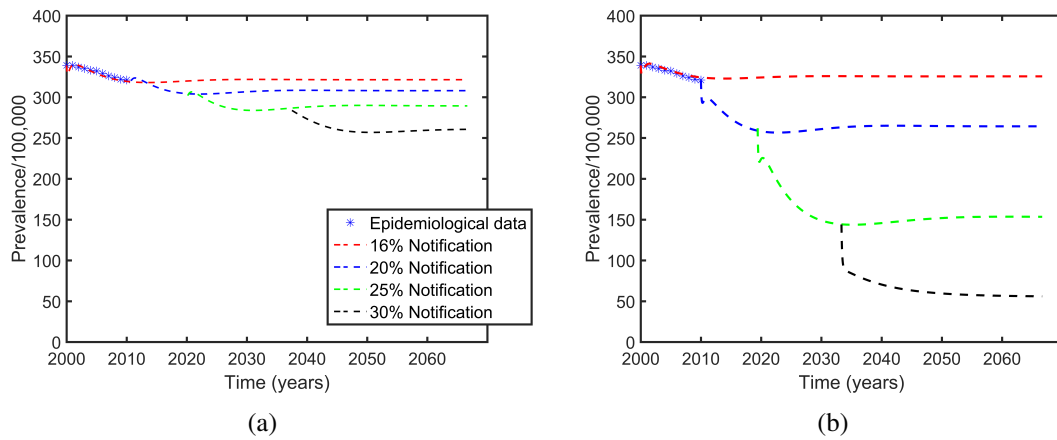


Fig. 3.6 Hypothesis 1: Effective strategies in a gradual increase in TB notification in the case of poor tuberculosis surveillance (a) Result of the prediction made by the model when the notification of new tuberculosis cases is gradually increased. The red curve shows the poor notification rate situation, which represents the present situation of tuberculosis in Nigeria (the blue stars). The rest of the curves represent a change in case detection and increasing in the notification of new cases by some percentages over a period of time. (b) Result of the prediction made when the notification of new tuberculosis cases is suddenly increased. The legend is valid for both plots

could keep the prevalence between 300 to 290/100,000 population. If no more effort is made to increase the notification further, the prevalence would remain in this range even after a period of 50 years. Increasing the notification further to 25% in another 10 years would reduce the prevalence to between 290 and 260/100,000 population, and a further increase in notification to 30% would bring down the prevalence to between 220 and 240/100,000. We repeated the experiment with a sudden increase in the notification from 16% to 30% as seen in 3.6(b) when the prevalence would remain in the range of 270/100,000 to 280/100,000.

For a successful and reasonable decline in TB cases in Nigeria to be achieved, case finding needs to be active and increased from period of time to time. Although a 30% notification rate is viable, a real TB prevalence decrease in Nigeria can only be achieved with a notification rate around 80%. Both strategies presented in Fig. 3.6(a) and 3.6(b) produce the same results at the long run in terms of TB prevalence decline. However, in terms of Public Health, a progressive decreasing strategy would be more feasible due to the economic resources that need to be allocated by Health authorities. In contrast, increasing the notification rate by a large percentage could be very ambitious and resource demanding. At this junction, the most important point is to be able to make interventions as soon as possible given the long time required to actually see changes in TB prevalence.

3.3.2 Hypothesis 3: Latent infected population

Although it is known that latently infected individuals are connected with the incidence of TB, it is not completely clear how this connection can influence the dynamics of the disease in a population. Our first hypothesis implied that poor case detection leads to higher number of un-notified TB cases, which leads to a high number of latently infected individuals that are at risk of developing active TB, thus results in a persistent high TB prevalence. Our model was also used to as a virtual experimental device to test the effect of the latently infected population in Nigeria, which can provide a significant information that can be very difficult or impossible to guess otherwise. We are going show why the prevalence of TB in Nigeria is substantially high and non-decreasing for a very long period of time.

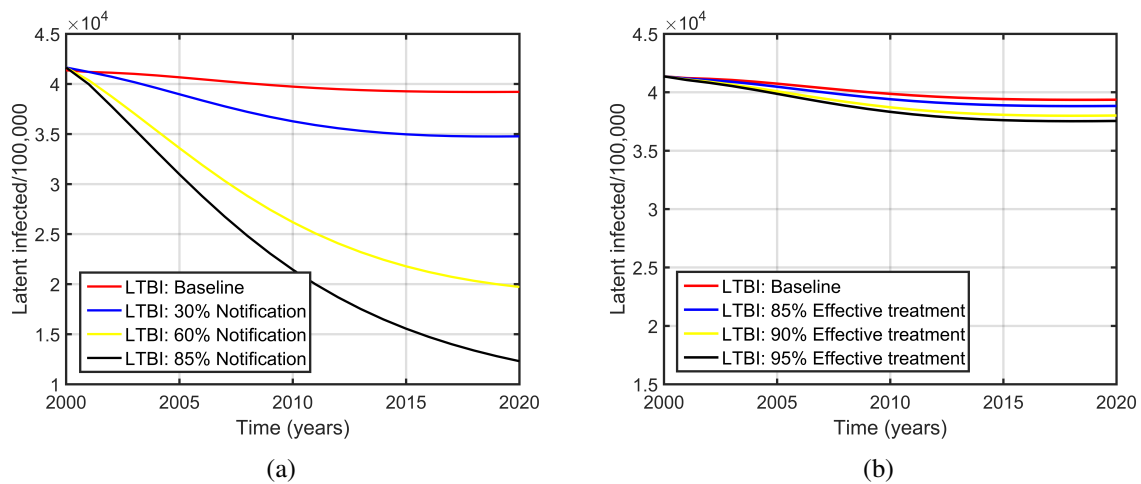


Fig. 3.7 Hypothesis 3: Changes in latent tuberculosis with respect to an increase in notification rate or in effective treatment (a) Simulated dynamics of the latently infected population with different notification levels, with all other parameters kept constant in the population of Nigeria. The red curve shows the latently infected population that remained constant for more than 20 years, thereby producing new sick constantly. With an increase in notification up to 30%, the pool of the LTBI starts decreasing (blue curve), and a further increase in notification shows a much higher decrease in the pool of LTBI (black curve, 85% notification rate). (b) Simulation of the latently infected population with improvement in effective treatment, with all other parameters kept constant in the population of Nigeria. There is little or no change in the number of the latently infected population before and after increasing the effective treatment.

We explored the effect of increasing the notification rate or the treatment success on the LTBI population (Fig. 3.7). Figure 3.7(a) shows the result of various simulations of the latently infected population carried out with different notification parameters. The latently infected population consistently declines with each increase in notification rate. This result

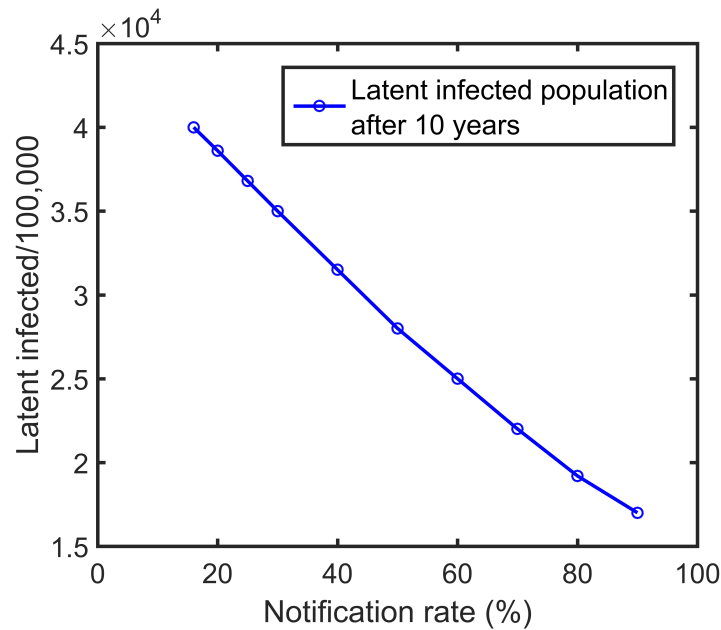


Fig. 3.8 Hypothesis 3: Latent infected population after 10 years of simulation of the model with different notification levels. Each point of the curve represents the corresponding population of the latent infected compartments, when the model is simulated with the corresponding notification rate on the x-axis.

explains why a population with a poor case detection program can fail to achieve a significant progress in reducing new TB cases, because the latently infected population remains the same for a very long period of time due to what is called the replacement principle, where each sick TB patient produce at least one sick patient before death or progress to receive treatment (the number of new cases that receive treatment is very few). This type of situation guaranteed the consistent production of newer TB cases if active case finding is not implemented.

The increase in notification rate does not only affect the active TB new cases but also the latently infected population. A 10% increase in notification was associated with a 38% decline in latent infected TB, while a 10% increase in effective treatment showed only a 2.9% decline in the latently infected population (Fig 3.7(b)).

3.3.3 Hypothesis 4: Poor notification and mortality in sick population

We finally tested the hypothesis that low notification rate has resulted in massive death in the sick population. This phenomenon might be a mystery in Nigeria as not every death is reported to the authorities. Only the deaths that occur in a notified case of TB are recorded as a death due to this disease. As the notification rate of tuberculosis in Nigeria is just 16%, one

can wonder how many deaths cases due to TB went unnoticed. As mentioned earlier, we used the notification of 16% to fit the data to the model.

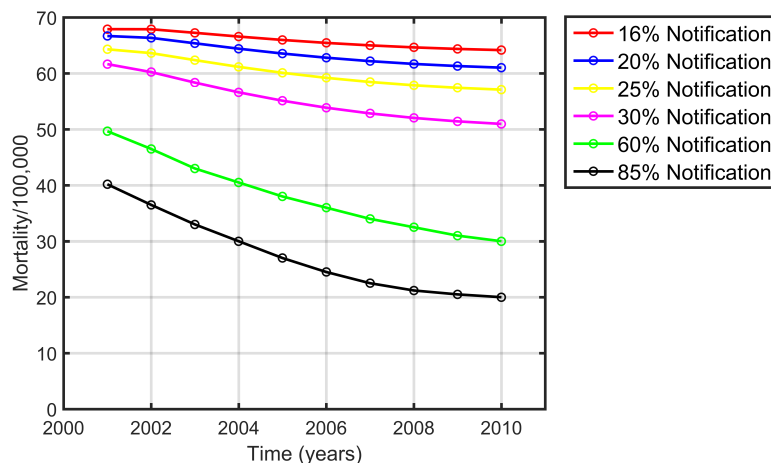


Fig. 3.9 Hypothesis 4: The figure shows the decrease in mortality due to TB in the sick population for a period of 10 years with different notification levels. The red curve represents the current mortality due to tuberculosis in Nigeria (baseline), and the rest of the curves shows what the situation would look like if the notification were to be different.

Fig. 3.9 shows the estimated TB-related mortality per hundred thousand populations in the population of Nigeria obtained from the model. The present-day mortality rate in Nigeria is shown by the red curve, which shows that mortality due to TB in Nigeria is between 70 to 65 individual per 100,000 populations. This result is very consistent with the report from WHO [71]. The model shows that an increase in the notification by 10% would lead to the decrease in mortality by 20%. Actually, we see from the Fig.3.9 that, if the notification is increased from 16% to 25%, the mortality declines to less than 60/100,000 individuals, which implies hundreds of thousands of lives that would be saved. In order to better understand these dynamics, we estimated the mortality rate in Nigeria after 10 years of simulation with the model by using different notification parameters, starting from the current situation in Nigeria (notification rate of 16%). The result of this experiment is shown in Fig. 3.10.

The expected cumulative number of death due to TB with the 16% notification after 10 years of simulation is 4,748,971. An increase in notification into a rate of 20% could have to save up to 292,244 lives within this period. If the notification rate was increased into 25%, up to 584,489 lives could be saved within this period.

Although the estimated TB prevalence in Nigeria from 1990 to the year 2016 remains the same, the estimated number of new cases increases annually. These are attributed to the increase in the population of the country, but at the same time, it shows a massive increase in

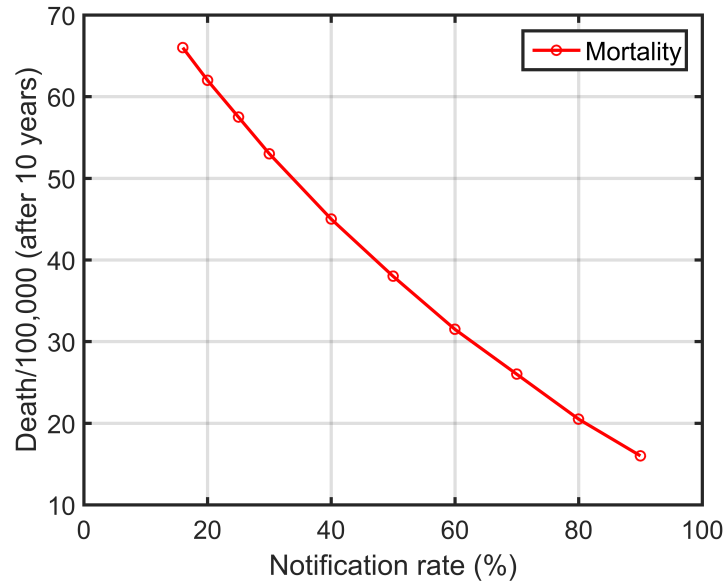


Fig. 3.10 Hypothesis 4: The result of 10-year simulation of the model using different notification rates. Each point on the curve represents the mortality in the sick population after the model is simulated for 10 years with the corresponding notification rate on the x-axis.

the number of deaths due to TB in the population as estimated by the model. The estimated new TB cases rise from 310,000 in 1990 to more than 600,000 in 2015, and the estimated number of TB-related deaths by the model rises from 62,000 in 1990 to more than 120,000 in 2015. If the population of Nigeria continues to grow at the rate of 3.2% annually [90], and the case finding remains passive at 16% notification, then TB-related deaths will reach up to 3,456,640 in the year 2030, this is twice more than the current number of deaths due to TB in the world reported by WHO in 2017 [101]. Depending on availability of funds, political willingness to implement active case finding in Nigeria, and the strategy implemented, millions of lives could be saved from 2018 to 2030. Globally, the model predicted a total death amount of 11,872,530 individuals from 2017 to 2030. An increase in case detection by 15% within this period could save up to 2,557,160 lives and would reduce the incidence by more than 30%.

3.4 Discussion

Epidemiological data from Nigeria shows that TB situation in this country is very persistent in time, in spite of the efforts made by all stakeholders battling the disease. While TB prevalence in Nigeria in 1990 was 323/100,000 population, today is around the same number

(330/100,000) according to WHO. The same situation is observed in the number of deaths due to the disease (70/100,000 today compared to 65/100,000 in 1990). The reason for this steady state of the TB epidemic in the country is still unclear. Many associate it with poverty and population growth. Nigeria recently overtook India as the world's poverty capital [97] with about 87 million people in abject poverty [97]. Its unprecedented demographic growth (from 95,62 million people in 1990 to 206,20 in 2018) is mainly due to a very high fertility rate (5.526 according to the World Bank in 2016). Others presume that there is a need for a better strategy in order to fight the disease towards a significant decline in the near future [27].

Nigeria is a particular setting where several health-care options (medical pluralism) including orthodox medicine (public, private, or drugstores), traditional medicine, spiritual healers, and so forth operate freely [27]. The public health facilities where the TB control programme operates is distanced from citizenship. More often than not, these health facilities are not the first choice during health seeking decisions. Unlike the theory of Dim and Dim [27] and Zwerling *et al.* [105], we believe that high TB prevalence in Nigeria has a lot to do with policies of its TB control program, as stated in other studies [45, 50, 55, 28]. Today, Nigeria is ranked as number 3 in the world in terms of TB burden [6] just behind China and India. The country has adopted the passive case finding as recommended by WHO [59], based on a study conducted in India [57, 13]. Nevertheless, the Indian context has completely different demographic, cultural and literacy settings from Nigeria. However, successful TB control only happens when social and cultural factors are taken into consideration [83]. It is therefore imperative to revise and consider the means of extending the NTBLCP strategy.

In this study, we applied epidemic models to real populations to draw conclusions about one of the leading health challenges in Nigeria. We numerically showed that an increase in the diagnosis rate together with a treatment's success of TB patients is an essential step towards the TB control in Nigeria. In order to increase notification rate of TB in Nigeria using the current passive case detection strategy, the people at the community level should be empowered with adequate knowledge of the growing burden of the disease and accessible potentials for a cure. Alternatively, active case-finding strategies could be explored and implemented, although the existence of many remote communities in some Nigerian areas would dramatically make difficult this task. In any case, the analysis reveals that, with current strategies, it will be impossible to actualize the aims of End TB strategy, with targets to reduce TB deaths by 95% and to cut new TB cases by 90% between 2015 and 2035 [93].

Even with a high increase in the notification, and without considering HIV/TB co-infection and the problem of multi-drug resistant strains (MDR), mortality from TB in Nigeria will still be dramatic for the next few decades to come. Two main facts will

contribute to this increase in mortality. First, Nigeria population growth of more than 3.2% annually [? 90]. Second, the resulting deaths due to TB which are expected to drastically increase by 2030 unless effective TB control is implemented in the country. The results we obtained in our study confirm this situation. This issue poses a great threat not only to Nigeria but also to the entire world TB program and to the ambitious *End TB Partnership* [93] by 2030.

The increase of TB diagnosis by applying active case finding along with specific actions to guarantee TB treatment's adherence and success could be used to achieve the goals set by WHO in Nigeria by reducing TB mortality significantly, TB prevalence and LTBI population. Perhaps the time has come for a critical reexamination of the costs, risks, and benefits of active case finding in this country. Indeed, the strategies evaluated in this paper would require substantial resources and an extraordinary mobilization of international attention, since the greater part of TB control programme funding comes from external sources [59]. We have not attempted to estimate the cost difference of the control strategies proposed. Initiating and sustaining such a large-scale effort of active case finding would be very challenging. However, the first step needs to include both a realistic assessment of implementing these ambitious initiatives and the identification of an optimal way to implement them. For that, the socioeconomic context of the different Nigerian territories and the social inequalities within and between them should be taken into account, so that they can be transformed into feasible policy recommendations. Furthermore, the success could not be guaranteed without both realistic evaluation of the costs of pursuing active case-finding and the treatment of new diagnosed TB sick.

Our results exploring different scenarios show the high relevance of increasing the notification rate in contrast to the modest effects of an improvement of treatment efficacy only. Nevertheless, the treatment efficacy would probably reveal as crucial if we included the problem of MDR, which would be addressed in further studies. Despite contemplating the absence of MDR as a limitation of the model, our conclusions cannot be questioned, since the main problem is still the pool of non-diagnosed TB sick. Future studies could also include the study of the reinfection effect in the model. This new setting would allow a detailed study of the LTBI population and its distribution between old and recent infections. Nonetheless, the reinfection effect would not change the main dynamics given by the current model and the recommendations arisen from it.

To sum up, the model shows that WHO targets stated by The End TB Strategy [93] would not be achieved with the current public management of TB in Nigeria. In fact, its forecasted situation for 2030 suggests being far away from ending the TB epidemics as aimed in the United Nations Sustainable Development Goals under Goal 3. The high percentage of

undiagnosed TB sick in this country is a huge obstacle for the target incidence's disease. It is essential to increase the number of diagnosed and successfully treated, thus requiring not only a significant economic effort but also a boost in social work in the community. This result is not only valid for Nigeria, but also for such countries where non-diagnosis is a serious problem. In all cases, a significant TB incidence's decrease can only be achieved by both increasing the notification rate to 80%-90% range and providing an appropriate management of diagnosed patients. With both measures, new infections can be prevented and the treatment's success can be guaranteed. This situation is the bottleneck that needs to be overcome in addition to the incorporation, improvement and/or strengthening of other essential TB control measures such as Directly Observed Treatment or contact tracing, among others.

Chapter 4

Analyzing socioeconomic factors that facilitate low notification rate in Nigeria. Case study: Gombe State

4.1 Introduction

This section is aimed at analyzing, interpreting and discussing the data obtained from the National TB control program center (NTBLCP) in Gombe state for better perspectives on demographic characteristics on TB dynamics. It also comprises the analysis of data obtained through in-depth interviews of 52 TB patients in Gombe state for better perspectives on socioeconomic characteristics, such as TB status and awareness, knowledge of the participants in TB control. The analysis is aimed at identifying and understanding the most important indicators that facilitate TB spread in the society of Gombe. It is particularly meant to shade more light on the ‘whys’ of the TB spatiotemporal pattern in the study area.

All the analyses have been performed using either MatLab, R, Minitab and/or Excel spreadsheet. In particular, these softwares allow to migrate Excel files to MatLab, Minitab or R, and thus to treat data as vectors or matrices, making really easy doing computations with them and finding relations or particular correlations.

4.2 Material and methods

4.2.1 Review of relevant literature and databases

The pre-field activities began with a review of relevant journals and monographs, unpublished dissertations, WHO fact sheets and other related documents in both hard and soft formats. Other outstanding sources include records from the relevant government agencies, principally the NTBLCP Gombe state branch, the Gombe State Hospitals Management Board and various hospitals in Gombe state. The data that were obtained in this section comprise of data of TB cases obtained from NTBLCP Gombe State between the year 2007 to 2016. It consists of 14,559 cases. The relevant information available from the data related to this study includes age, gender, origin, year of diagnosing and districts of the cases, among others.

4.2.2 Non-documentary data collection

Non-documentary data were simultaneously gathered in order to achieve objectives set in this chapter. Data on demographic, socio-economic, environmental and behavioral factors responsible for TB spread were gathered using In-depth Interviews (IDI) with patients on TB treatment, the healthcare personnel handling the TB patients, and the general population, as well as that on the level of awareness, knowledge and perception of the people on TB. In-depth interview of 52 patient was conducted between the 1st September 2018 to 30th October 2018. Some of the useful information includes age, gender, origin and risk factors, the year of diagnosis, delay time before going to the hospital, the time delay before diagnosis at the hospital, education level, income, job description, marital status, previous contact with a known TB patient, and the district where the patient live. The table containing detail of the interview questions is given in Appendix B.

4.2.3 Methods and instruments of data analysis

These methods form part of the post-field activities, and it comprised the entire procedures for the analysis of the data gathered during the field activities while the instruments comprised the principal techniques and tools employed in achieving the methods. We marrying up part of documentary data with the non-documentary data gathered from the field, and then analyzed it simultaneously. The data gathered were coded, categorized, analyzed iteratively and discussed. We employed the use of inferential statistics and Anova analysis of mean, with the help of machine learning techniques to make reasonable predictions and estimation of some of the parameters involved in the model. Time delay at the hospital and before going to the hospital was sougheed and analyzed using Matlab and python software.

4.2.4 Socio-demographic data of Gombe state

Gombe state is one of the 36 states in the Federal Republic of Nigeria. It was created on 1st October, 1996 out of the former Bauchi State, with headquarters in Gombe city as seen in Figure 4.1. It is located in the northeastern part of Nigeria, The boundaries of the state roughly correspond to those of the Tangale-Waja Chiefdom and Gombe Emirate. The state share common borders with the states of Borno, Yobe, Taraba, Adamawa and Bauchi. The Gombe state has an area covering 20,265Km² and has an estimated population of 3,256,962 people on 2016 [92].

As transmission of TB has a social aspect that can not be overlooked and requires the understanding of population distribution, it becomes imperative to mention some aspect related distribution and population size of Gombe state. The current population of Gombe state was estimated to be 3,400,066. The estimation was carried out with the help of information from National Bureau of Statistics (NBS), starting from 2006 with a population of 2,365,000 (by the NBS 2006 censures), at annual growth of 3.2% using logistic population growth. Figure 4.2 shows the population growth from 2006 to 2017 and Figure 4.3 shows the current gender and age group distribution of the state.

Gombe state has one Federal Medical Center, a state University Teaching Hospital and 22 secondary health facilities (General Hospitals). There are also a total of 508 primary health care centers ranging from Health post to Health Clinic and Comprehensive Health centers. Majority of the citizens (72.2%) live below \$1/day [92]. The potentials of the state are limited by an infant mortality rate of 20.7/1000 live births, maternal mortality rate of 1002/100,000 live births, an estimated HIV prevalence of 3.9%, and under 5 mortality rate of 104/1000 (2008 estimates); the principal causes of morbidity and mortality are malaria, tuberculosis, pneumonia, vaccine preventable diseases, snake bite, road traffic accidents, and Acquired Immune Deficiency Syndrome (AIDS)[92].

4.3 Results

In this section, we present and discuss result from various analyses conducted on different set of data collected with the vision of identifying the main reason for a low notification in Nigeria.

4.3.1 Results from the analysis of TB data from NTBLCP

The analysis of data from NTBLCP disclosed that males are more vulnerable to TB than females, and that the age group of 25-45 is the most affected group with 48% of TB cases.

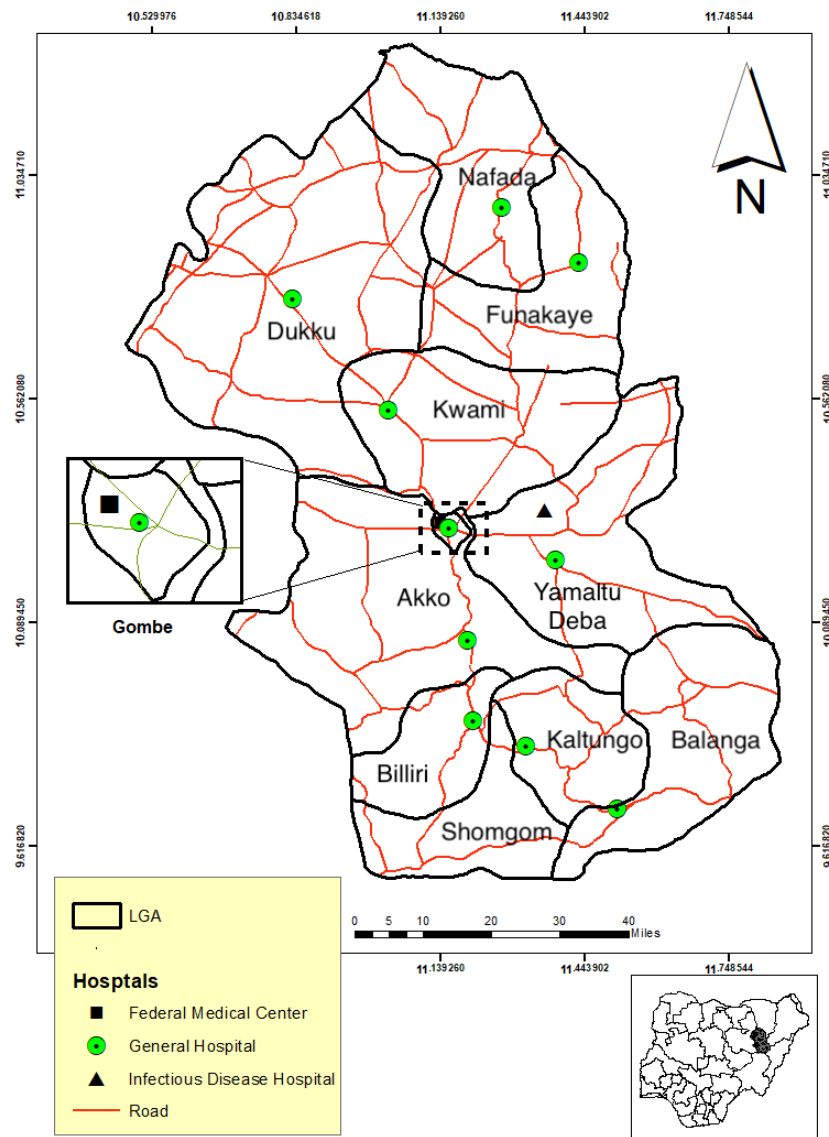


Fig. 4.1 The map of Gombe state, showing tertiary as well as secondary hospital in all the districts of the state

This is to say, the TB situation with regards to gender and age group agreed with the assertion of the most of literature, that the most active and productive age group is more vulnerable to TB. Moreover, males being more active and more productive are more vulnerable than their females' counterpart in the state as shown in Figures 4.4. The United States Embassy in Nigeria (2012) had specifically asserted that, the age groups commonly affected were the

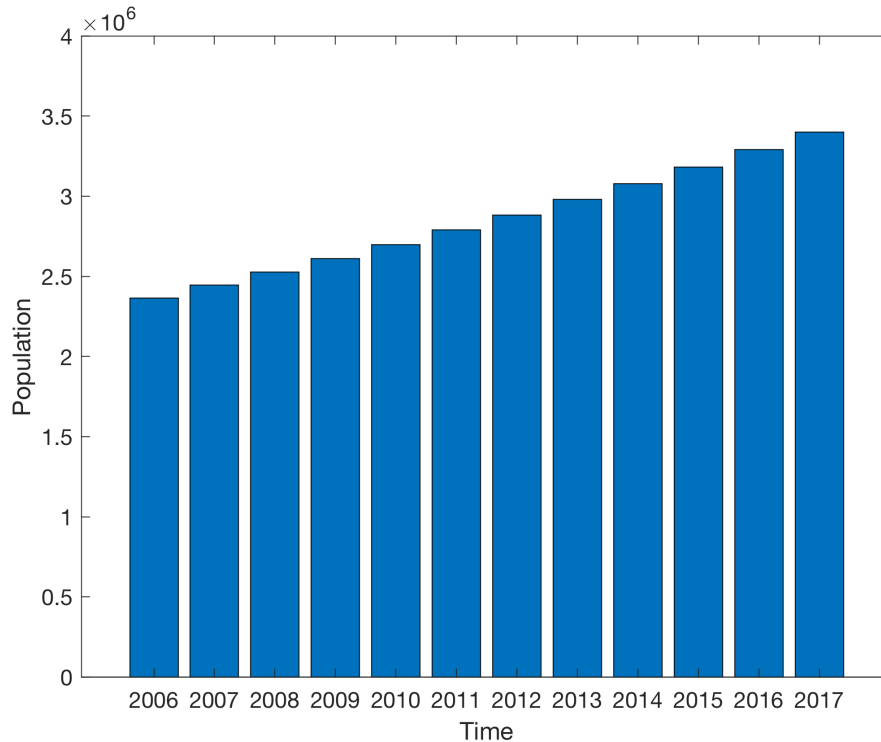


Fig. 4.2 Gombe State population: Estimated population of Gombe state from 2006 to 2017

most productive age groups, with the 25–34 age groups accounting for 33.6% of the smear positive cases registered in Nigeria. However, 61.5% and 38.5% prevalence were found in males and female patients respectively, with age group 31–40 years having the highest prevalence of 28.8% [65].

In summary, the age groups most commonly affected by TB in Gombe state are also the most productive age groups, with the 25–34 age group accounting for 26.7% and 35–44 age group account for 21.3% of all TB cases registered from 2007 to 2016 in Gombe state. Another close look at the data reveal that the age group 25–44 contributed 48% of all TB cases from 2007 to 2016, and male in this age group contributed 63% of the cases. In other words, with some level of confidence, we can say that male in the age group of 25–44, are probably an important source or agent of TB spread in Gombe state.

Although TB cases appeared to be highest in the age group of 25–44, TB incidence/100,000 is also very high in the age group of 45–65+. Figure 4.5 shows the trend of TB incidence/100,000 population in each age group. As we have seen in the Figure, although the population of people in the age group of 45–65+ is not very high (from Figure 4.3), the TB cases reported in this age group is still very high. Thus it results to a higher TB incidence in

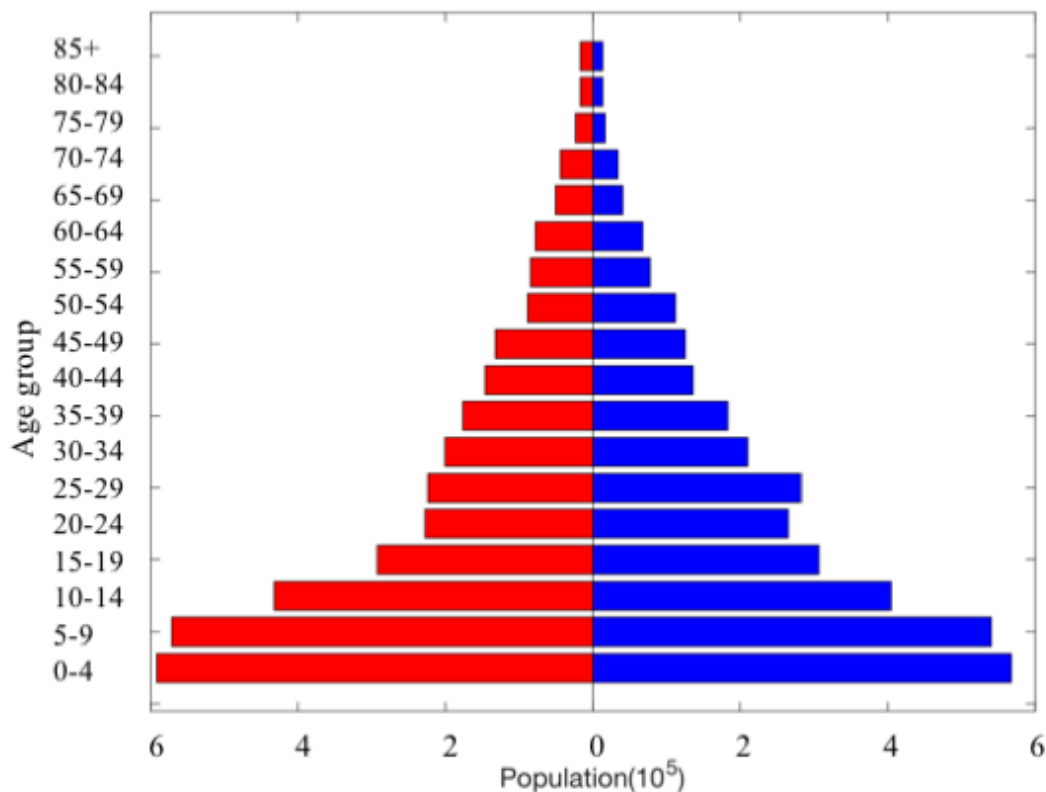


Fig. 4.3 Gombe state gender and age group of the population in the year 2017

the aforementioned age group as seen in Figure 4.5. This particular dynamics in TB trend is very unfamiliar and contrary to what was found in most literature. TB incidence in the age group of 25-34 reveals similar to TB incidence in the age group of 55-65 and 65+ in Gombe state according to notified cases.

22% of the reported cases are living with TB/HIV co-infection in Gombe state. The World Health Organization (WHO)[25] estimated that 27% of Nigeria's TB patients are HIV-positive; this gives an insight on the dangers of increasing TB cases in the country since the two diseases are opportunistically related. It is evidently clear that although Gombe state has an estimated very low HIV prevalence of about 3.9% [72] yet the percentage of TB/HIV co-infection is very alarming especially in female patient. Even though we saw in the previous figures, males contribute higher percentage of TB cases in the state, but Figure 4.6 indicates that female are at more risk due to their higher percentage of TB/HIV co infection. Figure 4.7 gives a more detailed description of TB cases and TB incidences in Gombe state

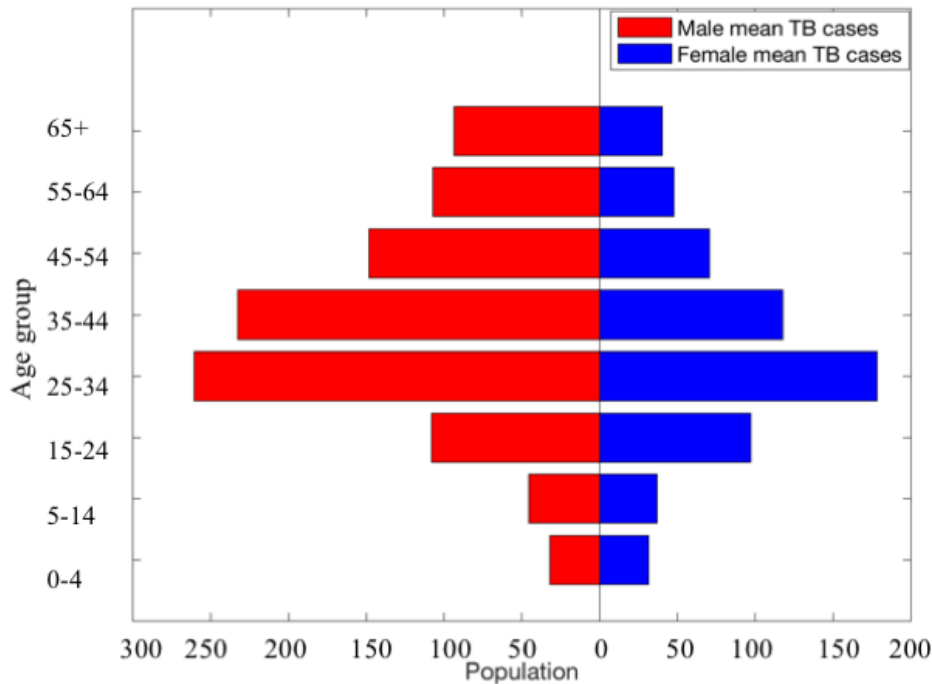


Fig. 4.4 The mean reported TB case by gender and age group from 2007 to 2016 in Gombe state Nigeria.

The mean incidence per 10^5 population in each district is given in Figure 4.7. It is clear that Gombe city alone contributed about 50% of all the TB cases reported in Gombe state with mean TB incidence of $202/10^5$. This can be attributed to the very high population density ($7,409Km^{-2}$ of the city compare to the other district and the fact that medical care is more available. The second highest incidence is Kaltungo with average TB incidence of $80/10^5$ and a population density of $244Km^{-2}$. Although there are other districts with a higher population density than Kaltungo, the high HIV prevalence in Kaltungo area could be one of many explanations that makes it possible for the district to have a higher TB incidence than other districts.

4.3.2 Analysis of data from field's work

It is evident from the summary of various interviews conducted in Gombe state that almost 0% of the patients are aware of TB symptoms and signs before they were sick with the disease. These implies that a great percentage of the public does not know how active TB looks like and would not take any precaution to protect themselves from TB. Hence, the greater percentage or even majority of the public are at a very high risk of being infected

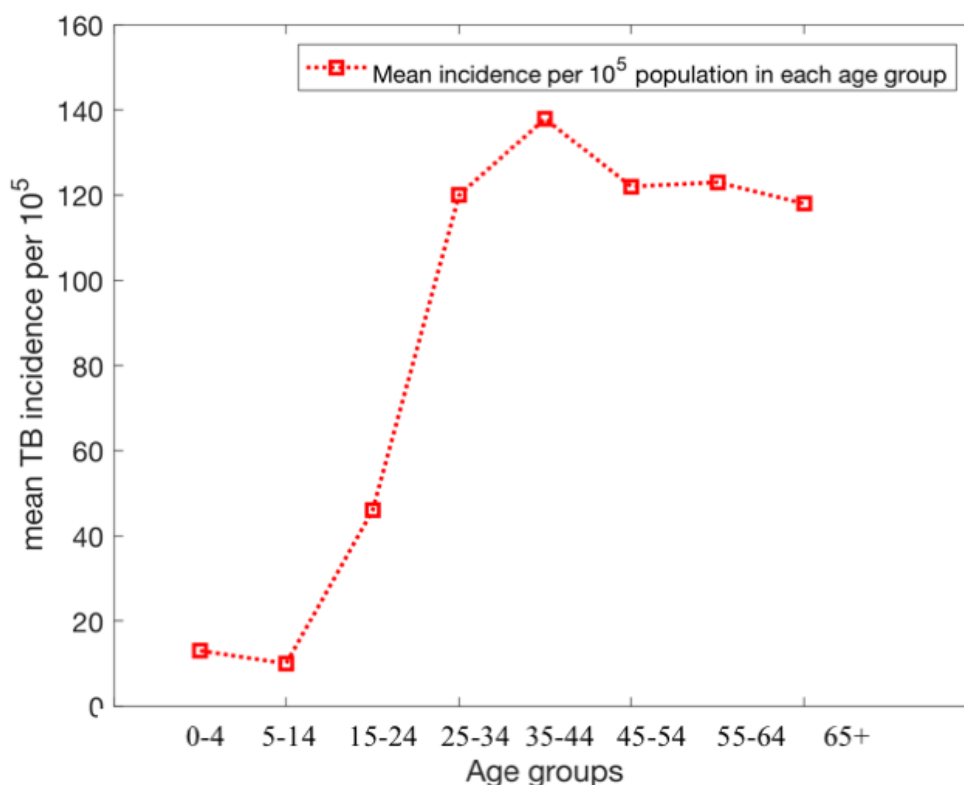


Fig. 4.5 The mean reported TB incidence by age group from 2007 to 2016 in Gombe state Nigeria.

with TB. Also, from the interviews held, only about 2% of the patient have an absolute understanding of TB transmission and spread, about 20% have some level of understanding of TB transmission and 78% have no idea or what so ever on TB transmission and spread. This is a very serious and alarming situation, there is an army of TB sick people in the mixed of the people and more than 78% of them have no idea about the disease transmission talk less of how to prevent others from being infected with the disease. These was illustrated in Figure 4.8.

It was also evident from the interviews that only a few health care workers have a full understanding of TB and its transmission; this is evident in the way little or no information on how one can protect others from being infected was given, only 37% of the patient were informed by the health care workers on TB transmission pattern as shown in Figure 4.8. Moreover, from some of the sessions held with the health workers during our visit to Gombe state, it is clear that most of these health workers does not know about Bovine TB, which some of the patient may be suffering from due to their interactions with herd of cattle and livestock (Many are herders). There is very little percentage (14%) of smokers amongst the

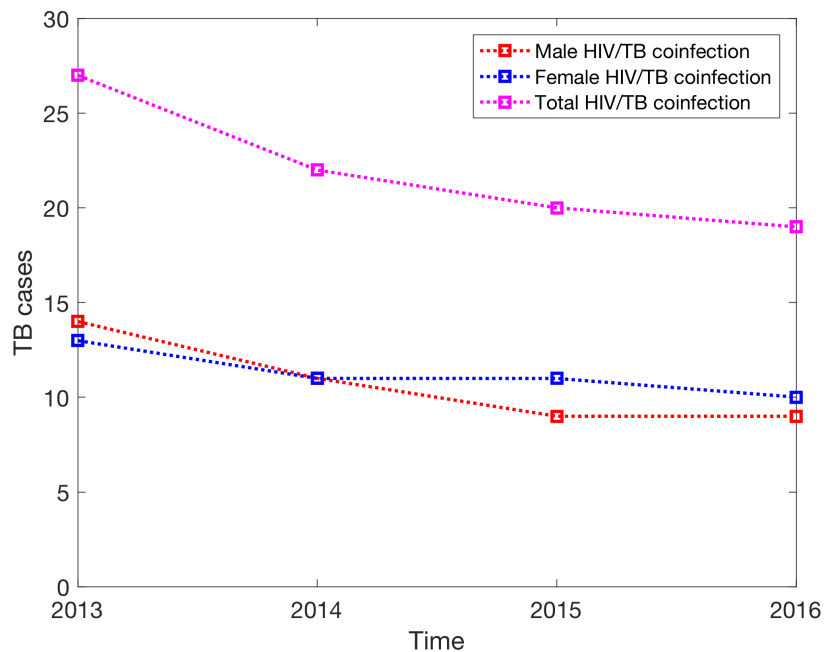


Fig. 4.6 HIV/TB Co-infection: The mean reported cases of HIV/TB co-infection from 2007 to 2016 in Gombe state Nigeria.

patient interviewed, and about 26% of the interviewed patient are HIV+. 20.4% are MDR and about 29% of them suffered from some kind of chronic disease. It was also deduce that the excessive long delay before going to the hospital in the average of 26 weeks is due to many factors that could be associated with the following:

- Lack of nearby health care facility
- Alternative medicine which most of the patient admit on trying before going to hospital
- Miss diagnosis in both hospitals and smaller clinics visited by the patient
- Poverty and lack of resources to access health care
- Constant visit by the patient to untrained health workers that have smaller rooms in the society operating as alternative to hospitals
- Ignorance and unawareness of the disease
- Ignorance and unawareness of free treatment for the disease

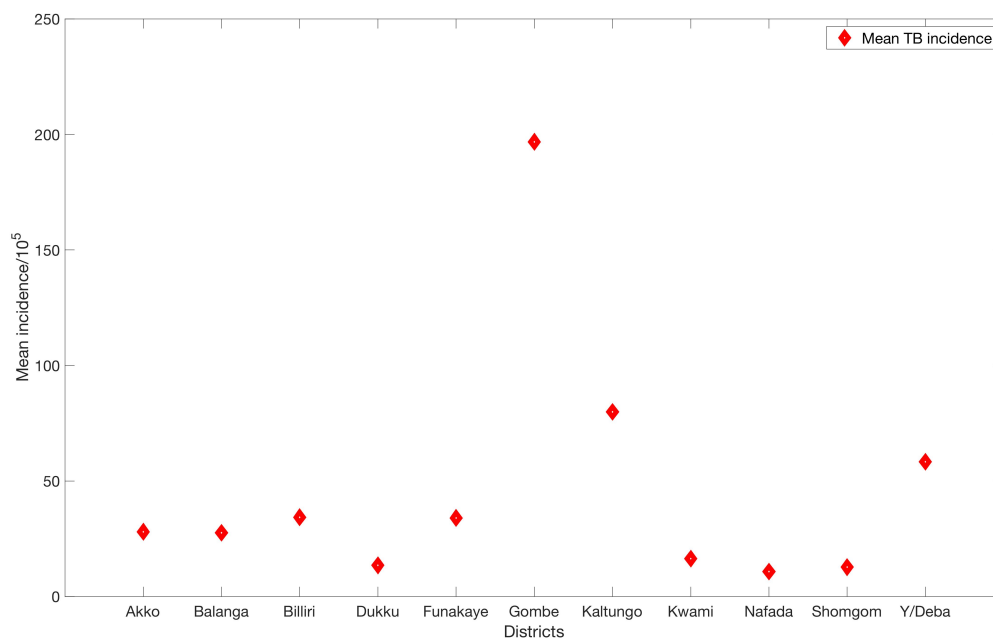


Fig. 4.7 TB incidence: The mean TB incidence/10⁵ reported of each district from 2007 to 2016 in Gombe state Nigeria.

To get a clear insight of how some of the above mentioned items affect the TB dynamics in Gombe state, we carried out a statistical analysis on the result of the interviews conducted (IDI). The results of the analysis is also used in parameter estimation for the development of an Agent-Based Model for tuberculosis spreading in Gombe state that is presented in the next Chapter.

4.3.3 Inferential Statistical Analysis of interviews

The determination of the exact proportion of TB infected and sick people of Gombe state population would require a complete study that would involve monitoring a large part of the population. A study of this magnitude would be very expensive and at the same time Herculean task. Nonetheless, we use the understanding of inferential statistics to make important estimates and conclusions about TB in the population, using the interviews conducted as our representative sample of TB patients in the population.

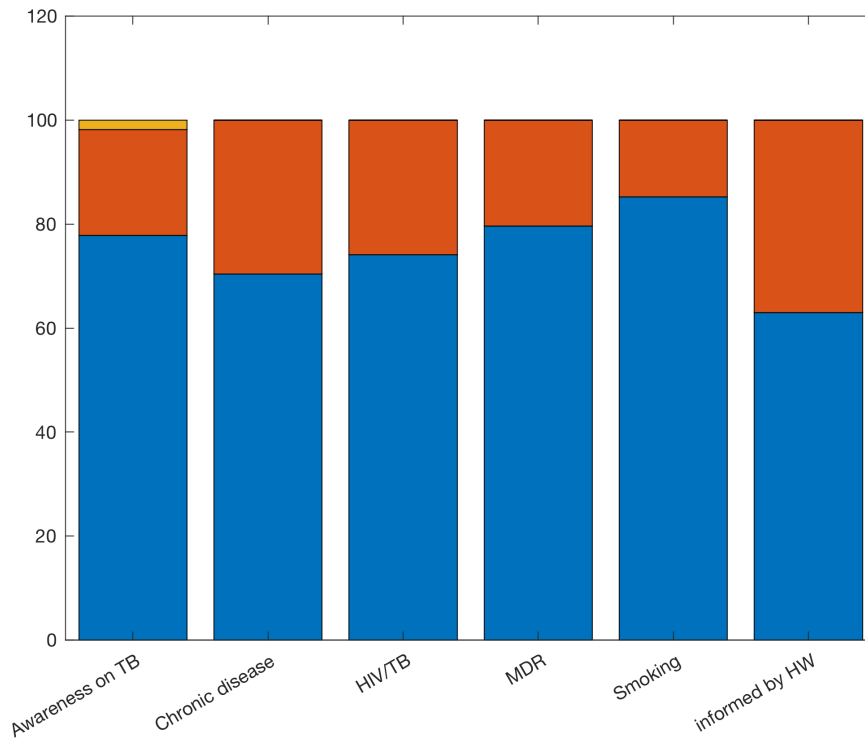


Fig. 4.8 Field work: Some of the result from the interview conducted. Awareness: 2% are fully aware, 20% are partially aware, 78% are un aware. Chronic disease: 29.6% of patient have some type of chronic disease. HIV/TB co-infection: 26% of patient are living with HIV. MDR: 20% of patient have MDR. Smoking: 15% of patient smokes. Informed by health worker(informed by HW): 37% of patient were informed by health workers

Analysis of surface area and population density of districts in Gombe state

As previously mentioned, we have obtained all the reported TB cases from 2007 to 2016, and we have calculated the mean TB incidence for each district per 10^5 populations. We start by identifying the districts with the highest TB incidence.

In an effort to explain and quantify the TB incidence in different districts of Gombe, we carried out an Anova analysis of the mean TB incidence by surface area of each district. Figure 4.9 presents the result of the analysis, with mean TB incidence as a function of surface area between the districts. As it can be observed, while most of the mean incidence base on the reported cases happens to be less than 50 cases per 10^5 populations, 3 district stands out namely Gombe city with 202 cases per 10^5 , followed by Kaltungo with 80 cases per 10^5 and lastly Y/deba with 54 cases 10^5 . Table 4.1 below summarizes the statistical analysis of the relationship between the mean TB cases and the surface area of each district. As seen in the

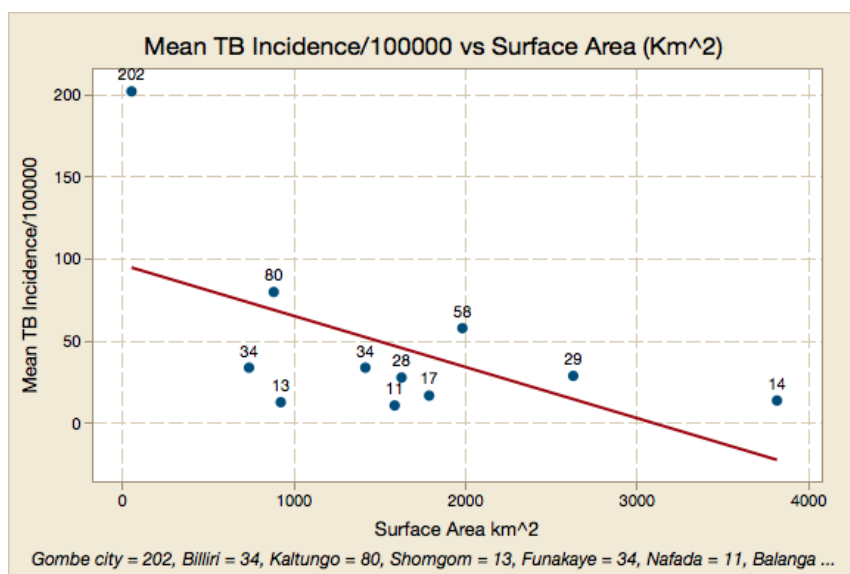


Fig. 4.9 Anova analysis of surface area: Relation between surface area and mean TB incidence of reported cases between 2007 and 2016 in each district.

Table 4.1 Summary result of Anova analysis of TB incidence in relation to surface area

Variable	No of Sample	Mean	Std	Minimum	Maximum
Mean TB Incidence/10 ⁵	11	47.27	55.41	11	202
Surface area (Km ²)	11	1584.5	1014.5	52	3815

table, TB reported incidence seems to decrease with an increase in surface area, with every increase of 1km, the mean TB incidence seems to decrease by 0.03109 per 10⁵. With 95% confidence, the decrease in mean TB incidence will remain between (-0.065 0.0028) per 10⁵ populations. It is also clear that the 32.4% variability in the meant TB incidence is due to the increase in the surface area, we could say that the surface area is quasi significant statistically with a P-Value of 0.0676. The larger the surface area, the less reported cases.

Furthermore, we investigate the relationship between mean TB incidence and population density (because figure 4.9 demonstrates a clear relation between TB incidence and the surface area of the inhabitant). We define population density as the number of inhabitant per Km² of the area. Results are shown in figure 4.10.

Table 4.2 Summary result of Anova analysis of TB incidence in relation to population

R-sq	Equation with 95% Confidence interval for the slope as (-0.064978 0.002773)
32.40%	Mean TB/10 ⁵ = 96.5303 – 0.0310880 * SurfaceArea
P-Value	0.0676

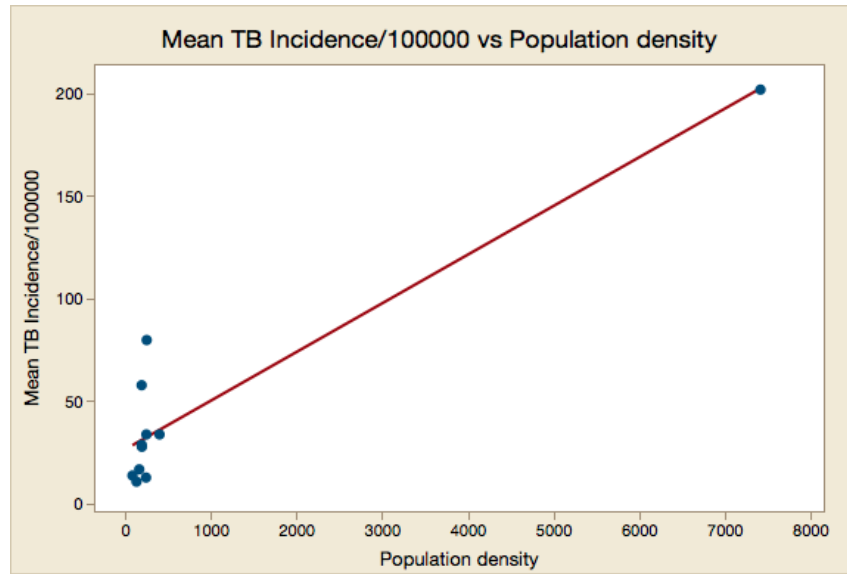


Fig. 4.10 Anova analysis of population density: Relation between population density and mean TB incidence in each district base on reported cases from 2007 to 2016 in Gombe state.

Table 4.3 Summary result of Anova analysis of population density

Variable	No of Sample	Mean	Std	Min	Max
Mean TB Incidence/ 10^5	11	47.27	55.41	11	202
Population density (Km^{-2})	11	585.3	2174.1	78	7409
Analysis of Variance					
Source	DF	Adj SS	Adj MS	F-Value	P-Value
Between groups	1	26583.0	26583.0	58.14	≤ 0.0001
Within groups	9	4115.2	457.2		
Total	10	30698.2			

Figure 4.10 shows the relationship between mean TB incidence and the density per unit Km^{-2} of the inhabitant of each district. Table 4.3 is associated with the figure and gives a summary of the statistical implication of the density of the population in relation to the mean TB incidence in each district. From Figure 4.10, we can say that 3 districts with the highest TB incidence seem to have higher population density as well. With every increase in 1 unit of population density, the mean TB incidence also increases by 0.02372 per 10^5 , this number is huge when converted to actual number of TB cases. Given a 0.05% confident level, we are 95% sure that the increase in the mean TB incidence will remain between (0.016679 and 0.0308) per 10^5 populations if all parameters remain the same. It is also clear from the results above that 86.6% of the variability in TB incidence in the respective district is due to

Table 4.4 Summary result of Anova analysis of TB incidence in relation to population density

R-sq	Equation with 95% Confidence interval for the slope as (0.016679 0.0308)
86.59%	Mean TB/10 ⁵ = 26.9191 + 0.0237146 * <i>populationdensity</i>

population density. The Anova analysis of variance with a P-Value of 0.0133 indicates that the test between mean TB incidence and the population density is statistically significant.

4.3.4 Estimate of some of the most relevant parameters extracted from the interview

Age

One of the most important parameters when investigating tuberculosis in any given population is the age distribution and the distribution of the TB patients across the age groups. We extracted the ages of all patients interviewed and estimated the mean age of TB patients in Gombe state.

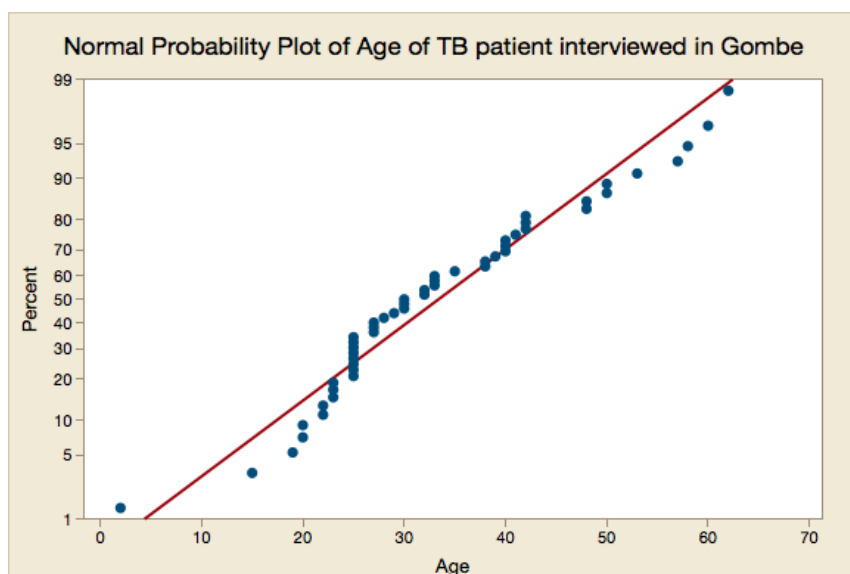


Fig. 4.11 Normality Test: Normal probability distribution test of age of patient interviewed in Gombe state (Nigeria).

From Figure 4.11, the mean age of patient interviewed is 33 with a standard deviation of 12.5, this is very consistent with data analyzes from NTBLCP in Section 4.3.4, which implies the sample is a good representation of the TB population in Gombe. The Anderson Darling test shows that the age of TB patient in the population is not normally distributed, this is

Table 4.5 Age distribution Normality test of patients interviewed

No of Sample	Mean	Std	Mini	Max
52	33.4	12.478	2	62
Anderson's test				
Null hypothesis	H ₀ : Data follow Normal Distr.			
Alternative hypothesis	H ₁ : Data do not follow Normal Distr.			
AD Value	P-Value			
0.97	0.0133			

expected as in most cases, the productive age of 25 to 40 are always the age groups with highest number of TB cases just as seen in section 4.3.4 of the data analyzes from NTBLCP. We estimate that, with a 95% confidence, the age group with the highest TB burden in Gombe state are between 29 to 37 years of age.

Number of dependent

Number of dependents is also considered to be a good marker for understanding the TB dynamics in the population. The type of healthcare policy practice in these areas is called over the counter payment policy where a patient has to pay for his medical expenses over the counter not by insurance policies, these system raise a significant concern as to how much an individual earns and the type of expenses and dependent attached to the income earn.

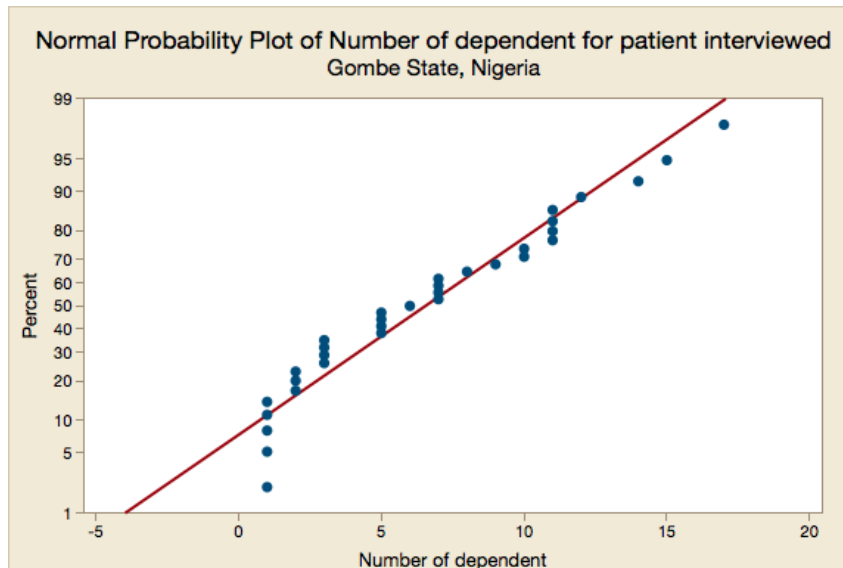


Fig. 4.12 Normality test of number of dependent per patient: Normality probability distribution test for number of dependent per TB patient interviewed in Gombe State Nigeria

Table 4.6 Estimated income distribution of patient interviewed (€)

No of Sample	Mean	Std	Mini	Max
52	66.591	57.012	7.229	289
Anderson's test				
Null hypothesis	H ₀ : Data follow Normal Distr.			
Alternative hypothesis	H ₁ : Data do not follow Normal Distr.			
AD Value	P-Value			
0.97	< 0.005			

From Figure 4.12, the number of dependent is somehow normally distributed. the P-Value for the test is 0.0630 which show quasi significant of the number of dependent. Another look at the data in relation to district shows that people in urban areas has less number of dependent compared with people in rural areas, the analysis is statistically significant with a P-Value of 0.0495 and 11.9% of increase in number of dependent could be explained by whether an individual reside in urban area or rural area. We predicted that, the mean number of dependent will remain between (5 8) with a 95% confidence level, with a standard deviation of 4.5 and mean value of 7.

Estimated income of patient

The estimated income of the patient was also analyzed, the result of this analysis is given in Table 4.6.

We predicted the mean estimated income to be between is (48 85)€ per month with a standard deviation of 57 and mean of 67€ with a 95% confidence level. Anderson's Darling test shows that the estimated income of the population is not normally distributed.

Ratio of patients to hospital in each district of Gombe state

We estimated the ratio of patient to secondary hospitals in each district of Gombe state by dividing the population of that district with the number of secondary facilities in the district. This will give us an idea of overcrowding and efficiency of TB control in each district.

The mean ratio of patients to hospital facility in Gombe State is 183,685:1, with a standard deviation of 99,990. The hospitals are normally distributed across the respective district, so we can assume a fairly similar type of behavior and facilities in the hospitals across all districts except for health workers.

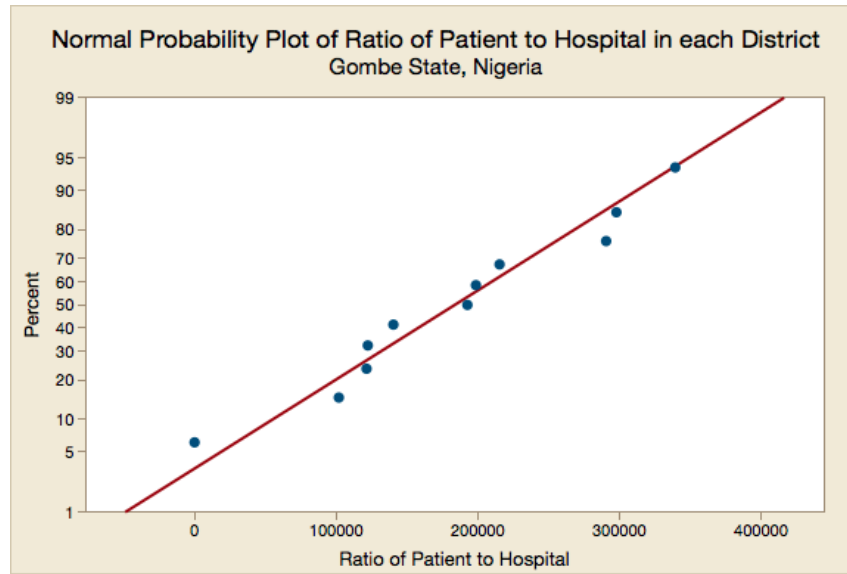


Fig. 4.13 Normality test: Normal distribution test for number of patients to hospital in each district of Gombe state Nigeria

Table 4.7 Ratio of patients to hospital facilities in the districts of Gombe

No of Sample	Mean	Std	Mini	Max
11	183685	99990	0	339407
Anderson's test				
Null hypothesis	H_0 : Data follow Normal Distr.			
Alternative hypothesis	H_1 : Data do not follow Normal Distr.			
AD Value	P-Value			
0.23	0.7321			

Diagnosis delay

The issue of diagnosis delay is one of the most fundamental focal point of these research, the longer the patient remain undiagnosed, the more he/she is likely to transmit the disease to rest of the population. To have a clear picture of what is causing delay in diagnosis of we divide this phenomena into two sub categories. The first is diagnosis delay before going to hospital, also known as *Patient Delay* ("time interval between onset of symptoms suggestive of TB and the patient's first contact with a health care point") [86] . This delay could be related to so many different factors ranging from lack of access to hospital, inadequate knowledge about tuberculosis and its symptoms, poverty due to many dependent of the patient, lack of education and understanding of the hospital system, traditional or religious beliefs, etc, as it has been summarizes from the interviews conducted. The second type of delay is what we

termed as diagnosis delay at the hospital, also known as *Diagnostic Delay* ("time interval between the first consultation with a health care point and diagnosis")[86]. This could be attributed to lack of expert and or training of the hospital staffs, lack or inadequate facilities at the hospital, over crowding of patient, expenses incurred due to treatment etc. There is a third category of delay that is not considered in this study, which is the delay on starting the treatment, also known as *Treatment Delay* ("time interval between diagnosis and initiation of anti-TB treatment"). *Diagnostic Delay* and *Treatment Delay*, together, conform the *Health System Delay*[86]. Finally, we can define *Total Delay* as the sum of the different delays.

Diagnosis delay before hospital (*Patient Delay, PD*)

The diagnosis delay before hospital of the patient interviewed was recorded and analyzed and summarized in table 4.8.

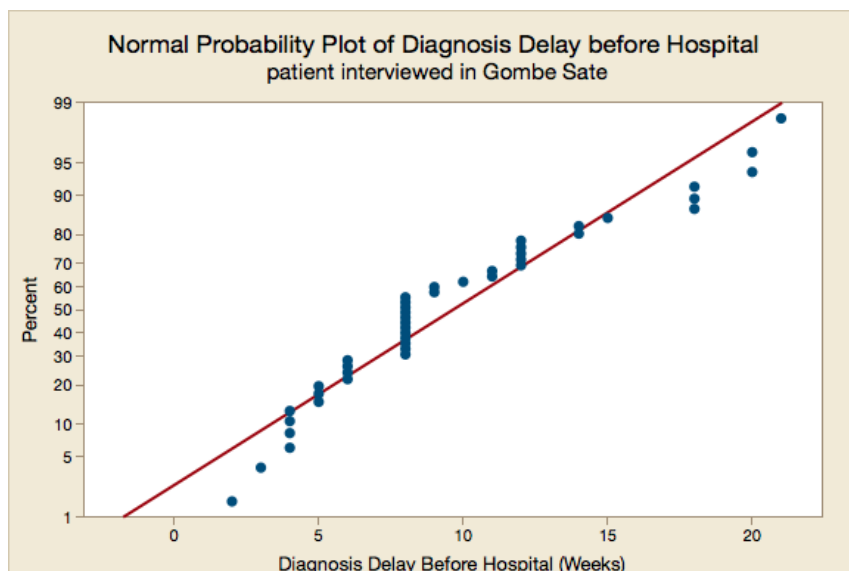


Fig. 4.14 Anova test: One way Anova analysis of PD of patients interviewed in Gombe state (Nigeria).

Figure 4.14 presents the one way Anova analysis for the diagnosis delay before going to the hospital of the patients interviewed, and table 4.8 presents the summary of the analysis. The PD was found to be 9.6 weeks with a standard deviation of 4.8, the Andersons test shows that the delay is not normally distributed, the P-Value of ≤ 0.005 shows a very strong significant of the diagnosis delay is statistically in relation to the TB dynamics. Later, we will analyze how the delay before going to hospital is related with several other factors such as poverty, location or district and Age using multiple Anova regression Analysis.

Table 4.8 Diagnosis delay before going to hospital of patient interviewed in the districts of Gombe, in weeks.

No of Sample	Mean	Std	Mini	Max
52	9.6364	4.8994	2	21
Anderson's test				
Null hypothesis	H_0 : Data follow Normal Distr.			
Alternative hypothesis	H_1 : Data do not follow Normal Distr.			
AD Value	P-Value			
1.33	< 0.005			

Diagnosis delay at the hospital (*Diagnostic Delay, DD*)

The delay at the hospital for all the patients interviewed was also recorded and analyzed.

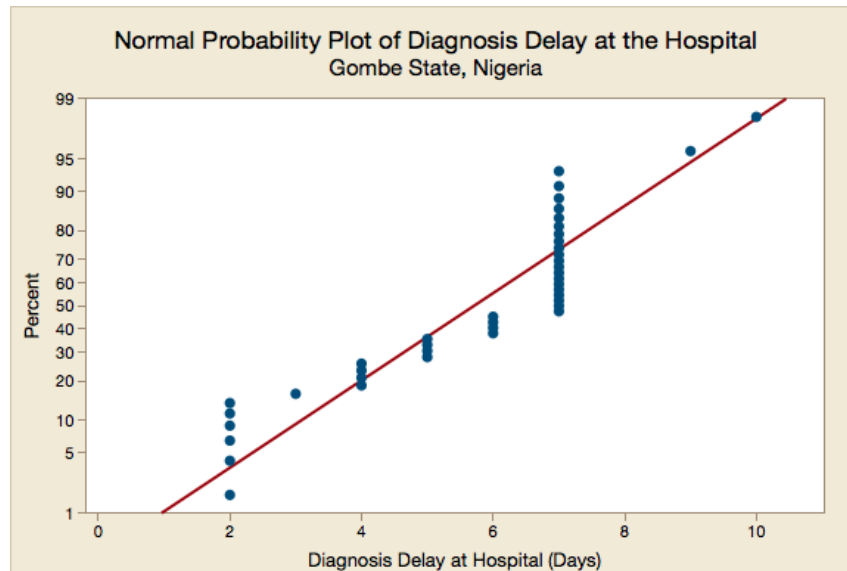


Fig. 4.15 Anova test: One way Anova analysis on diagnosis delay at the hospital of patient interviewed in Gombe Nigeria.

Figure 4.15 and table 4.9 describe and summarize the statistics of diagnosis delay at the hospital for the patient interviewed. The mean DD is 6 days with a standard deviation of 2. The Andersons test shows that DD is not normally distributed across hospitals of the different district in Gombe state. With 95% confidence, we can also infer that the mean delay at the hospital for diagnosis of TB in Gombe state is between 5 and 7 days for a TB patient across all districts.

Table 4.9 Diagnosis delay at the hospital of patient interviewed in the districts of Gombe, in days.

No of Sample	Mean	Std	Min(days)	Max()
52	5.7073	2.0401	2	10
Anderson's test				
Null hypothesis	H ₀ : Data follow Normal Distr.			
Alternative hypothesis	H ₁ : Data do not follow Normal Distr.			
AD Value	P-Value			
3.0	< 0.005			

Table 4.10 Diagnosis delay before hospital in weeks of patient interviewed in the districts of Gombe

No of Sample	Mean	Std	Min(weeks)	Max(weeks)
52	13.7	11.9	2	68
Regression Equation				
Diagnosis delay =	7.891 + 0.005525Area			
R-sq	P-Value			
24.07%	< 0.0002			

4.3.5 Significance and important relationships between some of the most relevant parameters extracted from the interviews

The next important questions to face is if the relationships between some of the parameters in section ?? are significant and important statistically. To answer such an endearing question, we employ the use of different types of statistical analysis such as Machine learning for regression and Decision trees, Anova for analysis of variance between variables and Correlation analysis. Our input variables are both numeric and categorical in nature (age of patient, population density, surface area of each district, number of dependent, estimated income per month of patient, ratio of patients to hospital, level of education, previous contact with TB patient, knowledge about TB symptoms and transmission, smoking, etc.) and we consider our response variables to be the diagnosis delay before going to hospital (PD) and at the hospital (DD).

A one way Anova analysis of the input variables to the response variables reveals that the most important variable that affects the diagnosis delay before hospital is the surface area of the districts. We have already seen that the diagnosis delay before hospital is not normally distributed, it varies from district to district. We hypothesize that the diagnosis delay before hospital depends on the size of the district. Patients from districts with large surface area are more likely to have a longer time before they visit hospital.

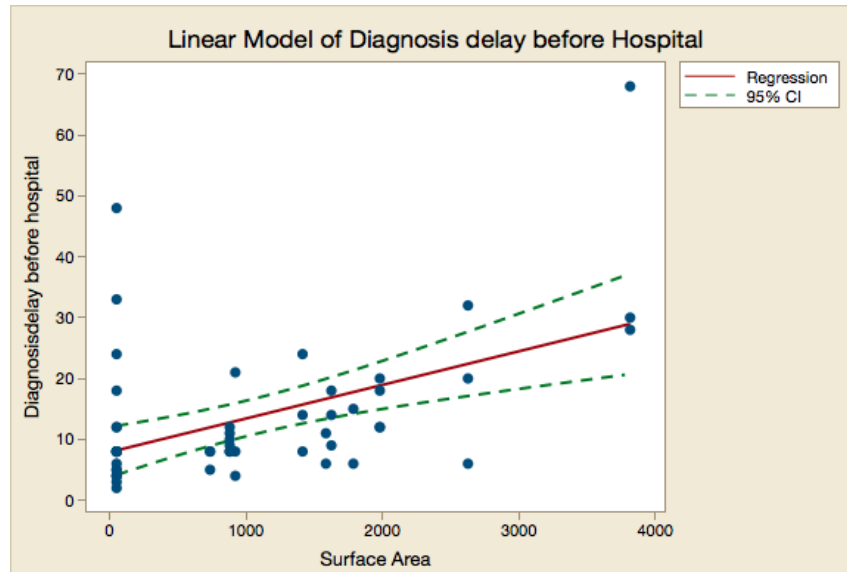


Fig. 4.16 Anova Analysis: One way Anova analysis on diagnosis delay before hospital and surface area of district in Gombe Nigeria.

Figure 4.18 and table 4.10 give evidence to support our hypothesis. 24.1% of the variation in diagnosis delay before hospital is due to an increase in surface area of the district. The P-value from the Anova analysis of variance supports the hypothesis that there is strong evidence that diagnosis delay before hospital is related with an increase in surface area of the district. For every 1Km^2 increase in surface area, the delay before hospital also increase by 1:08 hours (0.006 weeks). Table 4.11 gives both the Pearson correlation summary of all the variables involved in the model and there corresponding statistical significance based on their P-Values.

Figure 4.17(a) shows that patients with previous knowledge of infection have a lower diagnosis delay of about 6 weeks than patients without previous knowledge on TB infection, with a mean delay of about 25 weeks. Also, patients who have a previous contact with other TB patient seems to have higher delay time. A closer look on the interaction between previous knowledge of TB infection, previous contact with TB patient and diagnosis delay before Hospital (in Figure 4.17(b)) reveals that patients who are not knowledgeable about TB and have previous contact with other TB patient have the highest delay time, while patients with previous knowledge of TB and have no previous contact with TB patient have the lowest of delay time of 5 weeks.

Table 4.11 Pearson correlation coefficient for model variables (P-values are given under the respective correlation values in red)

	Age	No of dependent	Estimated income	Patient: hospital	Population density	Surface Area	Delay at hospital
No of dependent	0.4884 0.0002						
Estimated income	0.2536 0.0697	0.0917 0.5182					
Patient: hospital	-0.0217 0.8788	0.0539 0.7039	0.0999 0.4811				
Population density	0.0157 0.9122	-0.2361 0.0920	0.2005 0.1542	0.0527 0.7108			
Surface area	0.1255 0.3752	0.2324 0.0973	-0.2138 0.1281	0.0395 0.7809	-0.7575 < 0.0001		
Delay at hospital	-0.1476 0.2964	0.0827 0.5600	0.0135 0.9244	-0.0231 0.8711	0.1223 0.3877	-0.1011 0.4755	
Delay before hospital	0.1420 0.3151	0.2105 0.1341	0.03518 0.8047	0.15416 0.2755	-0.1589 0.2603	0.4906 0.0002	0.3077 0.0265
Cell content	Pearson's P-values						

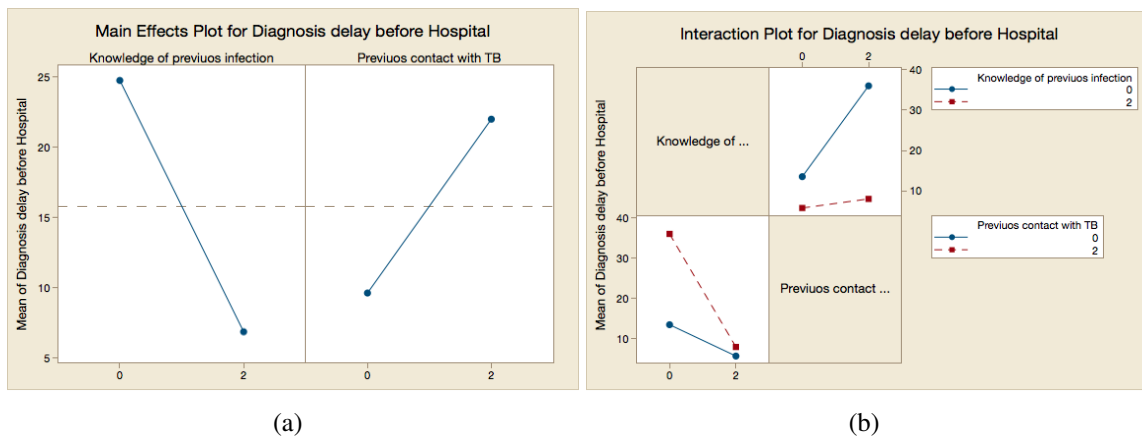


Fig. 4.17 Multiple regression Tree: Interaction plots (a) Interaction plot that signifies the individual effect of previous knowledge on TB and previous contact with TB patient on diagnosis delay before hospital. (b) Multiple interaction plot that shows the mutual effect of both previous knowledge on TB and previous contact with TB patient.

Distinction in diagnosis delay before hospital (PD) and diagnosis delay at hospital (DD) for high burden areas vs low burden areas in Gombe state

We have established in section 4.3.5 that diagnosis delay both at the hospital and before the hospital are the two most relevant and important parameters when it comes to TB dynamics

Table 4.12 Test for difference in diagnosis delay at hospital between areas with high reported incidence and areas low reported incidence in Gombe state

Null hypothesis	$H_0: \mu_1 - \mu_2 = 0$	
Alternative hypothesis	$H_1: \mu_1 - \mu_2 \neq 0$	
T Value	DF-Value	P-Vlaues
1.51	50	0.1375

Table 4.13 Test for difference in diagnosis delay before hospital between areas with high incidence and areas low incidence in Gombe stated

Null hypothesis	$H_0: \mu_1 - \mu_2 = 0$	
Alternative hypothesis	$H_1: \mu_1 - \mu_2 \neq 0$	
T Value	DF-Value	P-Values
-1.39	52	0.1728

in Gombe state. It is also clear that the most important factor that facilitates such delay is the surface area of the respective district and the population density. To this end, a distinction between districts can be made according to the reported incidence: those with a high mean TB incidence (reported cases) namely Gombe city, Kaltungo and Y/deba, with incidence between 54-202/10⁵, and those districts with low mean TB reported incidence between 0-49/10⁵s. In order to understand the cause of this distinction, we divided the data of the interviewed patient into two categories according to reported incidence, the high burden districts (HBD) and low burden districts (LBD). We first start by checking the diagnosis delay at the hospital between the LBD and the HBD, and we carry out a two way sample Anova test for equal mean.

First we had to test for equal variances between the high incidence and low incidence areas using Leven test, and equal variance was adopted for the result shown in table 4.12. The table shows that there is no statistically significant difference in diagnosis delay at the hospital for both HBD and LBD. This implies that the quality of service at the hospital across the district is the same. We might say that the delay at the hospital might not necessarily be associated to the quality of service at the different hospitals.

Now we look at the diagnosis delay before hospital in both high burden districts and low burden districts.

We also start by testing for equal variance in both high incidence and low incidence areas, where equal variance was also assumed for the result of the test in table 4.13. The table shows no statistically significant differences between the mean diagnosis delay before hospital of high incidence and low incidence areas. It is clear from this summary that delay before going to the hospital does not depend on the TB incidence but rather on the surface

area of the district. The result is also showing that both areas with high reported incidence and areas with low reported incidence seem to have similar type of delay before hospital. We can also infer from this result that low incidence areas (namely Akko, Balanga, Billiri, Dukku, Funakaye, Kwami, Nafada, Shongom) does not necessarily have low TB incidence but rather lower reported cases than the high incidence areas (namely Gombe, Kaltungo and Y/deba).

4.3.6 Multiple regression analysis

Still on the quest to identify the parameters that lead to TB diagnosis delay both at the hospital and before going to the hospital in Gombe State, we did a multiple regression analysis with all the parameters mentioned in section 4.3.4, considering only diagnosis delay at hospital and diagnosis delay before hospital as our response variables.

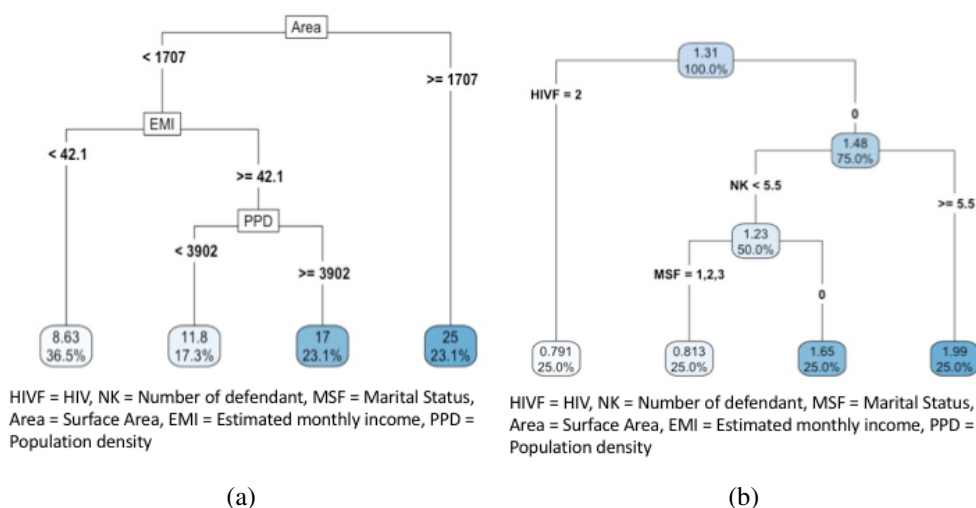


Fig. 4.18 Multiple regression tree: (a) Result of multiple regression analysis of results from the interview with diagnosis delay before hospital as a response variable. (b) Result of multiple regression analysis of results from the interview with diagnosis delay at the hospital as a response variable.

Figure 4.18(a) present the result for diagnosis delay before going to hospital as a response variable, while Figure 4.18(b) presents the result of the multiple regression analysis for the diagnosis delay at the hospital as a response variable. The diagnosis delay before hospital (Figure 4.18(a)) seems to be mostly associated with surface area of the district, estimated income of patient and population density, more than the rest of the parameters from the interview. From the regression tree, the most important of the 3 parameters mentioned is the surface area of the district. For patients that live in district with surface area bigger than

1707 Km^2 , their mean delay time before going to the hospital is 25 weeks. Patients that live in a district with surface area less than 1707 km^2 with estimated income of less than 43€ per month have a mean delay time of 9 weeks before going to the hospital, while patients that live above 43€ per month in the area with a population density higher than 3902 Km^{-2} have a mean delay time of 17 weeks. Otherwise, the mean PD of patients is 12 weeks.

The results also show that there are 3 important variables from the interviews that are mostly related with the diagnosis delay at the hospital. Starting from the root of the tree in Figure 4.18(b), HIV status is the most important variable; HIV positive individuals average diagnosis delay at the hospital is 0.791 weeks which is approximately 5 days. For HIV negative individuals, the next important parameter is the number of dependents of the patient, patient with more than 5 dependents seems to have an approximate DD of 14 days, while patient with less than 5 dependents and either married, divorce or widow have an approximate delay time of 6 days. Single patients have an approximate DD of 12 days.

4.4 Discussion

[63] Two types of data sets have been analyzed statistically: data from NTBLCP, and data from in-depth interview of TB patients in Gombe state. The analysis of data from NTBLCP highlighted information about some of the major characteristics of TB such as distribution of the disease amongst gender and age group of the inhabitants in the state. We have shown that male are more vulnerable to TB and the age group of 25-44 contributed 48% of the total reported TB cases. These data also indicate that the age group of 55+ has a similar TB incidence/10⁵ population to the most active age group of the society. HIV prevalence is high amongst the TB patient, while smoking, alcoholism and other types of risk factors are low with a total smear positive cases of about 32%. In the 11 district of Gombe state, reported TB cases amongst the district is not uniform, Gombe city has the highest mean incidence of 202/10⁵ population, followed by Kaltungo with incidence of 80/10⁵ and Y/deba with incidence of 54/10⁵. The data from the field work highlighted socioeconomic properties that allowed the propagation of TB in Gombe state. We found that 0% of the patient interviewed understand TB symptoms before they were sick, only 2% of the patient understand TB control and little or no information is communicated to the patient by the healthcare workers. Evidence of Bovine TB is on the high side which most of the healthcare workers does not know about, a very high percentage of HIV⁺ (26%) amongst the patient, and 20.4 % are MDR. Some of the obvious challenges facing these communities include but not limited to

- Lack of access to the health care facilities

- Present of alternative medicine which most of the patient admit trying before going to hospital
- Inaccurate diagnosis at the smaller clinics
- Poverty and lack of resources to access health care
- Constant visit by the patient to untrained health workers that have smaller rooms in the society operating as alternative to hospitals
- Ignorance and unawareness about the disease

The analysis also indicates that one of the major problems that facilitate the spread of TB is longer delay time before going to the hospital, which is related to the size of the respective district. The larger the district, the less likely for the patient to go to hospital on time. The population density of the respective district plays a role, in particular in relation to the delay time and inaccurate diagnosis in the district with a large population density. Number of dependent also happens to be another factor that contributed to the delay time which varies from urban areas to rural areas. Hospitals in all of the district are normally distributed with similar facilities except for Gombe city. The discrepancies in the mean TB incidence of each district may not be as a result of lower TB cases but as a result lower reported TB cases (lower notification in some areas lower than the others). Furthermore, a multiple regression analysis shows that the most important factor that affects the diagnosis delay of the patient before going to hospital is the surface area of the district, followed by the estimated income of the patient and the population density of the respective district.

Chapter 5

Using Agent-based simulations to test interventions in Gombe

5.1 Introduction

Some of the main findings from the previous chapter include the relationship between the diagnosis delay time and surface area of the district, the delay time and population density, and the distribution of the delay time among the interviewed patient. The very low notification of TB in this community makes it necessary to incorporate changes in the TB control program and hence the re-evaluation of the intervention strategies. An ABM is appropriate in this task due to its discrete nature and possibility to take individual diversity into consideration, whereas SEIR approaches can model the mean dynamics of the TB at higher level. Hence, in this chapter, we tackle the problem of TB propagation from an ABM approach. We adapt and implement an ABM for the dynamics of TB in the population of Gombe state, and we test its outcomes by means of virtual experiments.

The complex nature and particular characteristics of ABMs makes it difficult to provide a comprehensible description of the model that allows it to be replicated, which is a key to science. These prompt some experienced ABM modellers like [32] and [33] to develop a standard ODD protocol for ABMs. The ODD stands for Overview, Design concepts, and Details. This methodology is used in this chapter for describing the implemented model. The parameters used in the simulations are a combination of data extracted from NTBLCP, the field work in Gombe and literature. The simulator is created using C programming language, we simulated TB dynamics in the population of Gombe and data from WHO is used to validate the simulation. The results and conclusions of the simulations are used to further

design intervention strategies that could be adopted or tested in containing TB epidemics in Gombe state.

5.2 ODD description of the model

5.2.1 Overview

The objective of this ABM is to analyze the evolution of pulmonary TB incidence in a community and test interventions. It is fitted to Gombe state area of north east Nigeria, considering the population to be constant (this is because the available data from WHO used for fitting the model could only support the simulation in a constant population), and the possible effects of epidemiology control strategies and public health decisions are checked through virtual experiments.

5.2.2 Entities, state variables, and scales

The fundamental entities of the model are persons. They are considered to undergo through five infection states: healthy, infected, sick, under treatment, and recovered. Persons of any infection state different from healthy are simulated as agents, while healthy population is a property of the local space. Total population remains constant through out the simulation time, as well as the characteristics of the healthy population such as age distribution, male to female ratio and HIV⁺ percentage, among others. The healthy collective is much larger than the sick, infected, under treatment or recovered collectives. Hence, it is not necessary to control the healthy individuals one at a time, they are regarded to be property of the space. Once a healthy person is infected, then he/she will acquire individuality. This strategy is an integral and important optimization process to drastically reduce the computing time and the cost of the model. The model was previously tested against those models that considered healthy people as individuals [52].

The state variables of the agents mainly refer to their status in the TB infection cycle as well as the time spent in such phases and individual diagnostic time when getting sick. Other individual state variables and parameters are age, risk factors (e.g., diabetes, possible HIV infection, and other chronic diseases). Smear-positivity or negativity of infectious agents is also considered. A state diagram of the model is presented in figure 5.1. The total population of Gombe is about 3.4 million inhabitant in 2017, but the simulation is carried out per 100,000 population.

The model is partially spatially explicit, space is considered but it is not explicit, i.e., it does not mimic the exact real space of Gombe State. Simulation occurs in a discrete area of

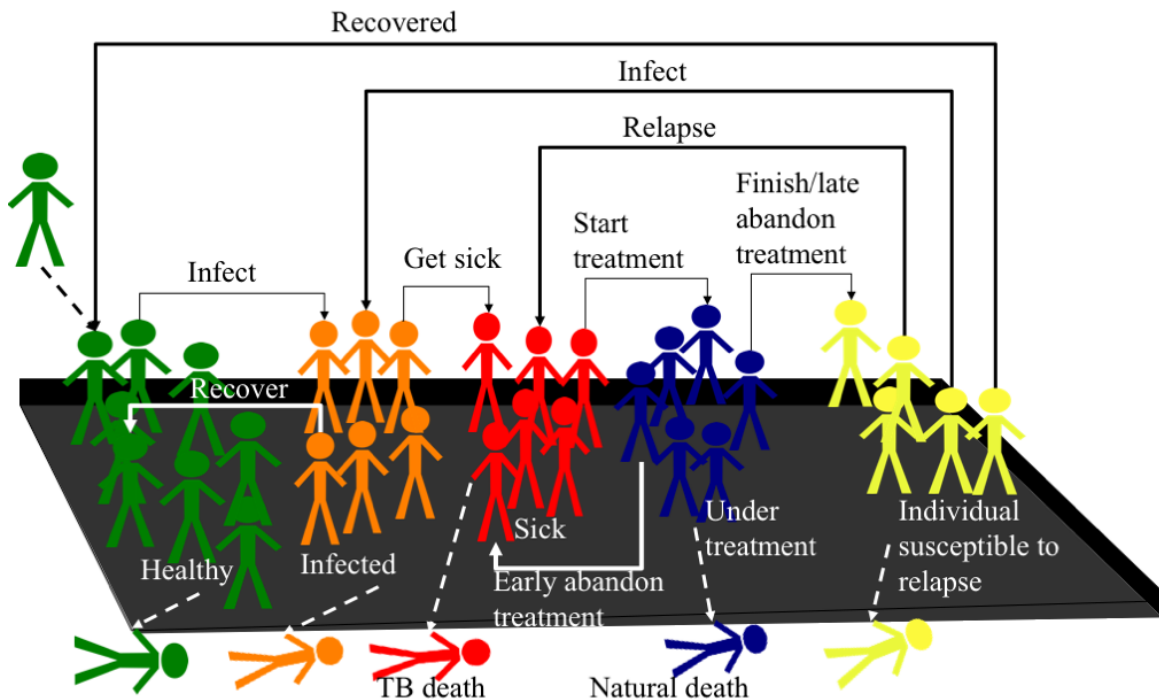


Fig. 5.1 State diagram of the model, where the five states of individuals and their possible transitions are shown. Output whit dotted arrows refer to deaths; input dotted arrows are the corresponding entrances of randomly selected individuals in order to keep population constant.

501 × 501 spatial cells. Each spatial cell represents a local abstract space where persons can interact and transmission of the bacilli can occur with certain probability in a time step. The time step is set to 1 day, and the simulation may cover up to a period of 1 year or more.

5.2.3 Process overview and scheduling

Our model was built in on C programming language, which is well suited for modeling a broad range of agent-based systems. The simulation starts with the set-up of the initial configuration, where the population is randomly generated according to the input distributions of parameters and randomly distributed in the 501 × 501 grid. The model assumes discrete time steps of 1 day, as mentioned. Each day, all individuals execute a series of actions, and their variables are updated immediately.

The individual actions may be: to increase age, to move, to get infected, to get sick, to be diagnosed and start a treatment, to abandon or finish the treatment, to recover, and to die. Some of the actions take place daily for all the individuals in the system (e.g., aging and

movement) and the other procedures are daily evaluated when necessary (e.g., the possibility of a sick individual to be diagnosed is daily assessed). When an individual dies, a new person is introduced with particular random characteristics according to the initial distribution of individual parameters, since general population heterogeneity is assumed to remain constant during the simulation. At the end of each time, step global variables are updated. figure 5.2 and 5.3 shows the flow diagram of the computational model.

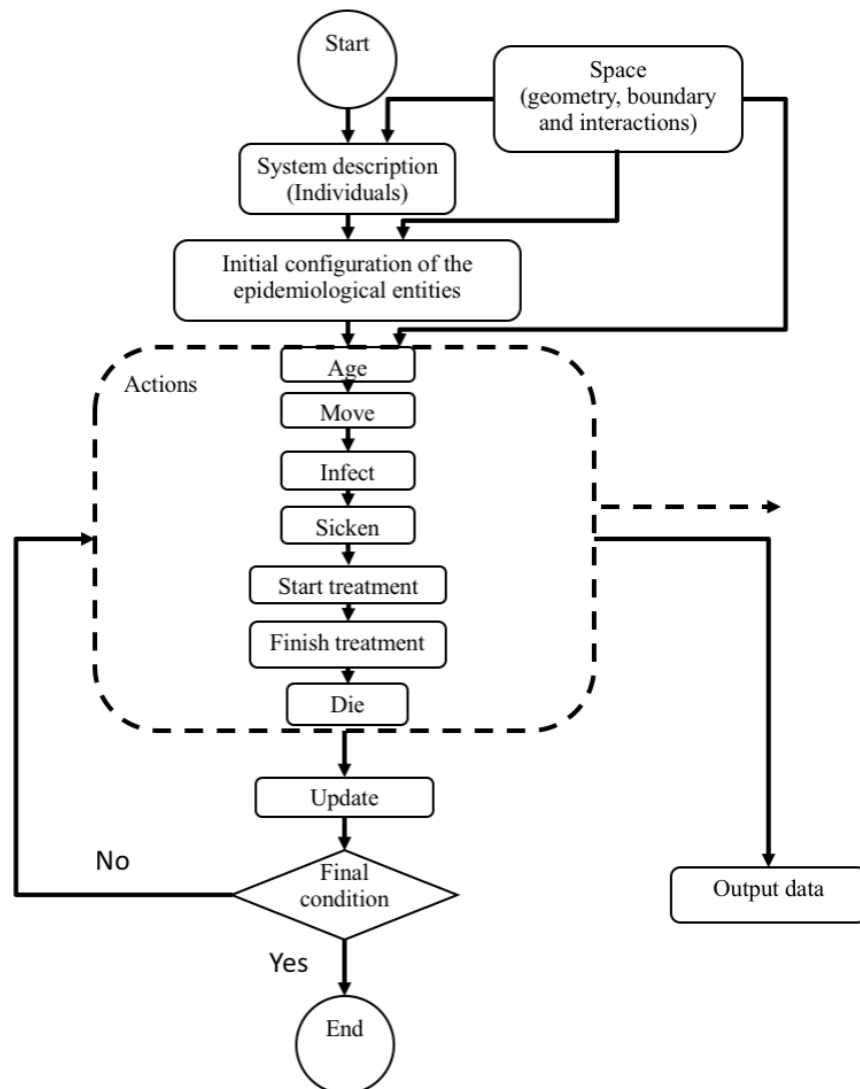


Fig. 5.2 Flow diagram of the model, with all the involved processes and sub-processes. step within the dotted box contained actions that required some certain conditions to be fulfilled, some of these steps are further detailed in Figure 5.3

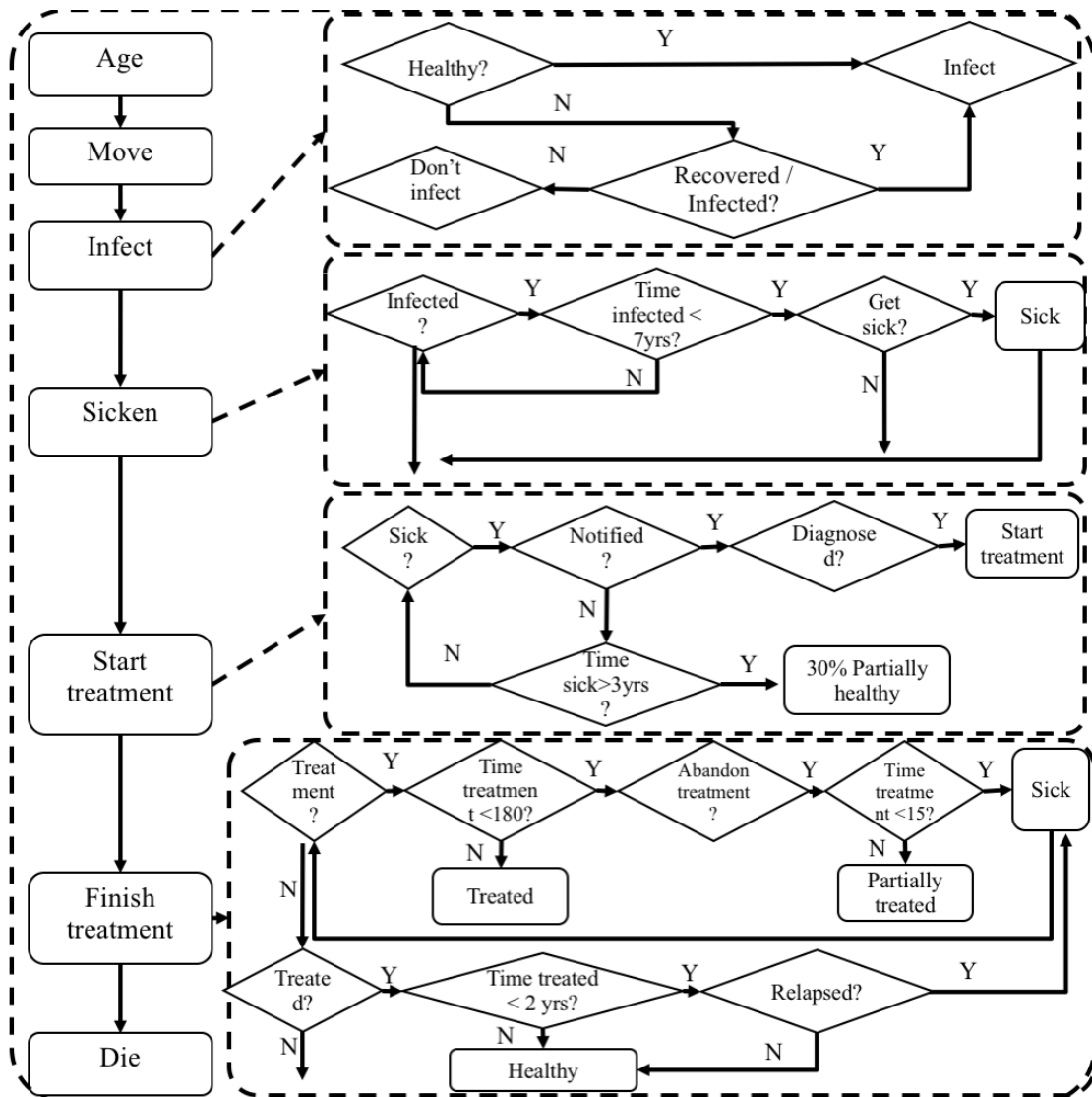


Fig. 5.3 Some more iterative steps from the flow chart in Figure 5.2 to explain some of the most complex functions.

5.2.4 Design concept

Basic principles

The model is based on general knowledge and understanding of the natural history of TB and TB transmission in a population. There are two essential characteristics of TB that must be taken into account in any epidemiological model:

1. An infected individual does not necessarily develop an active disease; on average, only 10% of infected people develop active disease. A person can remain infected for a very

long period of time and may develop active TB after several years, but the probability of developing the disease decreases with time [29, 18]. Infected people are usually not diagnosed, but they can be detected in a screening.

2. Only TB sick can spread the infection. The infection rate increases if the patient is smear positive. Once a TB sick is diagnosed, it is assumed that after two weeks of treatment, the patient could not be able to spread the disease.

The pharmaceutical treatment takes 6 months (World Health Organization [67]). Once the treatment is finished, the possibility of getting sick again because of a TB reactivation remains at 1% for 2 years. Moreover, in some experiments a treatment for detected infected individuals can be included. This treatment is longer than the one given to persons with active TB. It usually lasts 9 months and it is administered to infected individuals to prevent the development of an active disease once they have been detected during a screening exercise (World Health Organization [82]). There is also a probability of relapse for the infected state that is calculated similar to the first treatment.

Emergence

Emerging phenomena are mainly related to long-term dynamics of the infection at the population level. On one hand, only non-treated people with active TB can spread the disease as explained by [?]. Therefore, time before diagnosis is an essential component for the prevalence of the disease. On the other hand, infected persons may develop active TB even a few years after the infection. Therefore, global consequences of particular event at a particular moment may be detected some years later.

Interaction

Local interactions between individuals are explicitly modeled and are one of the most crucial component for the dynamics of the system. They refer to the contact of two persons favored by the spatial proximity between them and the possibility that one of those individuals with an active TB may infect the other person.

Stochasticity

Randomness is introduced at all levels of the simulation. The initial distribution of individual properties is randomly executed according to input distributions from the experimental data gathered. Movement is assumed to be random. Each action is associated with a certain probability and thus executed according to a random probability.

Observation

Output data shows the daily evolution of healthy people, infected people, sick people, people under treatment, and persons already treated. We also record the total number of new sick in a year, total diagnosed in a year and number of death due to TB in that year. Data are recorded annually and exported to an external CSV data file, where they are further analyzed.

5.2.5 Details

Initialization

The user can change some initial conditions at the beginning of the simulation in order to suit the purpose of the experiment. For this particular study, most of the input parameters are taken from official reports. The initial population is fixed at 10^5 individuals. All percentages shown in Table 5.1 are used for calculating the configuration of initial population: rates of sick, under treatment and recovered individuals per 100,000 inhabitants; mean diagnosis delay (from the fieldwork carried out); mean treatment abandon rate; notification rate; death rate due to TB and death rate in children and general population; age distribution in the population; age distribution in the initial population of the sick; gender distribution both in general and sick population; smear positive; individuals with risk factors and with HIV infection. Some other initial variables are assigned randomly: time spent in the infection state-assigned; spontaneous recovery time of the unnotified sick; etc.

Submodels

- Age: All individuals increase their age by 1 day each time step.
- Move: All persons can move randomly through the surrounding local space, once a day.
- Get infected: If there is a certain number of individuals susceptible to TB (healthy and treated) different from zero in the proximity of a sick individual, meaning one of the nine-neighboring spatial cells, including the one occupied by the sick person, this sick agent can infect one of them with a certain probability. The total of susceptible neighboring individuals is computed and then the infection process is repeated as many times as healthy and treated people are found. The infection probability depends on the type of TB disease that the sick person has, either smear positive or smear negative. A smear positive is considered to double the infection probability. The value of the infection probability is closely linked to the spatial and temporal scales, i.e.,

Table 5.1 Official data of Gombe state (Nigeria). Values used in simulations of the model. (*) Percentages with respect to the total number of TB sick people. (¶) Percentages with respect to TB affected people.

Gombe state	Value	Unit	Ref
Total population	3,257,000	persons	[1]
Infected population ¶			
Male	52.6	%	[92]
Age distribution			[92]
0-4	17.05	%	
1-14	28.65	%	
15-24	16.05	%	
25-34	13.50	%	
35-44	9.45	%	
45-54	6.75	%	
55-64	4.55	%	
65-74	2.50	%	
75-84	1.05	%	
85+	0.45	%	
Sick population *			
Male	60.1	%	NTBLCP data
Age distribution			NTBLCP data
0-4	3.8	%	
1-14	4.9	%	
15-24	12.33	%	
25-34	26.34	%	
35-44	21.45	%	
45-54	13.76	%	
55-64	9.29	%	
65-74	8.04	%	
75-84	8.04	%	
85+	8.04	%	
Daily mortality (natural)			
0-5 years	$3.69 \cdot 10^{-4}$	%	[96]
6-15 years	$7.01 \cdot 10^{-5}$	%	[95]
16+ years	$3.269 \cdot 10^{-5}$	%	[94]
Daily mortality (TB induced)			
Notified cases	$4.87 \cdot 10^{-4}$	%	[7]
Unnotified cases	$1.82 \cdot 10^{-3}$	%	Calibrated

the probability of infection is inseparable from the spatiotemporal scale. A change in any of these scales entails the revision of its value. Therefore, it is not a real infection

Table 5.2 Continuation of table 5.1 Values used in simulations of the model.(*) Percentages with respect to the total number of TB sick people.(¶) Percentages with respect to TB affected people.

Gombe state	Value	Unit	Ref
Diagnosis delay (days)			Field work
0-19	0	%	
20-49	0.76	%	
50-79	0.96	%	
80-109	0.45	%	
110-139	0.32	%	
140-170	0.19	%	
171-200	0.13	%	
201-230	0.13	%	
231-260	0.13	%	
261+	0.32	%	
Smear positive ¶	30	%	[61]
Treatment abandon rate ¶	15	%	[6]
Notification rate ¶	16	%	[6]
HIV ⁺ rate	4.9	%	[92]

probability when a sick individual meets a healthy person, but an effective infection probability given the particular spatio-temporal constraints. In this case (501 × 501 spatial cells and 100,000 individuals), the value of this probability was fixed at 0.76% daily for the smear positive 5.1. Once a person is infected, a newly infected individual is created with characteristics randomly assigned according to the data from Table 5.1. Assignment of gender, age group, HIV status, smear status and all other characteristics entirely depend on the initial official data. The infection time of the new individual is set at zero and starts increasing with each time step.

- Get reinfected: The process of reinfection is also considered, it is similar to the process of infection, but the individuals that qualifies for reinfection are the infected and the recovered. If there is a number of infected and or recovered different from zero in the proximity of the sick individual, the sick person may infect one of them with a certain probability. The details of this procedure are analogous to the Get infected ones.
- Get sick: Once infected, the individual may develop active TB according to a particular annual probability that decreases with infection time during the 7 years post-infection [29, 18, 84]. It is assumed that the infection lasts for a very long period of time, though after the first seven years of infection, an individual is evaluated daily with a given probability [7] for a spontaneous elimination of the disease. An immune factor of 0.1

is introduced in the case of Gombe population, this is due to excessive exposure of the population to the *Mycobacterium tuberculosis* obviously because of the very high prevalence of the disease, hence allowing some of the susceptible population to develop a spontaneous defense mechanism against the disease 5.3. For immunodeficient people, such as HIV⁺, a certain factor is multiplied by this probability to increase the chances of sickening. The same happens if there are other risk factors. The chance of becoming a TB sick individual is evaluated at each time step for all infected persons. Globally, the average of 10% of infected developing an active disease is satisfied. The possibility of relapse (getting sick again) for recovered patients is also evaluated daily according to the individual relapse probability (see below). Once a person gets sick, a notification property will be assigned to him according to the input notification rate, and the disease time counter starts running. After the first 19 days, the system will check if the person is notified and then evaluate the chance of diagnosis based on the distribution in table 5.1. Otherwise, he will remain unnotified until after 3 years of sickness, when the system will then check to see if he gets cured spontaneously.

- Get diagnosed and start treatment: if an individual has been assigned the chance to be diagnosed (notification rate), this chance is daily evaluated according to the diagnosis probabilities in table 5.1. Once diagnosed, medical treatment is assumed to start and to stop TB spreading after the first fourteen days of treatment. Individual's time under treatment is initially set to zero and then updated at each time step.
- Abandon treatment: There is a certain probability that an individual abandons the treatment before finishing it. This possibility is evaluated daily for each patient under treatment, according to the input abandonment probability. If a person leaves treatment during the initial 15 days post-diagnosis, he/she becomes ill again. If he/she abandons the treatment after 15–180 days post-diagnosis, the model will consider him/her to be recovered but with a certain probability of relapse during the following 2 years. This probability is assumed to decrease linearly from the 100% of a 15-day abandonment to the 1% of the 180-day treatment period.
- Recover: When a sick individual is diagnosed and treated for 180 days, he/she becomes recovered and a relapse probability of 1% is assigned (the chance of getting sick again during the following 2 years). After 2 years, the individual is considered to be healthy again.
- Die: Each individual has a certain probability of dying according to his/her age. These probabilities are fixed using demographic data from Gombe [2–4]. Accordingly, the

daily dying probabilities are considered to be $3.69 \cdot 10^{-4} \%$ for individuals under 5, $7.019 \cdot 10^{-5} \%$ for individuals between 6 and 15, and $3.269 \cdot 10^{-5} \%$ for individuals over 15, which is a simplification of the real mortality distribution. Furthermore, TB induced death probability for notified sick persons is given as $4.87 \cdot 10^{-4}$ [7] while for unnotified sick persons is calibrated to be $1.82 \cdot 10^{-3}$. This probability is evaluated daily for each sick individual, taking into account that 40% of non-treated TB sick may die in 5 years. Each time an individual dies, a new individual is introduced into the simulation world with the aim of maintaining a constant population. The individual's characteristics are fixed according to the distribution of the initial population as in table 5.1.

5.3 Model implementation and results

5.3.1 C implementation of the model

In this section, the features of the C implementation will be described. The reader is assumed to be familiar with the C language. For further information, refer to the source code.

Details

- **Space:** The space is considered to be a 2D grid represented by a 2D array. Although each individual can be in one of the points of the grid, only the healthy ones are considered as values in the array (i.e. a 3 in the array means that 3 healthy individuals are at that particular point). For the rest of the states, the position is a property of the individual (explained further in the following items).
- **Individuals:** All individuals but healthy are considered separately. That is, the individuals of each state are represented by different structures which contain different properties. These structures are implemented with structs. Some common characteristics for the non-healthy individuals are age (in days), gender, and position on the grid (i.e. an x and a y value). The particular properties of each state are as follows:
 - Characteristics of infected individuals: time infected (in days), HIV status, immune factor, and reinfection factor.
 - Characteristics of Sick individuals: time sick (in days), diagnosis time (in days), smear status, notification factor.

- Characteristics of individuals under treatment: time under treatment (in days) and smear status.
- Characteristics of treated individuals: time spent under treatment (in days), time since finishing the treatment (in days), smear status, probability of relapse before being considered healthy.

File configuration

For easier comprehension of the simulator, the source code was split into different files. This allows to classify these files based on their respective functions and thus makes it easier to debug, test and maintain the code. We give a detailed description of the different files:

- **IBMheader.h**: this is the header file that contains all declarations of functions and global variables of the simulator.
- **IBMparams.c**: the file contains all the global variables and parameters used in the simulator. This file is particularly important because it allows the change of any parameter easily without modifying the source code. Some of the parameters have been introduced in table 5.1, the rest of the parameters can be found in this file.
- **IBMsetup.c**: the file contains the definition of the functions needed for the initialization of the simulation, i.e., the functions that generate the individuals in each state according to the initial setup. It also initializes and generates the structure where the individuals are embedded in.
- **IBMdynamics.c**: the file contains the definition of the functions needed for the evolution of the simulation, i.e., the functions that represent the dynamics of the model such as infection, movement, diagnosis, etc.
- **main.c**: the main file of the program. This contains the function main and the flow of the simulation, it is where the functions in the two previous files are called.
- **random functions.c**: this contains the functions related with (pseudo)random numbers, from the generation of a pseudorandom number according to a certain distribution to the choice of an element between a collection with different probabilities.
- **list.h**: header file that contains the declaration of functions of the file list.c.
- **list.c**: file that contains the definition of the functions of the doubly linked lists used. This implementation is independent of the IBM model and could be used in other programs.

Assuming that one wishes to compile with gcc compiler (a commonly used C compiler), the line needed to compile the program from these files is: `gcc main.c IBMparams.c IBMsetup.c IBMdynamics.c randomfunctions.c list.c..`

5.3.2 Model validation

Once the simulator was ready, it was tested to see if it presents the same dynamics as the previous simulator [79] and mimics the real behavior of the system as well as which improvements, if any, are present with respect to the previous version [79].

In order to evaluate the dynamics of the system, 10 years simulation was performed starting with the parameters shown in table 5.1. Figure 5.4, 5.5 and 5.6 present the results of these test simulations with a common set of parameters. The mean of the simulations in each case was fitted to the official data as the case may be.

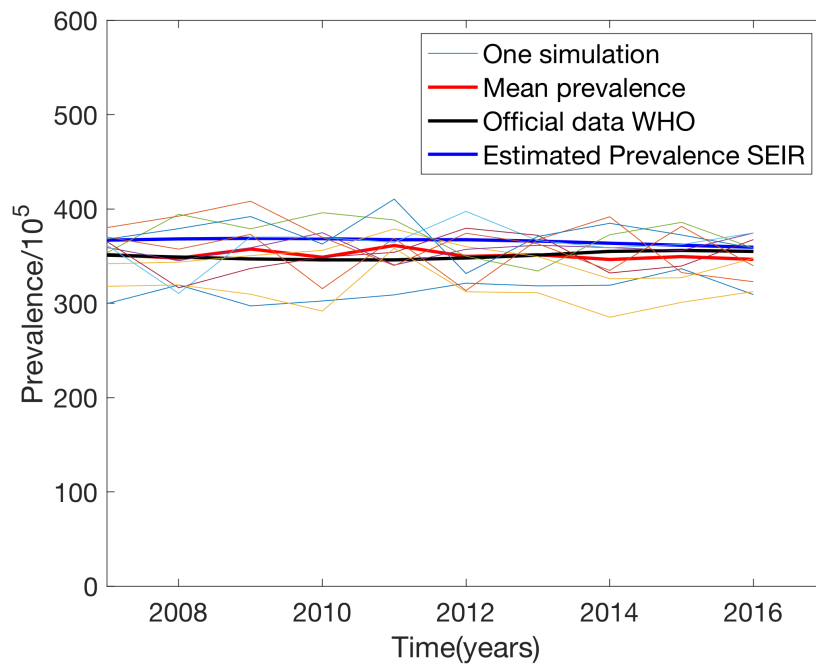


Fig. 5.4 Model validation prevalence: Result of 10 simulation compared to WHO official data of estimated prevalence between 2007 to 2016, and also the SEIR model estimates [7] for Gombe state. Light lines correspond to single ABM simulations, while red line shows the mean of 10 independent simulations.

The initial conditions 5.3 were estimated from the official data of NTBLCP of Gombe state and then adjusted until it fits the model, while the parameters used are described in table 5.1 (more details about the parameters are found in the source code file `IMBparams`).

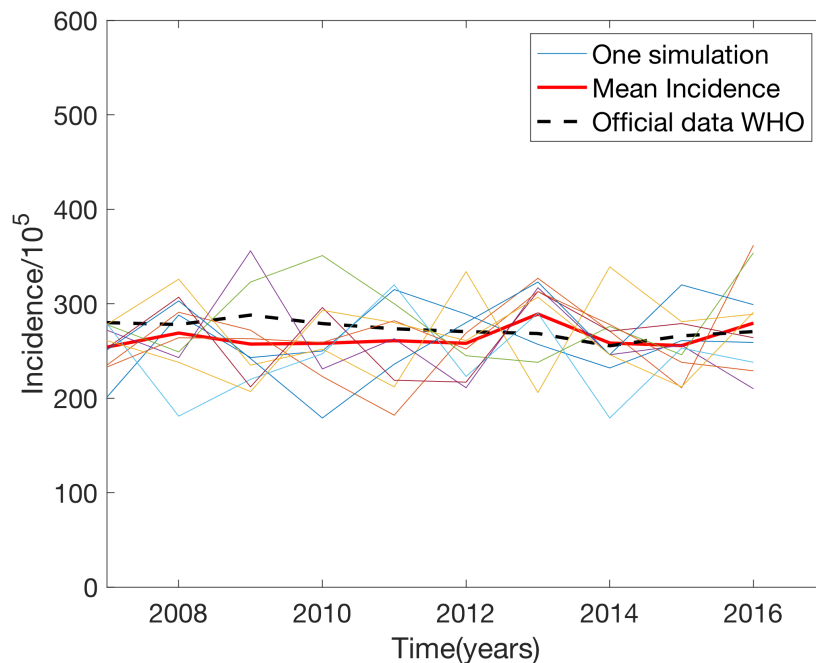


Fig. 5.5 Model validation of incidence rate: Result of 10 simulation compared to official data from WHO between 2007 to 2016. Light lines correspond to single ABM simulations, while red line shows the mean of 10 independent simulations.

Table 5.3 Initial conditions: Initial number of each state variable for the validation simulation of the model.

Healthy	63,760
Infected	36,000
Sick	170
Under treatment	20
Treated	50
Total	100,000

In Figure 5.4, 10 different simulations of prevalence in Gombe are presented. The mean of all the simulation is the red curve fitted between the official data of Nigeria's prevalence from WHO in a black curve, and the estimated prevalence of Gombe state in the blue curve (using techniques adopted from section 3 [7]).

Observation shows that simulation presents some noise level which is expected since we are dealing with stochasticity. The mean prevalence is between the real data and the SEIR-estimated data, and also follow the same behavioral tendency as the real data. In

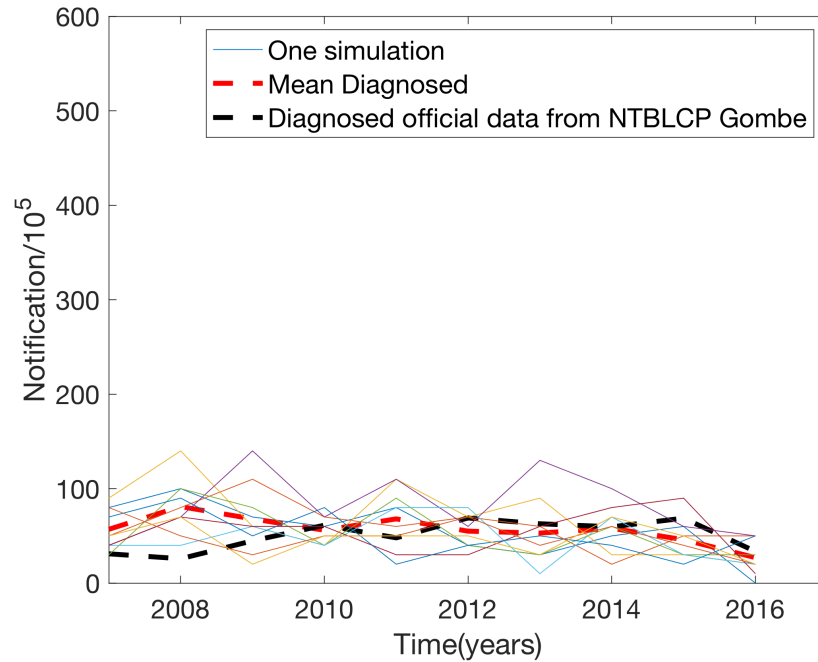


Fig. 5.6 Model validation diagnosed patient: Result of 10 simulation compared with diagnosed cases of Gombe state from official data of NTBLCP Gombe state between 2007 to 2016. Light lines correspond to single ABM simulations, while red line shows the mean of 10 independent simulations.

conclusion, it can be said that the implementation of the model in the C simulator was a success in term of the prevalence of TB observed.

Similar to the prevalence, the incidence in Figure 5.5 also present 10 simulations of TB incidence in Gombe, the mean of the simulation was also compared to the incidence of TB in Nigeria from official data of WHO (in a black curve) between 2007 to 2016. Observation also shows that the mean TB incidence tends to flow the behavior of the real data as in the case of the prevalence.

Figure 5.6 also presents 10 simulation of the diagnosed. The mean of the simulation tends to follow the behaviour of the diagnosed cases in Gombe state from 2007 to 2016. In conclusion, we can assert that the model is well posed and well behave and the implementation was successful.

5.3.3 Experiments testing different types of intervention in Gombe state

We have designed 3 different experiments to test different types of intervention in Gombe state. We start by testing the increase in notification alone which is the direct consequences of

educating the general public about TB and improving AAAQ of health facilities (intervention I); then, we test a decrease in diagnosis delay time alone (also consequence of higher awareness of population and increase in AAAQ of health system) and in combination with the increase in notification (intervention II); finally, we test the effect of a lower transmission due to a through education of the TB patients during diagnosis consultation with the health workers, also in combination with the previous interventions (intervention III).

Increasing notification rate is essential to reduce prevalence of TB in Gombe (Intervention I)

In the context of Nigeria, the country adopted passive case finding policies when it comes to TB [7]. For instance, in the case of Gombe state, an individual must go to the hospital by him or herself in order to become a reported TB case. Then, there is no follow up from the hospital to the patient domiciliary in order to fish out TB sick or infected people and also to check for treatment compliance. We use the simulator to test the effect of this policy on the TB dynamics. We assume the low notification of TB cases in Gombe and Nigeria at large as the direct consequences of this policy. We design an experiment to investigate the effect of increasing the notification parameter gradually from 16% to 20%, 25%, 30%, 60%, and 85% as shown in Figure 5.7.

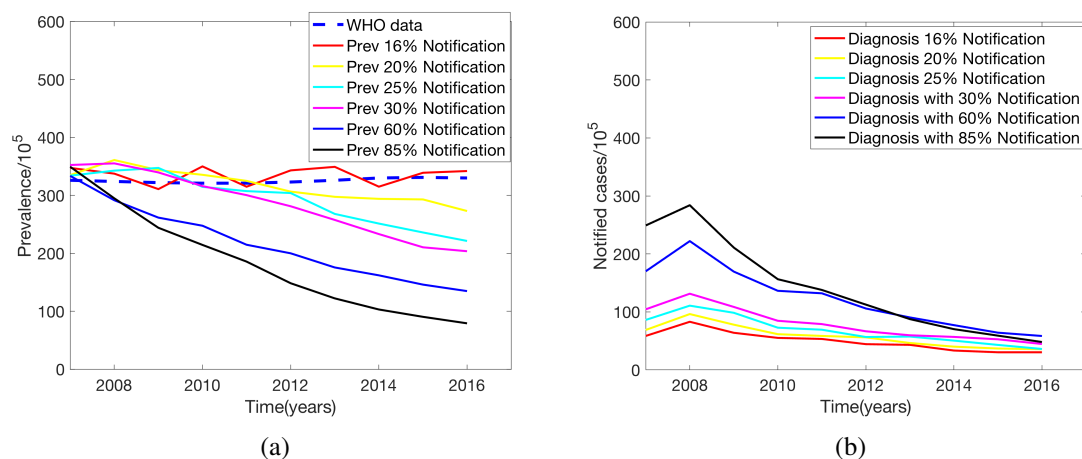


Fig. 5.7 Test for increase in notification: (a) Result of multiple simulations with a gradual increase in notification in the TB prevalence of Gombe state. (b) Result of multiple simulations showing the corresponding diagnosed persons with each increase in notification.

The experiment proves that the lack of active case finding or low notification explains the high prevalence of TB in the area. We carried out multiple simulations with a gradual increase

in the notification parameter from 16% up to 85%, while all other associated parameters remain the same (5.7).

The notification rate of TB in Nigeria is about 16% [7]. Therefore, the notification parameter in the baseline simulation was 16% as seen in table 5.1. Simulation results show that a 10% increase in the notification would decrease the prevalence by almost 35% in 10 years. Furthermore, simulation also shows that 44% increase in the notification will decrease the prevalence by 60%. It would be of interest to do a rough economic evaluation of the consequences of the increase in the notification. According to research carried out by [103], the cost-effectiveness analysis estimated that in the case where there is a reduction in mortality from 14% to 2%, the result in a cost per daily adjusted life year averted of \$330. 35,000 household were screened at the cost of \$363,275, which bring the cost to \$10.378 per household. The yearly cost of TB treatment under full DOT 80% coverage smear-positive cases is \$845.4 per patient [12] in Sub Sahara Africa. The difference in the accumulated number of TB cases in 10 years between the baseline simulation in Figure 5.7(a) and the increment of 10% in notification in Figure 5.7(a) is over 422. The total cost of treatment for these number will be \$356,590 while the cost of active case finding is \$4,379.5 thus a difference of \$352,211 [12]. The sum presented is only a rough estimate per 10^5 population. The cost of TB burden clearly outweighs the cost of investment in an increasing notification or active case finding. A full economic evaluation of the consequences of an increase in the notification would be explored in future research.

The role of the diagnosis delay combined with an increase in notification rate(Intervention II)

We carried out another experiment where we started by testing the effect of reducing the diagnosis delay alone and then combining with an increase in the notification rate. The result of the experiment is given in Figure 5.8

We can consider the increase in the notification rate as a consequence of educating the general public about TB. Besides, a decrease in diagnosis delay can be favoured from such education, although it also requires an improvement on the AAAQ of the health system.

As explained in Section 4, the diagnosis delay distribution among sick that is used in the model is a direct representation of the diagnosis delay time distribution obtained from the fieldwork in Gombe state. The corresponding mean delay time is about 169 days. In Figure 5.8, the green curve is the result of the decrease in the diagnosis delay time alone, using the baseline notification rate (16 %). We use a diagnosis delay distribution which is similar to the original data but reducing the length of period for diagnosis with 23% of the sick been diagnosed in the first 49 days and 77% of the patient been diagnosed within the next 30 days.

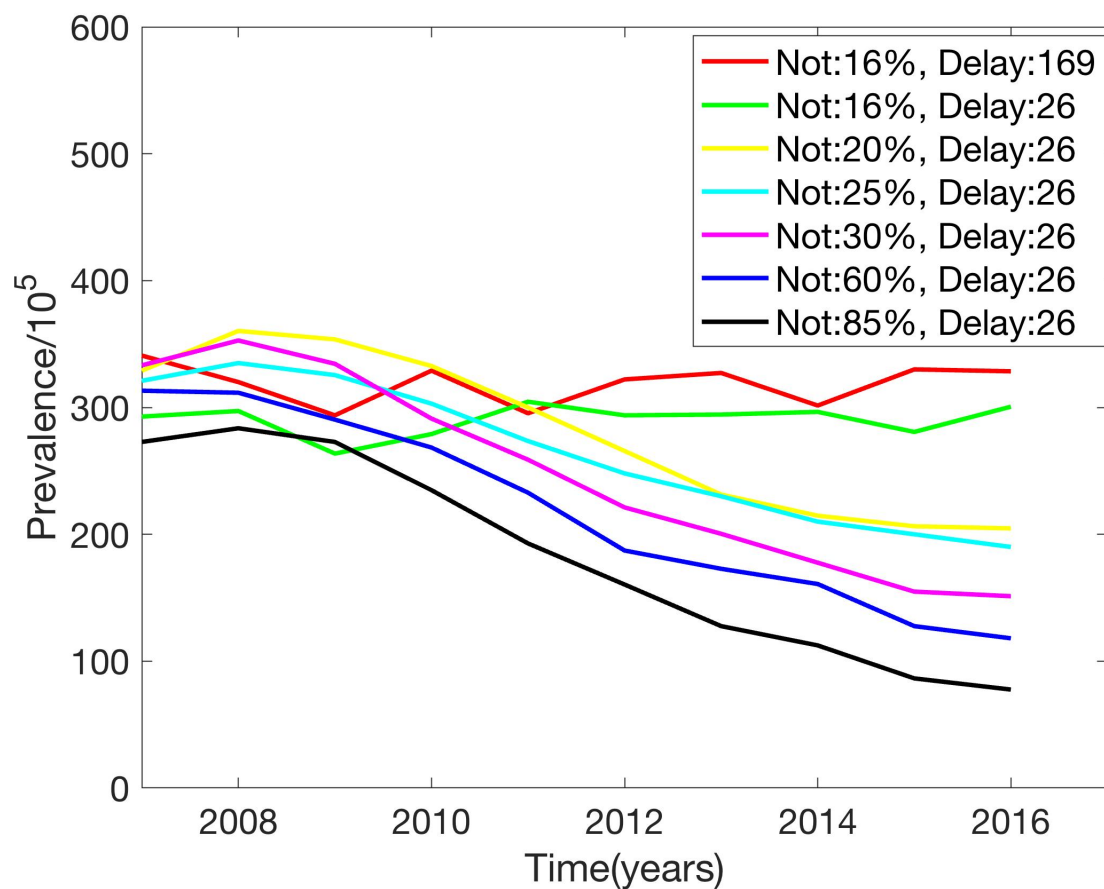


Fig. 5.8 Model test for increasing notification and decreasing diagnosis delay: Multiple simulation indicating the effect of a decrease in the diagnosis delay without and with an increase in notification rate. The Notation Not mean notification and Delay means diagnosis delay time

Then, the mean diagnosis delay time is 26 days. The simulation shows that after 10 years of this strategy there is only about 8% decrease in prevalence. A combined strategy of educating the general public thereby increasing the notification by 10% and reducing the diagnosis delay time shows a 42% decrease in prevalence after 10 years. Similarly, if we adopt the combined strategy of educating the public where the notification is increased by 44% and reducing the diagnosis delay as we did, prevalence shows a 65% decrease after 10 years.

The 3 ways intervention strategy (Intervention III)

We carried out another experiment where we combine: (1) education of the general public as a means of increasing the notification rate; (2) education of the general population and

improvement of the AAAQ of the health system to reduce the diagnosis delay; and (3) education of the patients during diagnosis consultation with the health worker as a means of reducing the time of infectiousness after starting treatment. The results of the experiment are given in Figure 5.8

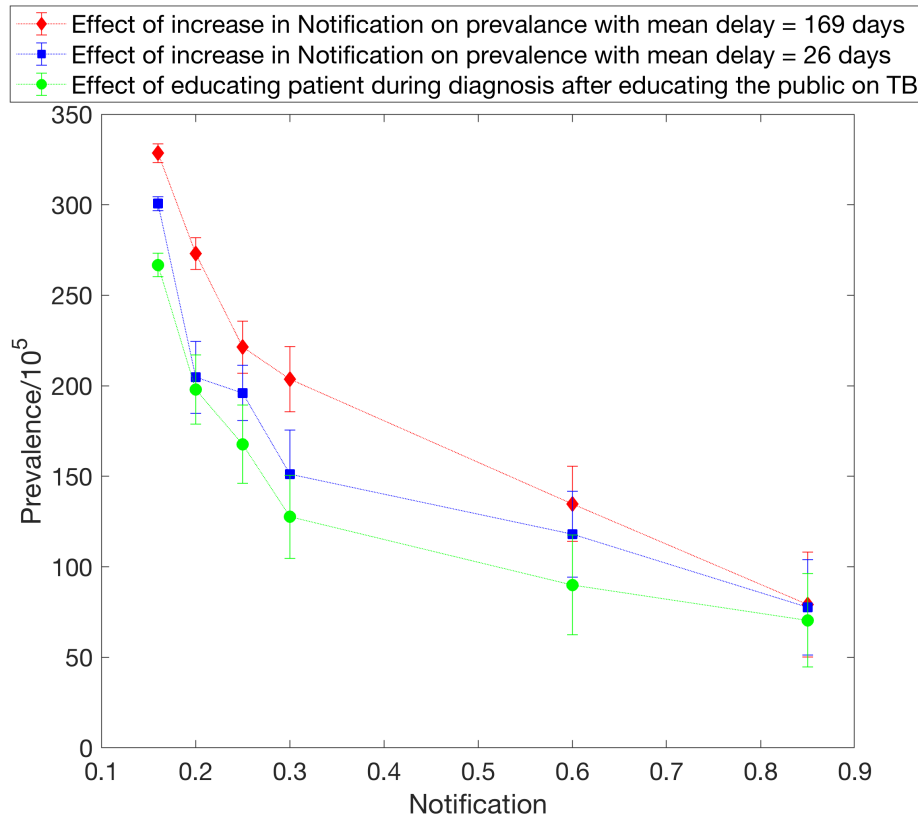


Fig. 5.9 Model Test for 3 ways intervention strategy in TB control: Multiple simulation indicating different type of education both for the general population and the patient with a reduced diagnosis delay.

As above-mentioned, we consider that the increase in the notification rate may be partially achieved with an increasing in education of the general public about TB. The control of TB spreading by the sick persons during the first 2 weeks after diagnosis is considered to be the consequence of educating the patient by the health workers; in addition, a decrease in diagnosis delay also contributes in controlling the spreading of the disease, since the infectious pre-diagnosis period of patients is reduced. The decrease in diagnosis delay may be achieved by a necessary combination of educating general population about the need for being diagnosed together with the improvement of the AAAQ of the health system and

facilities. In Figure 5.9, the red curve is the result of achieving an increase in notification rate only after 10 years of simulation, starting with from the baseline 16% to 20%, 25%...85%. The blue curve is a combination of reducing diagnosis delay and an increase in the notification rate. Finally, the green curve is the combination of both with a reduction in diagnosis delay. It is clear from the simulation that, once the highest notification is achieved (80%), the other two strategies are not necessary. Nevertheless, for intermediate notification rates they reveal to generate a significant decrease in prevalence. Adopting the first strategy by increasing notification by 10% will reduce prevalence by 35% in 10 years, while adopting the 3-way strategy will reduce prevalence by almost 50%. Thus, we can say that optimal strategy that will yield the fastest results is when we combine the 3 strategies. Table 5.4 shows a respective decrease in prevalence with respect to each of the strategies mentioned above.

Table 5.4 Comparison between interventions: Associated decrease in prevalence with respect to each intervention implemented after a period of 10 years.

Notification increase	4%	10%	15%	44%	69%
Intervention I	17%	33%	38%	59%	76%
Intervention II	38%	40%	54%	64%	77%
Intervention III	40%	49%	61%	73%	79%

5.4 Discussion

The current simulator is a step by step towards a heterogeneous, integrated and sophisticated platform for the simulation of tuberculosis spread in high burden areas. The simulator was built from a mechanistic perspective that accounts for the existing knowledge on the disease natural history and on the socioeconomic context taking environmental factors into account from the community of Gombe state. Factors such as immunological capacity, age and gender distribution, the delay time for diagnosis and its distribution were all deduced from the statistical analysis of data collected from Gombe state. The infected population and death due to TB were estimated using the SEIR model from section 2.1, while the probability of progression to active TB, natural death and other parameters were taken from literature.

All parameters have been calibrated to fit tuberculosis situation in Gombe state, the model is robust enough to fit any given situation with the implementation of few changes.

The simulator was implemented on the C program, and it is evident that it is a big improvement on the Netlogo counterpart version in term of computing time. The simulator in C does present a lower computational cost. It can simulate populations 100 times bigger in a reasonable time and perform smaller simulations more than 30 times faster than in Netlogo.

We have shown the capability of the simulator in reproducing both TB prevalence, incidence and notified cases in the population of Gombe with very high accuracy.

With model perfectly calibrated, we performed an experiment to show some intervention strategies. 3 types of intervention were proposed, an increase in the notification rate (Intervention I), an additional decrease in diagnosis delay (Intervention II) and an additional decrease in the infectiousness of diagnosed sick patients (Intervention III). All of them are expected results of interventions at the educational level, both of general population and of patients, as well as of an improvement of the AAAQ of the health system. For intermediate levels of notification rates, Interventions II and III provide significant improvements, while for highest levels of notification rates the benefits from interventions II and III would be residual.

Chapter 6

Conclusions and further work

In this chapter, the conclusions and further work are presented. Conclusions are organized as follows: first, specific conclusions to each specific objective of the thesis are detailed; afterwards, main conclusions that respond to the general aim of the thesis are presented.

6.1 Specific conclusions

Specific conclusions are detailed as individual response to each of the objectives from section 1.5.1.

Research objectives and conclusions

R1: To understand the main features of TB dynamics in low burden countries and high burden countries.

The features of both high burden and low burden context were studied, and differences in epidemiological behaviours were identified.

- Population as limiting factor: Models with population as a limiting factor tends to behave better in high burden setting while models with population as a non-limiting factor behave better in a low burden setting. This bring us to the role of the susceptible population in understanding TB dynamics in a given population.
- Model complexity: It appears that simple models could actually explain a lot about TB in both high burden and low burden setting, but to understand some specific features of these setting, the models have to be complex enough.
- Diagnosis delay: Diagnosis delay is one key parameter that is common to both high burden and low burden context. We observe that most low burden contexts with a good

public health system tend to have low diagnosis delay time while high burden contexts usually have a very high diagnosis delay time

- Latent infected population: Although stated by WHO that the 30% of the world population is infected with latent TB, we confirm this hypothesis in both low and high burden contexts. In the case of low burden contexts, about 20-25% of the population are latently infected with TB, while in the high burden contexts, about 37-60% of the population are latently infected with TB.
- Reinfection: Reinfection is an integral part of understanding TB dynamics, but it is only relevant in high burden contexts.

R2: To determine the cause of high incidence and its persistence in Nigeria

TB in Nigeria was extensively and rigorously studied, 4 major conclusions arise from the study.

- TB in Nigeria still remains a major health challenge due to poor notification and case detection, the notification of TB cases is estimated to 16% only.
- Current strategies focusing on treatment's adherence and efficiency will not yield to a positive results unless the notification of TB cases are drastically improved.
- The persistence of non-decreasing prevalence is also associated with a very large number of latent infected population.
- There is a lot of unreported death due to TB that is associated with low notification rate.

R3: To analyse the behavioural and socioeconomic reasons of TB persistence in Gombe (Nigeria)

To understand the reason behind the low or poor notification detected in chapter 3, we studied the behavioural and socioeconomic factors in one of the communities in Nigeria, namely Gombe state, by analysing data from National TB and Leprosy Control Program and from a field work study where 52 patients were randomly selected and interviewed.

- TB notification in Gombe state is averagely found to be similar to TB notification in Nigeria, which is around 16%. TB prevalence is also estimated to be similar to that one in Nigeria, which is around $330/10^5$ inhabitants.

- Looking at the particular characteristics of TB diagnosed cases in Gombe, we find that HIV/TB co-infection is about 26%, 20% of interviewed patient are MDR, male population are more vulnerable to TB (around 61%), 48% of TB cases occurs in the age group between 25-44 years and TB incidence/10⁵ in the age group 60+ is as high as TB incidence in the most productive age group.
- The population density is a key component in determining the high incidence of TB in different districts.
- Mean diagnosis delay of interviewed patients is 26 weeks on average. One of the main reasons associated with longer diagnosis delay before going to the hospital is the large surface area of the district in the state. Income level is partially associated with delay time at the hospital, which is 1 week on average.

R4: To propose interventions to improve TB control in Gombe (Nigeria)

- 3 main types of interventions have been proposed, aimed at increasing the notification rate, decreasing diagnosis delay time and decreasing the infectiousness of TB patients.
- Education of general population, education of patients and increasing the AAAQ of health system are three ways that can improve the three stated problems.
- Amongst the strategies proposed, the most effective at intermediate notification rates was identified to be the combination of the 3 strategies. Among them, the most important one that should be firstly implemented is the increase in notification rate.

Methodological objectives and conclusions

Different types of mathematical models have been developed and adapted to different contexts and levels of observation.

M1: To develop several compartmental models with increasing complexity adapted to study different context and address specific questions.

- We have developed several compartmental models with different levels of complexity. With each of them, we proposed specific questions that can be addressed such as those related with population structure of the susceptible, time since infection, diagnosis delay time or reinfection process, among others.
- A method for estimating parameters taking advantage of the successive increase in degrees of complexity of SEIR models was successfully developed.

M2: To adapt a general Agent-Based Model for the study of the specific situation in Gombe

- An Agent-Based Model was designed specifically to Gombe state demography, where information from the previous compartmental models and from interviews' statistical analysis was used for its development and fitting.
- The model provides results that are very consistent with the dynamics of TB in Gombe state, and has been used as a virtual experiments platform to help on the design of control interventions.

Public health objective and conclusion

PH: To look for possible in near future, real and feasible interventions to improve TB control in Gombe (Nigeria)

The proposed actions are focused on increasing the **AAAQ** (Accessibility, Availability, Acceptability and Quality) of local health system.

- The most feasible solution for improving TB notification and reduce the time before hospital in near future in Gombe sate is educating the general public, which will also increase Acceptability of the health system. This can be achieved by educating the general public about the importance of visiting the hospital once a person is sick through social media campaigns, posters at the hospitals and public places like schools, programs on TB in radio and TV stations, among others.
- A necessary action is the raising awareness of health workers so that they educate the patients about the main features of TB clinical symptoms and ways of infection. This should be accompanied with the design of graphic material that can be shown and/or distributed among patients.
- Other interventions should be focused on facilitating the diagnosis in remote areas with actions like active case finding (increasing Accessibility and Availability) and new cheap and portable diagnostics tools (increasing Accessibility, Availability and Quality).

6.2 General conclusions

The general aim of this thesis was to improve the control of TB in high incidence contexts. This aim has been pursued through the study of TB situation in Nigeria and, in particular, in

Gombe state. In fact, improving TB control in high burden contexts is necessarily associated with understanding the underlined problem of TB control programs in a specific area under investigation. There are no universal solutions for a global TB problem; conversely, particular solutions to specific contexts will provide the necessary ways towards the global TB problem.

Some of the major challenges of TB control in Nigeria have been tackled in this thesis. Among the main identified problems that prevent TB incidence to decrease in high burden contexts like Nigeria, we can highlight the following conclusions:

- **Notification rate:** Notification rate or un-notified cases of TB is one of the major problems of high burden contexts countries in terms of TB control. The methods and procedures of identifying and tackling the issue of un-notified TB cases were explored. It has been shown that the low notification rate is the bottleneck for the efficacy of other control measures.
- **Diagnosis delay:** Delay in diagnosis has also found to be another major area of concern in high burden contexts like Gombe state.
- **TB treatment adherence:** This was also found to be a problem in the context of high burden countries, but it has been shown that an increase in treatment adherence requires a previous drastic increase in notification rate in order to contribute to the TB control.

Among the main socioeconomic factors that are behind the stated problems in Gombe state, we have concluded that:

- Education of patients and of general population is essential in order to make progress on the notification rate and diagnosis delay, as well as to prevent sick patients of generating new cases. The education of patients requires a previous specific formation of health workers on this sense.
- The improvement of AAAQ of the health system is also essential in order to facilitate the access and TB diagnosis to Nigerian population, and it will also benefit from a powerful education program.
- An active case finding strategy may help on increasing the diagnosis but also on educating general population.

6.3 Perspectives and further work

The results gathered in this thesis have answered many questions, but they have also suggested new ones and opened new avenues to explore. Below are a few of the many possibilities for further developments on this field.

In the research presented, the TB situation in Nigeria has been investigated. The ABM simulator has done a good job in mimicking the TB dynamics in some areas of the country, but the parameter space for the model was not explored. The simulator could also be improved by introducing allowing the population to be an open population, introducing birth rate as well as death in stead of a constant population. Another twist into the model could also be the consideration of intensity of contact during the transmission of the disease, environment could also play a vital role since the disease is an air bone disease. We can improve the simulator by considering places where the transmission occurs such as school, house holds recreational centers etc. We can improve the result from chapter 5 by carrying out statistical analysis on the result from the virtual experiment. This could add to our understanding of which intervention strategy would be optimal use in containing TB in this area. Additionally, a cost analysis for carrying out such intervention should be consider, this will allow the stake holders an insight into the types of resources needed in containing the epidemic.

Finally, we hope to carry out a pilot program, to implement some of the interventions mentioned in chapter 5 in one of the district of Gombe state, and monitor the outcome.

References

- [1] GOMBE. <http://www.citypopulation.de/php/nigeria-admin.php?adminid=NGA016>.
- [2] GOMBE death probability1 ref. <https://www.indexmundi.com/facts/nigeria/mortality-rate>.
- [3] GOMBE death probability2 ref. <http://www.healthdata.org/nigeria>.
- [4] GOMBE death probability3 ref. <https://childmortality.org/data>.
- [5] Nigeria Development Data. {http://www.demographicdividend.org/country_highlights/nigeria/}.
- [6] Nigeria Tuberculosis profile. <http://www.who.int/tb/country/data/profiles/en/>.
- [7] Ahmad, N. M., Montañola-Sales, C., Prats, C., Musa, M., López, D., and Casanovas-Garcia, J. (2018). Analyzing policymaking for tuberculosis control in nigeria. *Complexity*, 2018.
- [8] Aliyu, A. A. and Amadu, L. (2017). Urbanization, cities, and health: The challenges to Nigeria—A review. *Annals of African medicine*, 16(4):149.
- [9] Allen, L. J., Brauer, F., Van den Driessche, P., and Wu, J. (2008). *Mathematical epidemiology*, volume 1945. Springer.
- [10] Asgedom, S. W., Teweldemedhin, M., and Gebreyesus, H. (2018). Prevalence of multidrug-resistant tuberculosis and associated factors in ethiopia: A systematic review. *Journal of Pathogens*, 2018.
- [11] Bala, S. I. and Ahmad, N. M. R. (2017). Bistability analysis in mathematical model of M-phase control in *Xenopus* oocyte extracts. *Computational and Applied Mathematics*.
- [12] Baltussen, R., Floyd, K., and Dye, C. (2005). Cost effectiveness analysis of strategies for tuberculosis control in developing countries. *Bmj*, 331(7529):1364.
- [13] Banerji, D. and Andersen, S. (1963). A sociological study of awareness of symptoms among persons with pulmonary tuberculosis. *Bulletin of the World Health Organization*, 29(5):665.
- [14] Benatar, S. R. (1995). Prospects for global health: lessons from tuberculosis. *Thorax*, 50(5):487.

- [15] Bhatt, K. and Salgame, P. (2007). Host Innate Immune Response to Mycobacterium tuberculosis. *Journal of Clinical Immunology*, 27(4):347–362.
- [16] Brust, B., Lecoufle, M., Tuailon, E., Dedieu, L., Canaan, S., Valverde, V., and Kremer, L. (2011). Mycobacterium tuberculosis Lipolytic Enzymes as Potential Biomarkers for the Diagnosis of Active Tuberculosis. *PLOS ONE*, 6(9):1–10.
- [17] C., H. (2011). *An Introduction to Epidemiology*. In: Teare M. (eds) *Genetic Epidemiology. Methods in Molecular Biology (Methods and Protocols) (Vol.713)*. Humana Press, Totowa, NJ.
- [18] Cardona, P. and Ruiz-Manzano, J. (2004). On the nature of mycobacterium tuberculosis-latent bacilli. *European Respiratory Journal*, 24(6):1044–1051.
- [19] Cardona, P.-J. (2018a). Pathogenesis of tuberculosis and other mycobacteriosis. *Enfermedades infecciosas y microbiología clinica (English ed.)*, 36(1):38–46.
- [20] Cardona, P.-J. (2018b). Patogénesis de la tuberculosis y otras micobacteriosis. *Enfermedades infecciosas y microbiología clinica*, 36(1):38–46.
- [21] Cardona, P.-J., Català, M., and Prats, C. (2020). Origin of tuberculosis in the paleolithic predicts unprecedented population growth and female resistance. *Scientific Reports*, 10(1):1–20.
- [22] cdc-2019-lesson3 (2019). Principal of epidemiology. [Measure of risk] <https://www.cdc.gov/csels/dsepd/ss1978/lesson3/section2.html>.
- [23] Centers for Disease Control and Prevention (US) and National Center for Prevention Services (US). Division of Tuberculosis Elimination and National Center for HIV and STD and TB Prevention (US). Division of Tuberculosis Elimination (1999). *Reported tuberculosis in the United States*. US Department of Health and Human Services, Public Health Service, Centers for Disease Control and Prevention, National Center for Prevention Services, Division of Tuberculosis Elimination.
- [24] Corberán-Vallet, A., Santonja, F., Jornet-Sanz, M., and Villanueva, R.-J. (2018). Modeling chickenpox dynamics with a discrete time bayesian stochastic compartmental model. *Complexity*, 2018.
- [25] Daniel, O., Oladapo, O., and Alausa, O. (2006). Default from tuberculosis treatment programme in sagamu, nigeria. *Niger J Med*, 15(1):63–67.
- [26] Davis, J., Clay, H., Lewis, J. L., Ghori, N., Herbomel, P., and Ramakrishnan, L. (2002). Real-Time Visualization of Mycobacterium-Macrophage Interactions Leading to Initiation of Granuloma Formation in Zebrafish Embryos. *Immunity*, 17(6):693–702.
- [27] Dim, C. C. and Dim, N. R. (2013). Trends of tuberculosis prevalence and treatment outcome in an under-resourced setting: The case of Enugu state, South East Nigeria. *Nigerian medical journal: journal of the Nigeria Medical Association*, 54(6):392.
- [28] Dodd, P. J., White, R. G., and Corbett, E. L. (2011). Periodic active case finding for tb: when to look? *PloS one*, 6(12):e29130.

- [29] Ferebee, S. (1970). Controlled chemoprophylaxis trial in tuberculosis, a general review. *Adv Tuberc Res*, 17:28–106.
- [30] Foppa, I. M. (2016). *A Historical Introduction to Mathematical Modeling of Infectious Diseases: Seminal Papers in Epidemiology*. Academic Press.
- [31] García-García, J.-M., Blanquer, R., Rodrigo, T., Caylà, J. A., Caminero, J. A., Vidal, R., Casals, M., Ruiz-Manzano, J., and the Working Group on Completion of Tuberculosis Treatment in Spain (2011). Social, Clinical and Microbiological Differential Characteristics of Tuberculosis among Immigrants in Spain. *PLOS ONE*, 6(1):1–7.
- [32] Grimm, V., Berger, U., Bastiansen, F., Eliassen, S., Ginot, V., Giske, J., Goss-Custard, J., Grand, T., Heinz, S. K., Huse, G., et al. (2006). A standard protocol for describing individual-based and agent-based models. *Ecological modelling*, 198(1-2):115–126.
- [33] Grimm, V. and Railsback, S. F. (2005). *Individual-based modeling and ecology*. Princeton university press.
- [34] Guzzetta, G., Ajelli, M., Yang, Z., Merler, S., Furlanello, C., and Kirschner, D. (2011). Modeling socio-demography to capture tuberculosis transmission dynamics in a low burden setting. *Journal of Theoretical Biology*, 289(1):197–205.
- [35] Haefner, J. W. (2005). *Modeling Biological Systems:: Principles and Applications*. Springer Science & Business Media.
- [36] Hassan, A. S., Garba, S. M., Gumel, A. B., and Lubuma, J.-S. (2014). Dynamics of Mycobacterium and bovine tuberculosis in a human-buffalo population. *Computational and mathematical methods in medicine*, 2014.
- [37] Hunter, R. L. (2011). Pathology of post primary tuberculosis of the lung: An illustrated critical review. *Tuberculosis*, 91(6):497–509.
- [38] Kermack, W. O. and McKendrick, A. G. (1927). A contribution to the mathematical theory of epidemics. *Proceedings of the royal society of london. Series A, Containing papers of a mathematical and physical character*, 115(772):700–721.
- [39] Kermack, W. O. and McKendrick, A. G. (1991). Contributions to the mathematical theory of epidemics—i. *Bulletin of mathematical biology*, 53(1):33–55.
- [40] Kim, P. and Lee, C. H. (2018). Epidemic spreading in complex networks with resilient nodes: Applications to fmd. *Complexity*, 2018.
- [41] Krasniqi, S., Jakupi, A., Daci, A., Tigani, B., Jupolli-Krasniqi, N., Pira, M., Zhjeqi, V., and Neziri, B. (2017). Tuberculosis treatment adherence of patients in kosovo. *Tuberculosis research and treatment*, 2017.
- [42] Krsulovic, F. A. M. and Lima, M. (2016). Tuberculosis Epidemiology at the Country Scale: Self-Limiting Process and the HIV Effects. *PLOS ONE*, 11(4):1–13.
- [43] Kyu, H. H., Maddison, E. R., Henry, N. J., Mumford, J. E., Barber, R., Shields, C., Brown, J. C., Nguyen, G., Carter, A., and Wolock, T. M. (2018). The global burden of tuberculosis: results from the global burden of disease study 2015. *The Lancet Infectious Diseases*, 18(3):261–284.

- [44] Lawson, L., Zhang, J., Gomgnimbou, M. K., Abdurrahman, S. T., Le Moullec, S., Mohamed, F., Uzoewulu, G. N., Sogaolu, O. M., Goh, K. S., Emenyonu, N., et al. (2012). A molecular epidemiological and genetic diversity study of tuberculosis in ibadan, nnewi and abuja, nigeria. *PLoS one*, 7(6):e38409.
- [45] Lienhardt, C., Glaziou, P., Uplekar, M., Lönnroth, K., Getahun, H., and Raviglione, M. (2012). Global tuberculosis control: lessons learnt and future prospects. *Nature Reviews Microbiology*, 10(6):407.
- [46] Liu, L. and Wang, Y. (2014). A mathematical study of a tb model with treatment interruptions and two latent periods. *Computational and mathematical methods in medicine*, 2014.
- [47] Liu, S., Li, A., Feng, X., Zhang, X., and Wang, K. (2016). A dynamic model of human and livestock tuberculosis spread and control in urumqi, xinjiang, china. *Computational and mathematical methods in medicine*, 2016.
- [48] Macal, C. M. and North, M. J. (2005). Tutorial on agent-based modeling and simulation. In *Proceedings of the Winter Simulation Conference, 2005.*, pages 14–pp. IEEE.
- [49] Martcheva, M. (2015). *An introduction to mathematical epidemiology (Vol.61)*. New York: Springer.
- [50] Menzies, N. A., Cohen, T., Lin, H.-H., Murray, M., and Salomon, J. A. (2012). Population health impact and cost-effectiveness of tuberculosis diagnosis with Xpert MTB/RIF: a dynamic simulation and economic evaluation. *PLoS medicine*, 9(11):e1001347.
- [51] Millet, J.-P., Orcau, A., Rius, C., Casals, M., de Olalla, P. G., Moreno, A., Nelson, J. L., Caylà, J. A., and the Barcelona Tuberculosis Working Group (2011). Predictors of Death among Patients Who Completed Tuberculosis Treatment: A Population-Based Cohort Study. *PLOS ONE*, 6(9):1–8.
- [52] Montañola-Sales, C., Gilabert-Navarro, J. F., Casanovas-Garcia, J., Prats, C., López, D., Valls, J., Cardona, P. J., and Vilaplana, C. (2015). Modeling tuberculosis in barcelona: a solution to speed-up agent-based simulations. In *Proceedings of the 2015 Winter Simulation Conference*, pages 1295–1306. IEEE Press.
- [53] Moualeu-Ngangue, D. P., Röblitz, S., Ehrig, R., and Deuffhard, P. (2015). Parameter Identification in a Tuberculosis Model for Cameroon. *PLOS ONE*, 10(4):1–20.
- [54] Murphy, B. M., Singer, B. H., and Kirschner, D. (2003). On treatment of tuberculosis in heterogeneous populations. *Journal of Theoretical Biology*, 223(4):391–404.
- [55] Murray, C. J. and Salomon, J. A. (1998). Modeling the impact of global tuberculosis control strategies. *Proceedings of the National Academy of Sciences*, 95(23):13881–13886.
- [56] Na, H. J., Eom, J. S., Lee, G., Mok, J. H., Kim, M. H., Lee, K., Kim, K. U., and Lee, M. K. (2016). Exposure to Mycobacterium tuberculosis during Flexible Bronchoscopy in Patients with Unexpected Pulmonary Tuberculosis. *PLOS ONE*, 11(5):1–10.

- [57] Nagpaul, D. R., Vishwanath, M. K., and Dwarakanath, G. (1970). A socio-epidemiological study of out-patients attending a city tuberculosis clinic in India to judge the place of specialized centres in a tuberculosis control programme. *Bulletin of the World Health Organization*, 43(1):17.
- [58] Nahid, P., Bliven, E. E., Kim, E. Y., Mac Kenzie, W. R., Stout, J. E., Diem, L., Johnson, J. L., Gagneux, S., Hopewell, P. C., Kato-Maeda, M., and the Tuberculosis Trials Consortium (2010). Influence of *M. tuberculosis* Lineage Variability within a Clinical Trial for Pulmonary Tuberculosis. *PLOS ONE*, 5(5):1–6.
- [59] Nigeria Tuberculosis Fact Sheet (2012). United States Embassy in Nigeria.
- [60] Nyendak, M. R., Park, B., Null, M. D., Baseke, J., Swarbrick, G., Mayanja-Kizza, H., Nsereko, M., Johnson, D. F., Gitta, P., Okwera, A., Goldberg, S., Bozeman, L., Johnson, J. L., Boom, W. H., Lewinsohn, D. A., Lewinsohn, D. M., for the Tuberculosis Research Unit, and the Tuberculosis Trials Consortium (2013). Mycobacterium tuberculosis Specific CD8+ T Cells Rapidly Decline with Antituberculosis Treatment. *PLOS ONE*, 8(12):1–10.
- [61] Obasanya, J., Abdurrahman, S. T., Oladimeji, O., Lawson, L., Dacombe, R., Chukwueme, N., Abiola, T., Mustapha, G., Sola, C., Dominguez, J., et al. (2015). Tuberculosis case detection in nigeria, the unfinished agenda. *Tropical Medicine & International Health*, 20(10):1396–1402.
- [62] Okuonghae, D. and Ikhimwin, B. O. (2016). Dynamics of a Mathematical Model for Tuberculosis with Variability in Susceptibility and Disease Progressions Due to Difference in Awareness Level. *Frontiers in Microbiology*, 6:1530.
- [63] Okuonghae, D. and Omosigho, S. (2011). Analysis of a mathematical model for tuberculosis: What could be done to increase case detection. *Journal of theoretical biology*, 269(1):31–45.
- [64] on HIV/AIDS, J. U. N. P. et al. (2010). *Global report: UNAIDS report on the global AIDS epidemic 2010*. Un aids.
- [65] Onubogu, C., Kunle-Ope, C., Onyejebu, N., Nwokoye, N., Raheem, T., Igbasi, U., Tochukwu, N., Omoloye, R., Ejezie, C., Musa, A., et al. (2010). Prevalence of tuberculosis and human immunodeficiency virus (tb/hiv) co-infections amongst patients with bronchopulmonary disorders in lagos. *African Journal of Microbiology Research*, 4(18):1904–1908.
- [66] Organization, W. H. (2013). *Global tuberculosis report 2013*. World Health Organization.
- [67] Organization, W. H. et al. (2010). Accelerating the implementation of collaborative tb/hiv activities in the who european region: report of the meeting of the world health organization (headquarters and regional office for europe) in collaboration with the tb/hiv working group of the stop tb partnership, 16-17 july 2010, vienna, austria. Technical report, Geneva: World Health Organization.
- [68] Organization, W. H. et al. (2018a). *Global tuberculosis report 2018*. World Health Organization.

- [69] Organization, W. H. et al. (2018b). Guidance for country-level tb modelling.
- [70] Organization, W. H. and Others (2016a). Global tuberculosis report 2016.
- [71] Organization, W. H. and Others (2016b). Media Centre: Tuberculosis, Fact sheet.
- [72] Oruonye, E., Abubakar, H., Ahmed, M., Dan, Y., et al. (2017). Hiv/aids interventions in gombe state nigeria; challenges of sustaining the gains. *International Journal of Asian Social Science*, 7(6):448–457.
- [73] Otu, A. A. et al. (2013). A review of the national tuberculosis and leprosy control programme (ntblcp) of nigeria: Challenges and prospects. *Annals of Tropical Medicine and Public Health*, 6(5):491.
- [74] Ozcaglar, C., Shabbeer, A., Vandenberg, S. L., Yener, B., and Bennett, K. P. (2012). Epidemiological models of mycobacterium tuberculosis complex infections. *Mathematical Biosciences*, 236(2):77 – 96.
- [75] Parhar, A., Gao, Z., Heffernan, C., Ahmed, R., Egedahl, M. L., and Long, R. (2015). Is Early Tuberculosis Death Associated with Increased Tuberculosis Transmission? *PLOS ONE*, 10(1):1–14.
- [76] Paulson, T. (2013). Epidemiology: a mortal foe. *Nature*, 502(7470):S2—S3.
- [77] Peyron, P., Vaubourgeix, J., Poquet, Y., Levillain, F., Botanch, C., Bardou, F., Daffé, M., Emile, J.-F., Marchou, B., Cardona, P.-J., de Chastellier, C., and Altare, F. (2008). Foamy Macrophages from Tuberculous Patients’ Granulomas Constitute a Nutrient-Rich Reservoir for *M. tuberculosis* Persistence. *PLOS Pathogens*, 4(11):1–14.
- [78] Porcelli, S. A. and Jacobs, W. R. (2019). Exacting edward jenner’s revenge: The quest for a new tuberculosis vaccine. *Science translational medicine*, 11(490):eaax4219.
- [79] Prats, C., Montañola-Sales, C., Gilabert, J. F., Valls, J., Casanovas-Garcia, J., Vilaplana, C., Cardona, P.-J., and Lopez, D. (2016). Individual-based modeling of tuberculosis in a user-friendly interface: understanding the epidemiological role of population heterogeneity in a city. *Frontiers in Microbiology*, 6(1564).
- [80] Punch (2019). Punch.
- [81] Railsback, S. F. and Grimm, V. (2019). *Agent-based and individual-based modeling: a practical introduction*. Princeton university press.
- [82] Raviglione, M. C. and Uplekar, M. W. (2006). Who’s new stop tb strategy. *The Lancet*, 367(9514):952–955.
- [83] Rubel, A. J. and Garro, L. C. (1992). Social and cultural factors in the successful control of tuberculosis. *Public health reports*, 107(6):626.
- [84] Sloot, R., Schim van der Loeff, M. F., Kouw, P. M., and Borgdorff, M. W. (2014). Risk of tuberculosis after recent exposure. a 10-year follow-up study of contacts in amsterdam. *American journal of respiratory and critical care medicine*, 190(9):1044–1052.

- [85] Song, B., Castillo-Chavez, C., and Aparicio, J. P. (2002). Tuberculosis models with fast and slow dynamics: the role of close and casual contacts. *Mathematical Biosciences*, 180(1):187–205.
- [86] Sreeramareddy, C. T., Qin, Z. Z., Satyanarayana, S., Subbaraman, R., and Pai, M. (2014). Delays in diagnosis and treatment of pulmonary tuberculosis in india: a systematic review. *The International Journal of Tuberculosis and Lung Disease*, 18(3):255–266.
- [87] Stead, W. W., Lofgren, J. P., Warren, E., and Thomas, C. (1985). Tuberculosis as an endemic and nosocomial infection among the elderly in nursing homes. *New England Journal of Medicine*, 312(23):1483–1487.
- [88] Susumu Ohno (1970). *Evolution by Gene Duplication*. Springer, Berlin, Heidelberg.
- [89] Telisinghe, L., Fielding, K. L., Malden, J. L., Hanifa, Y., Churchyard, G. J., Grant, A. D., and Charalambous, S. (2014). High Tuberculosis Prevalence in a South African Prison: The Need for Routine Tuberculosis Screening. *PLOS ONE*, 9(1):1–9.
- [90] USAID (2014). *HEALTH FACT SHEET*. web. <https://www.usaid.gov/sites/default/files/documents/1860/USAID-Nigeria-Health-Fact-Sheet.pdf>.
- [91] Wallis, R. S. (2016). Mathematical Models of Tuberculosis Reactivation and Relapse. *Frontiers in Microbiology*, 7:669.
- [92] web (2008). *National demographic health survey*. web. <https://datacatalog.worldbank.org/dataset/nigeria-demographic-and-health-survey-2008>.
- [93] web (2015). *WHO End TB Strategy*. web. http://www.who.int/tb/post2015_strategy/en.
- [94] web (2018a). *death rate between 15 and above*. web. <https://childmortality.org/data>.
- [95] web (2018b). *death rate between 5-14*. web. <http://www.healthdata.org/nigeria>.
- [96] web (2018c). *death rate under 5 years*. web. <https://www.indexmundi.com/facts/nigeria/mortality-rate>.
- [97] web (2018d). *Vanguard News paper*. web. <https://www.vanguardngr.com/2018/06/nigeria-overtakes-india-as-worlds-poverty-capital-report/>.
- [98] WHO (2017). *Estimated TB incidence rate*. web. <http://gamapsrver.who.int/mapLibrary/app/searchResults.aspx>.
- [99] WHO (2019). Global tuberculosis report 2019.
- [100] World Health Organization (2016). *Global Tuberculosis Report 2016*. WHO, Geneva, Switzerland.
- [101] World Health Organization (2017a). Global Tuberculosis Report.
- [102] World Health Organization (2017b). *Global Tuberculosis Report 2017*. WHO, Geneva, Switzerland.

-
- [103] Yadav, R. P., Nishikiori, N., Satha, P., Eang, M. T., and Lubell, Y. (2014). Cost-effectiveness of a tuberculosis active case finding program targeting household and neighborhood contacts in cambodia. *The American journal of tropical medicine and hygiene*, 90(5):866–872.
- [104] Zumla, A., George, A., Sharma, V., Herbert, R. H. N., Oxley, A., and Oliver, M. (2015). The who 2014 global tuberculosis report—further to go. *The Lancet Global Health*, 3(1):e10–e12.
- [105] Zwerling, A., Shrestha, S., and Dowdy, D. W. (2015). Mathematical modelling and tuberculosis: advances in diagnostics and novel therapies. *Advances in medicine*, 2015.

Appendix A

Mathematical analysis

Analysis of the model

The model analysis is based on the SE8IR described in section 2.2. The general or total population $N(t)$ is given by,

$$N(t) = S(t) + E_1(t) + E_2(t) + E_3(t) + E_4(t) + E_5(t) + E_6(t) + E_7(t) + E_{>7}(t) + I(t) + R(t).$$

Now if we let $x_1(t) = S(t)$, $x_2(t) = E_1(t)$, ... $x_{11}(t) = R(t)$, the SE8IR model is given by

$$\frac{dx_1}{dt} = \Pi - (\Lambda - \mu)x_1, \quad (\text{A.1})$$

$$\frac{dx_2}{dt} = \Lambda(x_1 + x_{11}) + \Lambda_1 \sum x_i - (v_1 + p_1 + k)x_2, \quad i = 3, 4, 5 \dots 9. \quad (\text{A.2})$$

$$\frac{dx_3}{dt} = kx_2 - (v_2 + p_2 + k)x_3 - \Lambda_1 x_3, \quad (\text{A.3})$$

$$\frac{dx_4}{dt} = kx_3 - (v_3 + p_3 + k)x_4 - \Lambda_1 x_4, \quad (\text{A.4})$$

$$\frac{dx_5}{dt} = kx_4 - (v_4 + p_4 + k)x_5 - \Lambda_1 x_5, \quad (\text{A.5})$$

$$\frac{dx_6}{dt} = kx_5 - (v_5 + p_5 + k)x_6 - \Lambda_1 x_6, \quad (\text{A.6})$$

$$\frac{dx_7}{dt} = kx_6 - (v_6 + p_6 + k)x_7 - \Lambda_1 x_7, \quad (\text{A.7})$$

$$\frac{dx_8}{dt} = kx_7 - (v_7 + p_7 + k)x_8 - \Lambda_1 x_8, \quad (\text{A.8})$$

$$\frac{dx_9}{dt} = kx_8 - (v_8 + p_8 + k)x_9 - \Lambda_1 x_9, \quad (\text{A.9})$$

$$\frac{dx_{10}}{dt} = \sum (p_i x_i) - (\mu + k)x_{10}, \quad (\text{A.10})$$

$$\frac{dx_{11}}{dt} = k_3 x_9 + k_1 x_{10} - (\mu + \Lambda)x_{11}. \quad (\text{A.11})$$

Consider the model (A.1 - A.11) with death as natural death in all of the compartment.

Proposition 1.0:

The following biologically feasible region of model (A.1 - A.11)

$\Omega = \{(x_i \in \mathbb{R}^{11} : \sum x_i \leq \frac{\Pi}{\mu}, i = 1, 2, 3 \dots 11)\}$ is positively invariant and any solution of the model with initial condition in Ω will remain in the region for $t \geq 0$

Proof. Adding the equations in model (A.1 - A.11) gives

$$\frac{dN(t)}{dt} = \Pi - \mu N(t) \Rightarrow N(t) \leq \frac{\Pi}{\mu} \quad (\text{A.12})$$

by (A.12) and Gronwall [36, 11], Ω is positively invariant and it is enough to consider the dynamics of the model (A.1 - A.11) in Ω epidemiologically and biologically well posed. \square

Positivity of Solution

if $x_i(0) > 0$, for $i = 1, 2, 3, \dots, 11$ then the solutions $x_i(t)$ of the model (A.1 - A.11) are positive for all $t \geq 0$.

Proof. : It is clear from equation A.1 of the model that

$$\frac{dx_1}{dt} \geq -(\Lambda - \mu)x_1, \quad (\text{A.13})$$

so that

$$\frac{dX_1}{dt} = \Pi - (\Lambda - \mu)X_1, \quad (\text{A.14})$$

$$x_1(t) \geq x_1(0)e^{\int(-(\Pi+\mu)dt)}. \quad (\text{A.15})$$

Using a similar approach, it can be shown that $x_i(t) \geq 0$ for all i and hence the proof. \square

Disease-Free Equilibrium (DFE)

The DFE of the model (A.1 - A.11) is given by

$$E_0 = (x_1^*, x_2^*, x_3^*, x_4^*, x_5^*, x_6^*, x_7^*, x_8^*, x_9^*, x_{10}^*, x_{11}^*) = \left(\frac{\Pi}{\mu}, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0\right) \quad (\text{A.16})$$

this type of equilibria occurs when there is no infection. At disease free equilibrium, we set the reinfection Λ_1 to zero, we have $x_2 = x_3 = x_4 = \dots = x_{11}$, $\Rightarrow N = x_1$. The linear stability of E_0 can be established using the next generation operator method on the model (A.1 - A.11). The only occurrence of the variable x_1 in equation 2 to 11 of model (A.1 - A.11) is $(\alpha x_{10} x_1)/N$ in equation 2 which becomes αx_{10} at equilibrium, therefore linearization of equation 2 to 10 is close because it doesn't involve the deviation of x_1 from its steady state.

$$\mathbb{B} = \begin{pmatrix} -w_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ k & -w_2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & k & -w_3 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & k & -w_4 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & k & -w_5 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & k & -w_6 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & k & -w_7 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & k & -w_8 & 0 \\ p_1 & p_2 & p_3 & p_4 & p_5 & p_6 & p_7 & p_8 & -(k_2 + \mu) \end{pmatrix}, \quad (\text{A.20})$$

where $w_i = p_i + v_1 + k$. This mean all event that lead to the new infection are integrated via \mathbb{T} and all other event via \mathbb{B} . Let $\mathbb{T} = -\mathbb{T}\mathbb{B}^{-1}$, then the dominant eigenvalue of \mathbb{T} is our basic reproduction number, which is given as

$$R_0 = \frac{\alpha\eta}{\zeta}, \quad (\text{A.21})$$

where $\eta = p_8k^7 + w_8p_7k^6 + w_8w_7p_6k^5 + w_8w_7w_6p_5k^4 + w_8w_7w_6w_5p_4k^3 + w_8w_7\dots w_4p_3k^2 + w_8w_7\dots w_3p_2k + w_8w_7\dots w_2p_1$. and $\zeta = (\mu + k_2)w_8w_7\dots w_1$, $w_i = (v_i + k + p_i)$, $i = 1, 2, 3, \dots, 7$, $w_8 = (v_8 + k_3 + p_8)$

Definitions of terms in the R_0 : the threshold quantity, R_0 , represents the average number of *Mtb* secondary cases in the population that one *Mtb*-sick can generate if introduced into a completely susceptible population [11, 36]

Lemma 1.

The DFE E_0 of the model (A.1 - A.11) is locally asymptotically stable (LAS) if $R_0 < 1$ and unstable if $R_0 > 1$ [11, 36].

Global Stability analysis

Consider the model A.18 without reinfection Λ_1 , the DFE E_0 is globally asymptotically stable(GSA) in Ω .

Proof. Consider the linear Lyapunov function $\dot{F} = a_0 x_2$ where $a_0 = (\alpha \mathfrak{R}_0 - \alpha)$, with derivative given as $\dot{F} = a_0 \left(\frac{\alpha x_{10} x_1}{N} - w_1 x_2 \right) \leq a_0 \alpha x_{10} \leq 0$ if $\mathfrak{R}_0 \leq 1$ for all $t > 0$ in Ω .

Since all the parameters and variables of model (A.1 - A.11) are non-negative (Theorem 2), it follows that $\dot{F} \leq 0$ for $R_0 \leq 1$ with $\dot{F} = 0$ if and only if $x_{10} = 0$. Thus, it follows, by LaSalle's Invariance Principle [36], that $x_{10}(t) \rightarrow 0$ as $t \rightarrow \infty$. Therefore $\limsup x_{10}(t) = 0$, $t \rightarrow \infty$, it follows that, given a small $\omega > 0$, there exist $M > 0$, such that $\limsup x_{10}(t) \leq \omega$ for all $t > M$. Hence using the second equation of the model (A.1 - A.11), for $t > M$, $\dot{x}_2 \leq \alpha \omega - w_1 x_2$. Thus, by comparison theorem [36, 11], $\limsup x_2 \leq (\alpha \omega) / w_1$, so that, if we let $\omega \rightarrow 0$,

$$\limsup x_2 \leq 0, t \rightarrow \infty. \quad (\text{A.22})$$

Similarly it can be shown that

$$\liminf x_2 \geq 0, t \rightarrow \infty. \quad (\text{A.23})$$

Hence it is obvious to see that

$$\lim x_2 = 0, t \rightarrow \infty. \quad (\text{A.24})$$

and following similar argument we can show that the result hold for all x_i , $i = 3, 4, 5 \dots 11$, and we have seen that

$$\lim x_1 = \frac{\Pi}{\mu}, t \rightarrow \infty. \quad (\text{A.25})$$

Thus, we say that all the solution of the model(A.18) goes to 0 except for the first equation which goes to $\frac{\Pi}{\mu}$. Hence the proof. \square

Appendix B

Table of interviews from fieldwork

Fig. B.1 Summary of in-depth interview conducted in Gombe State Nigeria

LC	TS	GD	AG	MS	NS	NK	LW	EM (€)	OC	P/R	A1	A2	HW	DH	DD	HIV	MDR	CD	SMK	Stg
Gombe	UB	F	25	MR	1	1	None	0	Wife	6/2	1	2	0	5	7	0	0	0	0	1
Gombe	UB	M	42	MR	1	5	NCE	119	Trike	22/4	1	3	0	2	6	0	0	0	0	0
Gombe	R	M	33	MR	1	5	SSCE	71	Trike	11/4	1	2	1	4	5	1	0	0	0	0
Gombe	UB	M	23	S	0	0	SSCE	47	Mason	11/5	1	2	1	8	2	1	0	0	0	0
Gombe	UB	M	22	S	0	0	SSCE	36	Mason	11/4	1	1	1	8	2	0	1	0	0	0
Gombe	UB	M	57	MR	1	11	SSCE	17	Trader	11/3	1	2	0	8	4	1	0	0	1	0
Gombe	UB	M	25	S	0	0	SSCE	24	Trader	11/5	1	2	1	8	14	1	1	0	0	0
Gombe	UB	M	40	MR	1	1	PS Cert	48	Iron bender	12/3	1	1	1	48	7	1	1	0	1	1
Gombe	UB	F	27	D	0	3	None	0	Tailoring	5/2	1	1	0	4	7	0	1	1	0	0
Gombe	R	M	32	MR	2	10	None	95	Labor	13/4	1	1	0	48	60	0	1	1	0	0
Gombe	UB	M	50	MR	1	11	None	120	Car wash	13/5	1	2	1	6	7	0	0	0	0	0
Gombe	UB	M	23	S	0	0	None	26	Pure water	10/4	1	2	1	4	2	1	1	0	1	1
Gombe	UB	M	38	S	0	0	SSCE	48	Car wash	3/1	1	2	0	24	7	0	0	0	0	0
Gombe	R	M	20	S	0	0	PS Cert	36	Trike	20/7	1	1	1	12	30	0	0	0	0	0
Gombe	UB	M	30	S	0	0	SSCE	95	Iron Bender	9/6	1	2	1	8	12	0	1	1	1	0
Gombe	R	M	25	S	0	0	None	60	Butcher	4/3	1	1	0	12	9	0	0	0	0	1
Gombe	UB	M	40	S	1	1	PS Cert	60	Iron Bender	12/3	1	1	0	8	12	0	1	1	1	0
Gombe	UB	F	27	D	0	3	None	0	Dependent	2/3	1	1	0	5	7	0	0	0	0	0
Gombe	R	M	39	MR	3	7	Diploma	286	Butcher	11/2	1	2	0	33	7	1	0	0	0	0
Gombe	R	M	70	W	0	2	None	12	Trader	4/3	1	1	1	3	4	0	0	0	0	0
Akko	R	M	58	MR	1	5	None	120	Groceries	7/2	1	1	0	20	7	1	0	1	0	0
Akko	R	M	62	MR	1	1	None	24	Famer	6/4	1	1	0	32	6	0	1	1	0	0
Akko	R	M	40	S	0	0	Diploma	83	Civil servant	8/4	1	1	1	6	3	0	0	0	0	0
Kaltungo	R	M	48	MR	1	8	JSS3	75	Mason	11/5	1	1	0	8	2	1	0	0	0	0
Akko	R	M	26	S	0	0	SSCE	36	Mason	11/4	1	1	1	154	2	1	0	1	0	0
Kaltungo	R	F	50	W	0	5	None	18	Trader	6/3	1	1	0	9	4	0	0	0	0	0
Kaltungo	R	M	25	S	0	0	None	24	Driver	11/3	1	1	0	10	14	0	0	0	0	0
Y/Deba	R	M	15	S	0	0	SSCE	NA	None	15/3	1	1	0	20	7	0	0	0	0	0
Y/Deba	R	F	19	S	0	0	PS Cert	10	Tailor	7/4	1	1	0	12	7	1	0	1	0	0
Kwami	R	F	41	MR	8	25	None	0	Wife	25/5	1	1	0	48	14	0	0	1	0	0
Dukku	R	F	48	W	9	10	None	0	Wife	35/5	1	1	0	68	7	0	1	1	0	0

Fig. B.2 Summary of in-depth interview conducted in Gombe State Nigeria

LC	TS	GD	AG	MS	NS	NK	LW	EM (€)	OC	P/R	A1	A2	HW	DH	DD	HIV	MDR	CD	SMK	Stg
Dukku	R	M	2	S	0	0	None	0	NA	60/15	1	1	0	30	7	0	0	0	0	0
Shomgom	UB	M	23	S	0	0	None	7	Pure water	40/4	1	1	0	25	5	0	0	0	0	0
Shomgom	R	F	29	MR	1	1	None	NA	Wife	5/2	1	1	0	8	7	0	0	1	0	0
Y/Deba	R	M	33	MR	1	7	SSCE	71	Trader	20/3	1	1	0	14	5	0	0	1	0	0
Nafada	R	M	22	MR	0	0	SSCE	36	Mason	10/2	1	1	0	11	7	0	0	0	0	1
Balanga	UB	M	42	MR	1	11	None	72	Trader	22/4	1	1	0	9	6	0	0	0	0	0
Balanga	R	M	33	MR	2	15	None	72	Trike	20/3	1	1	0	14	5	0	0	1	0	0
Kaltungo	R	M	25	S	0	0	Diploma	54	Restaurant	11/3	1	1	1	8	2	1	0	0	0	0
Billiri	R	M	25	S	0	0	SSCE	0	Student	40/5	1	2	1	8	5	0	0	0	0	0
Billiri	R	M	33	MR	1	2	MSC	238	Civil Servant	4/3	1	1	0	5	4	0	0	1	0	0
Billiri	UB	M	30	MR	0	17	SSCE	54	Trader	11/5	1	1	1	8	6	1	0	0	0	0
Y/Deba	UB	M	25	MR	1	1	PS cert	48	Plumber	12/3	1	1	1	18	7	0	0	0	0	0
Somgom	R	F	27	D	0	3	None	8	Trader	5/2	1	1	0	4	7	0	0	0	0	0
Dukku	R	F	60	W	0	9	0	None	Dependent	25/5	1	1	0	28	7	0	0	1	0	0
Nafada	R	M	42	MR	2	12	None	95	Labourer	16/4	1	1	0	6	7	0	0	1	0	0
Balanga	UB	M	53	MR	1	14	None	46	Farmer	17/5	1	1	1	18	7	1	0	0	0	0
Funakaye	UB	M	28	MR	2	2	None	71	Plumber	10/4	1	1	1	14	2	0	0	0	1	1
Funakaye	UB	M	38	MR	3	11	SSCE	90	Car wash	15/3	1	1	0	24	7	0	0	0	0	0
Kaltungo	R	M	20	S	0	0	None	24	Vendor	20/2	1	1	1	12	30	0	0	0	0	0
Funakaye	UB	F	30	MR	1	7	PS Cert	0	Wife	15/6	1	1	1	8	12	0	0	1	0	0
Kaltungo	R	M	25	S	0	0	None	NA	NA	10/4	1	1	0	11	10	0	0	0	0	0
Kwami	R	F	35	MR	1	6	None	NA	Wife	10/3	1	1	0	6	13	0	0	0	0	0

Key

LC: Location

TS: Type of Settlement (UB = Urban, R = Rural)

GD: Gender (M = Male, F = Female)

AG: Age

MS: Marital Status (MR = Married, S = Single, W = Widow, D = Divorce)

NS: Number of spouse

NK: Number of Children

LW: Level of Western education

EM: Estimated monthly income

OC: Occupation

P/R: Person per room

A1: Awareness on TB signs, symptoms and treatment (on a scale of 1-3)

A2: Awareness on TB transmission and spread (at a scale of 1-3)

HW: Health Care Worker

DH: Delay before hospital

DD: Delay at hospital

HIV: HIV status

MDR: Multi-drug resistance TB

CD: Chronic disease

SMK: Smoking

Stg: Stigma