ASSESSING THE SPATIAL-TEMPORAL SPREAD OF DRUG-RESISTANT TUBERCULOSIS IN KENYA

SULEIMAN HUDU ABDULLAHI

W62/6501/2016

A RESEARCH PROJECT SUBMITTED AS A PARTIAL FULFILLMENT FOR A MASTERS OF SCIENCE DEGREE IN MEDICAL STATISTICS AT THE UNIVERSITY OF NAIROBI

DECEMBER 2017

DECLARATION

In submitting this research project, I certify that this is my original work and that it has not been submitted elsewhere for a degree. I understand the rules of plagiarism and that any work done by others should be properly cited. I declare that I am the sole author of this work and other people's work has been properly acknowledged in the text.

Suleiman Hudu Abdullahi

Signed …… Date …….

CERTIFICATE OF APPROVAL

This thesis was approved for submission in partial fulfilment for the award of the degree in Masters of Science in medical statistics at the University of Nairobi.

Suleiman Hudu Abdullahi

Signed …… Date …….

Supervisors

1. Dr Hillary Kipruto, MSc, PhD
   Statistician
   World Health Organization,
   Nairobi, Kenya

Signed …… Date…….

2. Mrs Anne Wang'ombe, MSc
   Senior Lecturer
   School of mathematics
   University of Nairobi

Signed …… Date……..

DEDICATION

To my wife Salimatu and children: Umar-farouq and Abdulrahman.

To the memory of my parents.

To all patients suffering from drug-resistant tuberculosis and their caregivers.

ACKNOWLEDGEMENT

Special thanks to my dear wife Salimatu for her patient support as I laboured over this piece of work during family time.

I am grateful to my mother in-law (Mama) and father in-law (Baba) for their enduring support and understanding; to my Siblings for their support and prayers, particularly Umar. S. Hudu for his uninterrupted encouragement.

Thanks to my supervisors; Dr Hillary Kipruto and Mrs Anne Wang'ombe for their ideas and support in making this piece of work a reality; To many colleagues Abdulhafeez Adisa, Ale Boni Maxime and Andrew Mwale for making PAMAPS experience a memorable one; to Mian Anatole for his time and ideas: to Abubakar Zaki Dange for being family in Nairobi

I am highly indebted to the European Union's Intra-IACP mobility program through the Postgraduate academic mobility for African physician-scientist (PAMAPS) for providing funding for the degree.

Thanks to Dr Julius Oyugi (UNITID deputy director) for his enduring help and listening ears; to Professor M'Imunya James Machoki (UNITID director) for making the program successful; to Professor Walter Mwanda (former UNITID director) for his support during earlier months of the program; to Kenya Tuberculosis, Leprosy and Lung Disease Control Program and Kenya National Bureau of Statistics (KNBS) for proving the data used in this thesis.

ABSTRACT

Background: Tuberculosis (TB) remains one of the world biggest public health threat. Over the last three decades there has been a remarkable decline in the TB incidence and mortality globally. However, the emergence and increasing spread of a drug resistant strains of *Mycobacterium tuberculosis* is threatening to derail the effort to eradicate tuberculosis. Kenya is among the 27 high MDR-TB burden countries that account for more than 85% of estimated MDR-TB cases in the World. Also ranked 13th among the 22 high-burden nations, that collectively account for up to 80% of the global TB Cases. Drug-resistant TB is not evenly distributed across Kenya, therefore identifying the Spatio-temporal pattern and areas of high risks of DR-TB will help government prioritise resources and allow for efficient deployment of interventions that are often in limited supply to the areas where they are most urgently needed.

Objectives: The purpose of this study was to utilized spatial methods to assess and predict the spatial risk distribution of drug resistant TB in Kenya.

Study design and sites: A retrospective cohort study using longitudinal data of notified cases of drug-resistant TB from the Kenyan national DR-TB surveillance database. The study covered all the 47 counties of Kenya, using county as spatial unit of analysis.

Material and methods: Exploratory spatial data analysis (ESDA), and Bayesian spatial model were employed to estimate the spatial risk pattern of drug resistance TB using county level data obtained from the national DR-TB surveillance database for the period of five years (January, 2012 to December, 2016).

Results: Between 2012 and 2016, there has been a remarkable change in the distribution of the empirical Bayes Smoothed notification rate and excess risk of DR-TB. The EBS maps revealed a significant temporal pattern in the distribution of DR-TB cases over the five years period (2012-2016). The local Moran test for the year 2016 has identified a significant clustering of counties with high risk of DR-TB.

TABLE OF CONTENTS

LIST OF TABLES

Table **Page**

LIST OF FIGURES

Figure **Page**

LIST OF APPENDICES

Appendix **Page**

LIST OF ABBREVIATONS

AFB	acid fast bacilli
ART	antiretroviral therapy
BCG	Bacille-Calmette-Guérin
CAR	conditional autoregressive
CXR	chest X-ray
CI	confidence interval
DIC	deviance information criterion
DLTLD	division of leprosy TB and lung disease
DOT	directly observed therapy
DOTS	directly observed therapy short course
DRS	drug resistance surveillance
DST	drug susceptibility test
DR TB	drug resistant TB
EBSR	Empirical Bayes smoothed rate
EDA	exploratory data analysis
ESDA	exploratory spatial data analysis
GIS	geographic information system
GLMM	generalized linear mixed models
GPS	geographical positioning system
HBC	high burden country
HIV	human immune-deficiency virus
INLA	Integrated Nested Laplace Approximation
KNBS	Kenya national bureau of statistics
LISA	local indicators of spatial association
LTBI	latent TB infection
MCMC	Markov chain Monte Carlo
MDG	Millennium Development Goal
MDR-TB	multidrug-resistant TB
MDR/RR-TB	RR-TB cases including MDR-TB cases
NTP	national TB programme
PMDT	programmatic management of drug- resistant TB
RR	rifampicin-resistant
SDG	Sustainable Development Goal
SES	socio economic status
SSA	sub Sahara Africa
TB	tuberculosis
UHC	universal health coverage
WHO	World Health Organization
WRD	WHO-recommended rapid diagnostic
XDR-TB	extensively drug-resistant TB

DEFINITION OF TERMS

Mono-resistance: Is defined as a resistance to one first-line anti-TB drug only

Poly-resistance: resistance to more than one first-line anti-TB drug, other than both isoniazid and rifampicin

Multi-drug resistance (MDR): Is defined as a resistance to at least both isoniazid and rifampicin

Extensive drug resistance (XDR): Is defined as a resistance to any *fluoroquinolone*, and at least one of three second-line injectable drugs (*capreomycin, kanamycin and amikacin*) in addition to multidrug resistance

Rifampicin resistance (RR): Is defined as a resistance to rifampicin detected using phenotypic or genotypic methods, with or without resistance to other anti-TB drugs. It includes any resistance to rifampicin, in the form of mono-resistance, poly-resistance, MDR or XDR

CHAPTER ONE: BACKGROUND

1.0 Introduction

This chapter provides background of the study, highlighting global and national epidemiology of drug resistant tuberculosis. It also provides insight on the signs and symptoms, risk factors, diagnosis and treatment of drug-resistant TB. The anti-TB drug resistance surveillance system is briefly discussed, and the chapter ends with the highlight of the problem statement, justification and study objectives.

Tuberculosis (TB) remains one of the world biggest public health threat. It causes ill-health in millions of people every year. In 2015, it was one of the 10 leading causes of death worldwide, ranking above HIV/AIDS as one of the leading causes of death from infectious disease, with an estimated 1.8 million deaths. There were estimated 10.4 million incident cases of TB globally in the same year (WHO 2016a)

The emergence and increasing spread of drug-resistant TB continues to threaten the global tuberculosis control efforts and remains a major health concern in many countries including Kenya. The numbers of reported cases of drug-resistant TB in Kenya has been on the rise. According to WHO global TB report 2015, Kenya is one the 27 high MDR-TB burden countries that account for more than 85% of estimated cases of MDR-TB in the world. It also ranked 13th among the 22 high-burden nations, that collectively account for 80% of the global TB cases (WHO 2016a).

To achieve a universal access for diagnosis and treatment, as well as monitoring achievement in prevention and control of drug-resistant TB, there is need to strengthen surveillance systems to ensure sufficient detection and monitoring of the epidemiological profile of cases of drug resistant TB. (WHO 2010). Identifying Spatio-temporal patterns and areas of high risks of drug-

resistant TB will help Kenyan government prioritise resources and enable efficient deployment of interventions that are often in limited supply to locations where they are most urgently needed. It will also provide insight for targeted researches aimed at understanding the local drivers of emergence and spread of drug-resistant TB.

1.1 Drug resistant TB

Drug resistant TB is a form of TB infection with a strains of *Mycobacterium tuberculosis* that is resistant to at least one of the anti-TB medicines. Although, the occurrence of drug-resistance TB is being largely attributed to misuse or mismanagement TB drugs, there has been increase in the incidence of directly acquired form of drug-resistant TB. Different type of DR-TB exist and include the following; Mono-resistance, Poly-resistance, Multi-drug resistance (MDR), Extensive drug resistance (XDR) and Rifampicin resistance (RR) tuberculosis (WHO 2016a)(WHO 2015).

1.2 Risk factors for DR-TB

Although both the drug susceptible TB and drug resistant TB share most of the risk factors, several studies were conducted globally to evaluate the factors that increase the odds of developing drug resistant TB. A multi-country prospective epidemiological case control study conducted in France, Germany, Italy, and Spain has found the following factors to have significantly increased odds of developing DR-TB; use of intravenous drug *(OR 4.68)*; asylum-seeker caregivers *(OR 2.55)* as income factor; living in a nursing home *(OR 2.05)*; previous tuberculosis with pulmonary location *(OR 2.03)*; prison *(OR 2.02)*; a contacts of a known tuberculosis patient *(OR 2.01);* immunosuppression other than human immunodeficiency virus (HIV) *(OR 1.96);* acquired immunodeficiency syndrome (AIDS) *(OR 1.96)*; current tuberculosis with pulmonary location *(OR 1.77);* and health-care worker *(OR 1.69)* (Casal M, Vaquero M, Rinder H, Tortoli E, Grosset J, Rüsch-Gerdes S, Gutiérrez J 2005). Similarly, a systematic

review of published reports of risk factors associated with MDR-TB in Europe have found that, previous treatment *(OR 10.23),* been a foreign born *(OR 2.46),* younger than 65 years *(OR 2.53),* male *(OR 1.38),* and HIV positive *(OR 3.52)* to have significantly increased the risk of developing DR-TB (Faustini et al. 2006). Poor living and working conditions such as overcrowding, poor ventilation and poor practices of infection control in healthcare facilities and other congregate setting increased the risk of DR-TB transmission at the population level. Also factors that are capable of impairing the host's defence against TB infection, such as HIV infection, smoking malnutrition, diabetes, alcohol abuse, silicosis, wide range of systemic diseases, indoor air pollution and treatments with immunosuppressant are associated with high odd of developing drug-resistant TB (Lönnroth et al. 2009) (WHO 2014).

1.3 Causes and transmission of drug resistant TB

Generally there are two main pathways leading to the development of a drug-resistant TB.

i. ***Initial or primary drug resistance:*** This occurs when a person has been infected with a drug-resistant TB strain. In this case, transmission of drug-resistant TB occurs in similar way as the drug susceptible TB. It is spread from one person to another through the air. When a person with pulmonary TB cough, sneeze or spit, he propel the TB germs into the air. Another person needs to inhale only a few of these germs to become infected (Grandjean et al. 2015) (WHO 2014).

ii. ***Acquired drug resistance:*** The main cause of drug resistant TB is a mycobacterial characteristic. The mycobacteria‘s genetic machinery is programmed to mutate at a certain rate and still keep on growing, enabling them to survive otherwise effective anti-TB drugs. Other causes include;

- Health care providers: Inadequate regimen due to lack of effective guidelines
- Drugs: Inadequate supply or quality

- Patients: Inadequate drug intake for any reason.

1.4 Symptoms and Diagnosis drug-resistant TB

Early and rapid diagnosis of TB and drug resistance will have a significant benefits for the patients and public health, including increased survival, better prognosis, prevention of acquisition of further drug resistance, and reduced spread of drug-resistant TB to the vulnerable populations.

Common symptoms of active pulmonary TB are *cough with sputum and blood at times, chest pains, weight loss, weakness, fever and night sweats*. Definitive diagnosis of drug-resistant TB requires that a Mycobacterium tuberculosis bacteria be detected and resistance to anti-TB drugs determined. This can be achieved by isolating the bacteria in culture, identifying it as belonging to the Mycobacterium tuberculosis complex (MTBc), and conducting a *drug susceptibility testing (DST)* using a solid or liquid media or by performing a WHO- endorsed molecular test to detect tuberculosis DNA and a mutations associated with resistance (Lönnroth et al. 2009) WHO 2014). The following laboratory diagnostic methods are recommended for the diagnosis of DR-TB by WHO:

i. ***Phenotypic DST (conventional DST):*** The phenotypic testing determines if an isolate of Mycobacterium tuberculosis bacteria is resistant to an anti-TB drug by evaluating a growth (or metabolic activity) in the presence of drug.
ii. ***Genotypic DST (molecular DST):*** The genotypic testing detects mutations in the TB genome associated with specific drug resistance.
iii. ***The Xpert MTB/RIF assay:*** Is the only WHO-recommended rapid diagnostic (WRD) test for detection of TB and rifampicin resistance. The method achieves an equivalent sensitivity for the detection of TB resistance as culture on solid media. The *Genotype*®

MTBDR may be used as the initial test, instead of phenotypic culture-based DST to detect resistance to *fluoroquinolone*s and the second- line injectable drugs.

iv. ***Smear microscopy:*** This is a low-cost and frontline method for TB (but not drug-resistant TB) diagnosis. The main purpose of smear microscopy for drug-resistant TB is to assess the initial bacterial load, specimen triage to different diagnostic algorithms, monitor response to therapy, and to confirm the presence of acid fast bacilli (AFB) rather than contaminants in the culture media, before performing a rapid identification tests.

v. ***Clinical diagnosis:*** This is TB diagnosis based on symptoms, chest X-ray abnormalities or suggestive histology. Generally associated with low specificity, which may result in false diagnoses of TB and is not capable of detecting drug resistance.

In Kenya, patients in whom DR TB is suspected should have their sputum sent for GeneXpert as the first test. Any sample found to have DR-TB should be followed by a culture and susceptibility testing. Any patient with Rifampicin resistance is MDR-TB, and should then undergo a second-line DST.(PMDT-Kenya 2014)

1.5 Treatment of drug resistant TB

The treatment of drug resistant tuberculosis is more expensive, require longer therapy and more toxic medication. The recommended treatment for MDR-TB could either be shorter or longer treatment regimens based on specific conditions (WHO 2016b). In Kenya the treatment of MDR-TB is generally based on a standardized regimen which consist of 8 month intensive phase of 5 drugs *(Kanamycin, Prothionamide, Levofloxacin, Cycloserine and Pyrazinamide)* followed by 12months continuation phase of 4 drugs *(Prothionamide, Levofloxacin, Cycloserine and Pyrazinamide)* - ***8 Km-Pto-Lfx-Cs-Z / 12 Pto-Lfx-Cs-Z***. Individualized regimen is also used in some special situations based on individual resistance pattern of the infecting strain (PMDT-Kenya 2014). The following are the different WHO regimens for treatment of drug-resistant TB globally (WHO 2016b).

i. Shorter MDR-TB regimen for adults and children:

- In patients with RR-TB or MDR-TB who were not previously treated with the second-line drugs and in whom resistance to *fluoroquinolones* and second-line injectable agents was excluded or is considered highly unlikely, a shorter MDR-TB regimen of 9–12 months may be used instead of the longer regimens (WHO 2016b).

ii. Longer MDR-TB regimens for adults and children:

- In patients with RR-TB or MDR-TB, a regimen with at least five effective TB medicines during the intensive phase is recommended, including *pyrazinamide* and four core second-line TB medicines – one chosen from Group A (*levofloxacin, moxifloxacin, gatifloxaci*), one from Group B(*amikacin, capreomycin, kanamycin, streptomycin*), and at least two from Group C2 (*ethionamide or prothionamide, cycloserine or terizidone, linezolid, clofazimine*) (WHO 2016b)
- If the minimum number of effective TB medicines cannot be composed as given above, an agent from Group D2 (*bedaquiline, delamanid*) and other agents from Group D3 (*p-aminosalicylic acid, imipenem–cilastatin, meropenem, amoxicillin/clavulanate, thioacetazone*) may be added to bring the total to five. *Bedaquiline* is only recommended for adults; *delamanid* may also be used in patients aged 6–17 years.
- In patients with RR-TB or MDR-TB, it is recommended that the regimen be further strengthened with a high-dose *isoniazid* and/or *ethambutol* (a conditional recommendation with a very low certainty in the evidence).
- It is recommended that any patient (a child or adult) with RR-TB in whom isoniazid resistance is absent or unknown be treated with a recommended MDR-

TB regimen (either a shorter MDR-TB regimen, or a longer MDR-TB regimen to which isoniazid is added).

iii. Surgical interventions in patients with MDR-TB:

- In patients with RR-TB or MDR-TB patients, elective partial lung resection (lobectomy or wedge resection) may be used alongside the recommended MDR-TB regimen

iv. Treatment of XDR-TB:

Kenya, has adopted individualized treatment regimen for all patients with XDR-TB. While treatment should be individualized, the following empiric regimen is recommended in cases where first and second line DST is not yet available: 12 months intensive phase of ***CM-Mfx-PAS-Cfz-Amx/Clv*** and 18 months Continuation phase ***of Mfx-PAS-Cfz-Amx/Clv*** (PMDT-Kenya 2014).

1.6 Anti-TB drug resistance surveillance

The need for strengthening surveillance for drug-resistant TB has been reiterated by the *2009 World Health Assembly resolution WHA62.15* "Prevention and control of multidrug-resistant tuberculosis and extensively drug-resistant tuberculosis". The resolution urges all Member States to "achieve universal access to diagnosis and treatment of multidrug-resistant and extensively drug-resistant tuberculosis", including by means of "strengthening health information and surveillance systems to ensure detection and monitoring of the epidemiological profile of multidrug-resistant and extensively drug-resistant tuberculosis and monitor achievement in its prevention and control".

The Global Project for Anti-TB Drug Resistance Surveillance was initiated in 1994 with the aim of collecting and evaluating data on anti-tuberculosis drug resistance in a systematic and ongoing manner all over the world. Over the last 20 years this project has collected and

analysed drug resistance data from national surveillance systems and from surveys of sampled patients from 144 countries. Since 2012, the data have been published annually in the Global Tuberculosis Report (WHO, 2016b).

In Kenya, routine drug resistant surveillance based on culture and DST testing particularly, among patients at risk of drug resistance has been in existence for nearly two decades. Although, the system was interrupted for some years due to some logistic challenges, it was revived and been fully functioning since 2003. To achieve the ambitious targets of ending global TB epidemic by 2035, there is need for a robust surveillance system capable of ensuring timely, accurate and complete data, and free from the conventional challenges of a paper based system. In 2012, Kenya TB program develop a unique nationwide electronic data transmission that would ensure data of highest quality. This new system was named "TIBU" standing for "Treatment Information from Basic Unit" and also means "Treat" in Kiswahili language. Since its inception, the database has been able to manage more than 370,000 TB cases including hundreds of drug-resistant TB cases (WHO, 2016a) (Sitienei, 2016).

Despite the implementation of a robust TB information system TIBU, Kenya is still far from the ultimate goal of directly measuring TB incidence from case notification. To achieve this, combination of strengthened surveillance and better quantification of under-reporting (the number of cases missed by surveillance systems) has become essential, hence, methods that provide indirect estimate such as spatial modelling has become handy. Spatial analysis has the ability to identify problem locations that might be difficult to detect otherwise. Additionally, the recent deployment 70 GeneXpert MDR/RIF machines, 5 culture laboratories and 150 LED microscopes operating within a network of 1,860 AFB microscopy sites (1:25,000 population) are among the government effort to strengthen the routine drug resistant surveillance system (Sitienei, 2016) (NTLD-Program 2014).

1.7 Statement of the problem

Tuberculosis (TB) remains one of the world biggest public health threat. Over the last three decades there has been a remarkable decrease in TB incidence and mortality globally. However, the emergence and increasing spread of a drug resistant strains of *Mycobacterium tuberculosis* was threatening to derail the effort to eradicate tuberculosis. There were an estimated 580 000 incident cases of MDR/RR-TB in 2015, with cases of MDR-TB accounting for 83% of the total. Kenya is among the 14 countries enlisted as having a high burden of TB, MDR-TB and HIV/TB (WHO, 2016a). Drug resistant tuberculosis is more expensive to manage and has great potential to bankrupt patients largely due to more complicated and lengthy therapy involved and inability of the patient to go to work (Zumla et al. 2012). Its management require longer therapy and more toxic medication. Furthermore, only 50% of the patients on MDR-TB treatment were successfully treated globally. This is largely due to high rates of mortality and loss to follow-up (WHO, 2016a) (Chang & Yew 2013).

Spatial analysis can assist in identifying problem locations that might have been difficult to identify otherwise (Zhang et al. 2014), and will help Kenya government achieve effective monitoring and successful implementation of control and prevention programmes for TB/MDR-TB.

1.8 Justification

The interest in mapping the spatial distribution of drug-resistant TB is motivated considering the fact that the incidence rates of drug-resistant TB is not evenly distributed across Kenya.

Detecting the spatial pattern will provide possible insight on the influence of environmental and sociocultural factors on the distribution of drug-resistant TB. This will help in a successful implementation of prevention and control programs for drug-resistant TB. Exploratory spatial

data analysis (ESDA) methods and spatial modelling are emerging as useful approaches to achieve this understanding (Associates 2002). Although there were few literature on spatial epidemiology of TB in Kenya (Kipruto et al. 2015), no such information on the spatial distribution of drug resistant TB. Additionally, the study will provide insight for targeted researches aimed at understanding the local drivers of emergence and spread of drug-resistant TB.

1.9 Study hypothesis

There is no spatial-temporal correlation on distribution of drug-resistant TB in Kenya.

1.10 General objective

The study aims to evaluate the spatial-temporal distribution of drug-resistant TB in Kenya

1.11 Specific objectives

1. To determine the distribution of notified cases of drug-resistant TB in Kenya
2. To evaluate the trends of spatial correlation for the distribution of drug-resistant TB in Kenya
3. To model the effect of covariate on Spatio-temporal distribution of county notified cases of drug-resistant TB in Kenya

CHAPTER TWO: LITERATURE REVIEW

2.0 Introduction

This section reviews existing literature on GIS and spatial analysis in health research, spatial epidemiology of TB/DR-TB and Bayesian model. It looks at what has been done by other researchers on these topics.

2.1 GIS and Spatial Analysis in Health Research

Although GIS have been in existence in for several decades, a combination of several factors has reinforced its development and emergence as multidisciplinary field in 1970s. Its rapid transformation from what had been an expensive, slow, and difficult, to inexpensive, fast, and easy to use has open the door for its application in several disciplines including health, forestry, transportation planning, hazard planning, marketing, archaeology, surveying and criminal justice (Clarke et al. 1996).

The concept of space in epidemiology can be traced back to the 4th century B.C. Hippocrates was among the first to relate disease occurrence to the local environment in his book Air, Water, and places. In 1854, John Snow examine a classical example of how geographical space relates to the occurrence of diseases during London cholera outbreak investigation (Gotway 2004). Over the last three decades GIS has increasingly been used in the field of epidemiology and public health. The development of more refined geographic information systems (GIS) has provided a new set of tools for public health professionals. It helps them pinpoint cases and exposures, identify spatial trends and disease clusters, correlate different sets of spatial data, and test statistical hypotheses. It also enable researchers to locate high prevalence locations and populations at risk, identify areas in most need of resources and decisions on resource allocation (Carroll et al. 2014) (Clarke et al. 1996).

Like GIS, the spatial analysis is increasingly being utilized in many disciplines including health (Messner et al. 1999) (Zenk et al. 2005). It enable researchers to explore both the spatial interaction (dependence) and the spatial structure (heterogeneity) linking risk factors (covariates) to outcome rates.

Exploratory spatial data analysis (ESDA) is been considered as one of the powerful spatial analysis tools aimed at improving our understanding of the data. As collection of statistically robust techniques, ESDA enable researcher to describe and visualize spatial distributions, identify atypical locations (spatial outliers), discover patterns of spatial association (clusters or hot spots) and suggest spatial regimes or other forms of spatial instability (spatial heterogeneity) or spatial non-stationarity (Anselin, 1994, 1996 1998, 1999) (Wise et al. 1997) (Anselin 1995).

Epidemiologists are increasingly incorporating ESDA into health-related research for both communicable and non-communicable diseases. For example Empirical Bayes smoothing, Moran's I statistic and the Local Indicator of Spatial Association (LISA) statistic have been used to assess spatial patterns of low birth weight prevalence and spatial clusters in the State of Georgia, USA. The researchers found significant geographical variation in LBW prevalence (Tu et al. 2012). For more use of ESDA in non-communicable see (Tian et al. 2002) (Shrestha et al. 2012).

On the other hand, Fulcher et al employed ESDA to examine spatial distribution of HIV service providers in Toronto (Fulcher & Kaukinen 2005). Also Hendricks et al. examine the spatial spread of Lyme disease in both high incidence and low incidence states of the USA using yearly county-level case report numbers for probable and confirmed Lyme disease cases. Local indicators of spatial autocorrelation (LISA) and spatial empirical Bayesian smoothed rates found significant clustering in some states (Hendricks et al. 2016).

2.2 Spatial epidemiology of TB and drug-resistant TB

Although efforts were made by researcher to understand the spatial epidemiology of TB globally (Tiwari et al. 2006) (Álvarez-Hernández et al. 2010) (Couceiro et al 2011) (Harling et al 2014) (Nunes, 2007), only handful of such studies were conducted in Africa. In South Africa, analysis of point pattern and spatial statistics were used to identify clustering of TB cases in the areas of high incidence. The researchers found significant association between number of notified TB cases and factors such as unemployment, overcrowding and number of *shebeens* (local drinking places) per enumerator sub-district (Beyers et al. n.d.). Also in a multi-country study Moran's I statistics and Poisson regression model were used to explore the spatial and temporal patterns of TB distribution in Africa. The researchers found significant clustering of cases and suggested that 25 Africa countries were at increased risk of tuberculosis, and ten countries could be grouped as "hot spots" (Uthman et al 2008). In another study, researchers used GPS coordinates, clinical and demographic characteristics of patients attending chest clinic between 03/2007 and 02/2008 to identify spatial pattern of TB among permanent residents of Greater Banjul Gambia.

In Kenya, very few literature on the application GIS and spatial analysis on TB epidemiology exist. For example; Sitienei et al evaluated the effect of risk factors on the spatial temporal distribution of TB using data obtained from randomly sampled patients in 3 Kenyan provinces (Nairobi, Nyanza & Rift valley). The study found significant clustering in Nairobi and 7 districts of Nyanza province which demonstrated a higher median relative risk than the rest of the districts(Sitienei 2014). In another study, Kipruto et al. 2015 applied the concept small area estimation to assess spatial and temporal distribution of TB, using 47 Kenyan counties as spatial reference units. The researchers found a significant clustering of TB cases in a number of counties over a period of 3 years (2012-2014), with 11 counties (*Marsabit, Isiolo, Makueni,*

Nairobi, Lamu, Mombasa, Machakos, Kajiado, Siaya, Kisumu and Homa bay) appeared to have the highest estimated risk of case notification rates per 100,000 (Kipruto et al. 2015).

While handful studies have used GIS and spatial analysis to describe the pattern of drug-resistance TB globally (Jenkins et al. 2013) (Liu et al. 2011) (LIN et al. 2011), none of such studies was carried out in Kenya and Africa at large. For example, In Moldova researchers analysed national TB surveillance data collected between 2007 and 2010 to assess spatial variation in MDR-TB. A significant geographical variation in MDR-TB burden and hotspots was identified. It was found that locations with a higher proportion of previously incarcerated TB cases were at greater risk of being MDR-TB hotspots (Jenkins et al. 2013).

Even though few researchers have tried to evaluate the spatial epidemiology of TB in Africa, the literature on spatial epidemiology of DR-TB is meagre and possibly non-existence. Therefore, the goal of my study is to focus on application of ESDA and spatial model to evaluate the spread of drug resistant in Kenya. The research will take into consideration the impact of factors such as social deprivation index (includes information on income level, school education, housing quality, overcrowding), unemployment rate, proportion of migrant population, and other socio-demographic factors on the spatial distribution of drug resistant TB in Kenya.

2.3 Bayesian spatial model in disease mapping

The Bayesian model is increasingly being utilized in various fields including epidemiology, business, ecological studies and social sciences. This is largely due to its ability to account for uncertainties (Bivand et al 2008). Mapping spatially aggregated count data of diseases often show a noisy pattern which makes interpretation difficult. To overcome this problem, the Besag-York-Mollie (BYM) model is being increasingly employed in epidemiology. BYM model is the most robust and most commonly used Bayesian Hierarchical Model (Besag et. al.

1991) (Haining, 2003). It is computationally efficient conditional autoregressive (CAR) model that incorporate both spatially structured and spatially unstructured random effects to produce results in a smooth risk surface and prediction variance that changes mainly as a function of the predicted risk (Pascutto et al.2000) (Ma et al. 2007) (Latouche et al. 2007) (Auchincloss et al. 2012).

BYM model is highly computational and this make parameter estimation difficult. To overcome this challenge two approaches namely Markov chain Monte Carlo simulation (MCMC) and the Integrated Nested Laplace Approximation (INLA) have been used for Bayesian inference. Basically, MCMC methods generate a simulations of parameters of the model which after a suitable burn-in period, become realisations of posterior distributions (Roberts 1996). The MCMC methods, however, involve more computationally and time intensive simulations to obtain the posterior distribution for the parameters. To overcome this, the INLA approach has recently been developed as efficient and a reliable estimations method with lower computational time than MCMC (Rue et al. 2009). The use of INLA in BYM model has increased exponentially in the recent time.

CHAPTER THREE: METHODOLOGY

3.0 Introduction

This chapter describes the study design, study area and populations. It also provides detail of the statistical methods, software, data source and limitation of the study. It ends with ethical considerations and study timelines.

3.1 Study area

The study covered all the 47 counties of the republic of Kenya, using county as spatial unit of analysis.

Kenya is a country in Africa and a founding member of the East African Community (EAC). Its capital and largest city is Nairobi. Kenya's territory lies on the equator (between latitude 5^{o} north and 5^{o} south and between 24^{o} and 31^{o} east longitude) and overlies the East African Rift covering a diverse and expansive terrain that extends roughly from Lake Victoria to Lake Turkana (formerly called Lake Rudolf) and further south-east to the Indian Ocean. Kenya is bordered by Tanzania to the south and southwest, Uganda to the west, South Sudan to the north-west, Ethiopia to the north and Somalia to the north-east. Kenya covers 591,971 sq. kilometre (228561 square mile) out of which 580,609 km is land area and 11,362 km water area. The population estimate was approximately 43 million people in 2014 (KNBS).

Fig. 1 Kenya administrative boundaries *(Source: https://www.knbs.or.ke/downloads)*

3.2 Study population

The target population for this study will include all the notified cases of drug-resistant TB captured in the national DR-TB surveillance database for the period of five years (2012-2016), and who meet the study criteria. Five years duration was chosen considering the fact that the counts aggregated over a small number of years will prevent extreme heterogeneity (Messner et al. 1999).

3.2.0 Population characteristics

3.2.1 Inclusion criteria

- All notified drug-resistant TB cases diagnosed using methods recommended by WHO
- Both newly diagnosed and those who developed resistant while on treatment
- Patient on treatment with second line and injectable anti TB treatment

3.2.2 Exclusion criteria

- Patient susceptible to all first line anti-TB drugs

3.3 Data source

The study extracted and utilized the data of all the notified cases of drug-resistant TB captured in the Kenyan national TB surveillance database (TIBU). Since its inception in 2012, the system has been collecting a high quality data all over Kenya, from 4,325 peripheral health facilities, 270 basic management units and 12 regions. To date, the database has been able to manage more than 370,000 TB cases including hundreds of drug-resistant TB cases, and capture patient level data including demographic, laboratory and clinical data.

The most recent population estimate, administrative boundaries, socio-demographic data was obtained from Kenya national bureau of statistics (KNBS).

3.4 Data source verification

Efforts were put in place to ensure that the data obtained is verified to be true. Six counties were randomly selected as sample of the 47 counties and the data collected from them was verified.

3.5 Study variables

The variable of interest include:

i. ***Demographic;*** County, registration number, Age (years), Sex (M/F)
ii. ***Socio-economic Status (SES);*** Occupation, education and income, Social deprivation index (SDI), proportion of migrant population
iii. ***Clinical profile;*** Height, weight, nutritional support, HIV status, type of patient (new DR-TB or previously treated)
iv. ***Laboratory profile;*** Type resistance (Mono resistant, RR, MDR, XDR, Others), method of diagnosis (geneXpert or culture)

3.6 Study design

A retrospective cohort study using a longitudinal data of notified cases of drug-resistant TB from the Kenyan national DR-TB surveillance database for the period of 5 year (2012-2016).

3.7 Sample size determination

The risk of DR-TB will be modelled using multiple regression. The study utilised the entire cases of drug resistant TB captured in the national drug resistant surveillance database (2012-2016).

3.8 Statistical Analysis

The analysis begin with extraction of data from the Kenyan national DR-TB surveillance database and imported to the Microsoft Access 2013. The data was cleaned and variables coded accordingly using the latest version of R statistical software. Below is the schematic presentation of the entire analysis process conducted in a successive passion.

Spatial Analysis process

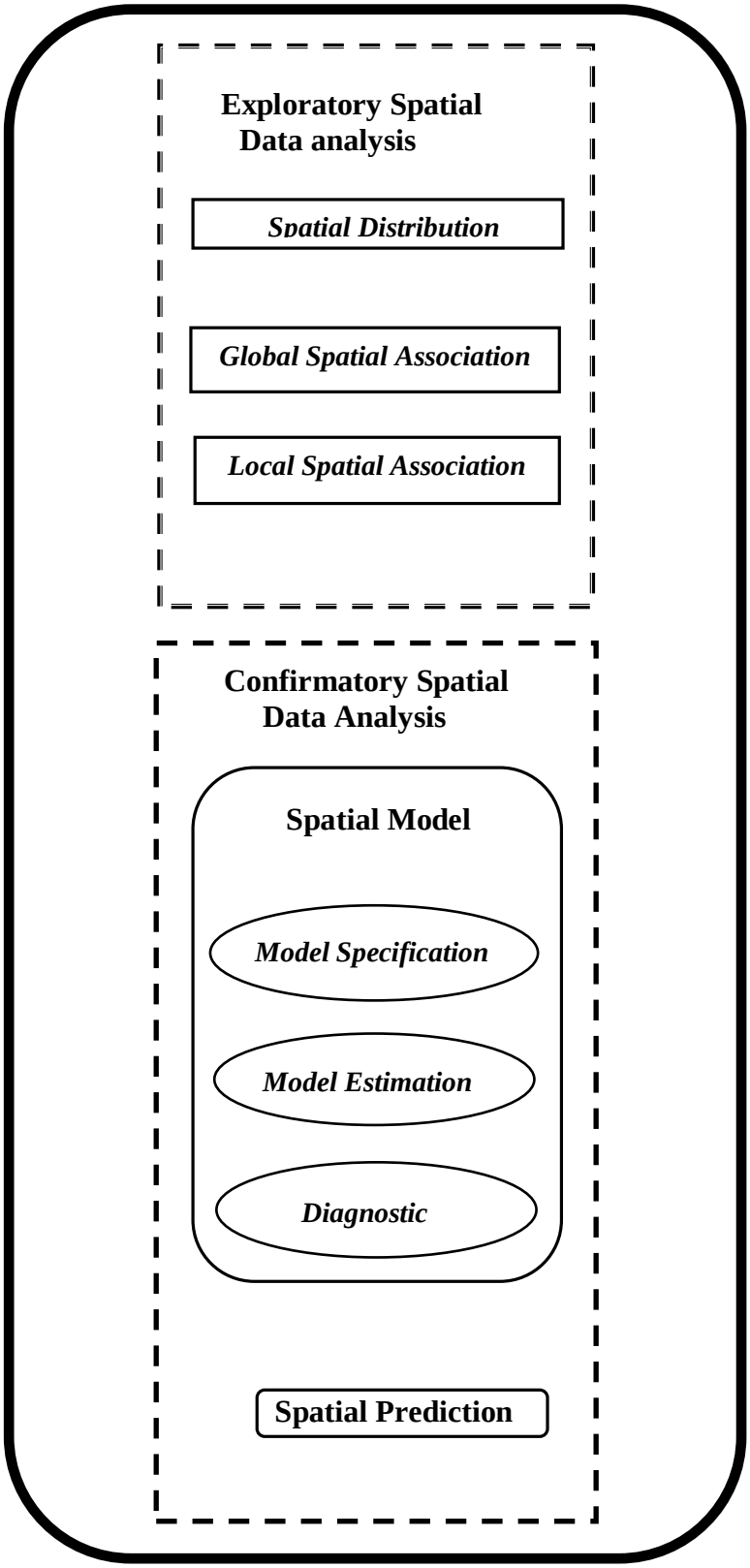


Fig. 2: Spatial analysis process (*Source: L. Anselin 1999****)***

3.8.1 Exploratory data analysis (EDA)/basic descriptive

After data cleaning and coding, series of basic descriptive analysis were performed to evaluate non-spatial characteristics of the data using the latest version of R statistical software. The mean, median, range, quartiles and histograms were used for continuous variables, while frequency counts, percentages, tables (one-way and two-way), bar charts and pie charts were used for categorical variables.

3.8.2 Exploratory spatial data analysis (ESDA)

The exploratory spatial data analysis was used to evaluate the "spatial" aspects of the data to enable identification of local patterns of spatial association. It identified spatial clustering, and spatial outliers of cases of drug-resistant TB among the counties (Zhang et al. 2014). Global and local measure of spatial autocorrelation was evaluated. The maps of Excess risk and Empirical Bayes smoothed notification rate DR-TB cases were presented. The latest version of *GeoDa* spatial analysis software was used for the ESDA.

Raw Rate and Excess Risk Maps: The ESDA began by comparing maps of RR and ER. The excess risk is one of the outlier maps used for exploration of outliers in a *choropleth* map depicting rates or proportions. Both excess risk map and raw rate map are purely for visualization and does not indicate whether or not the observed excess is “significant” in a statistical sense (Anselin et al. 2003).

Spatial Relative Proximity: Essential to the analysis of areal data is the concept of *spatial weights matrix* (W) which expresses the spatial arrangement (topology) of the data. Here each *spatial weight* w_{ij} typically reflects the “spatial influence” of county ***i*** on county ***j.*** Various spatial weight matrices often used in practice, however, recognizing Kenya’s topology, a contiguity weights, where counties are considered neighbours only when they share a border

was used in this analysis. Where; $\mathbf{w_{ij}}$ = weight between county ***i*** and ***j***, $\mathbf{w_{ij} = 1}$ when counties share border, $\mathbf{w_{ij} = 0}$ when they do not share border (Smith n.d.) (Haining, 1990; Upton & Fingleton, 1985)

Empirical Bayes Smoothing: Shading disease rate on a regional map ***(choropleth mapping)*** using a **"raw" rates** may provide a misleading impression of the underlying **"true risk".** This is due to the fact that when rates are estimated from unequal populations (such as widely varying county populations), the results are inherently unstable as a result of high degree of variance instability between spatial units. For example, an unusually high or low rate for a county may be due to very few population at risk from one year to the other, so that the chance fluctuations may cause the rate to be high in one year and low in the next. Therefore, in order to compare counties spatially, allowance must be made for non-constant spatial variability of rates. *The Empirical Bayes smoother* takes into account the spatial inhomogeneity of variances (the effect of the varying population at risk), as well as any spatial dependence between counties. It uses *Bayesian principles* to guide the adjustment of the raw rate estimate by taking into account information from the surrounding locations. The principle is referred to as ***shrinkage***, in the sense that the raw rate is moved *(shrunk)* towards an overall mean, as an inverse function of the inherent variance (Cressie 1992) (Gotway 2004).

Bayesian method drive a ***posterior*** (the current state about certain event) by using ***prior*** (prior knowledge of the event) and the ***likelihood*** (our data). It treats a disease risk x_i as a random variable with *a prior mean* $\boldsymbol{\tau}$ and *prior variance* $\boldsymbol{\gamma}$. The Bayes estimator provides an approach that ***borrows strength*** from the prior mean, where the amount of strength borrowed depends on the stability of the crude local estimate as measured by the prior variance. The Empirical Bayes Estimator $\boldsymbol{\Psi}$ can be expressed mathematically as;

$$\Psi = \phi + \omega\ (r_i + \phi)$$

Where;

ϕ – method-of-moments estimator of the overall mean (weighted sample mean)

ω - method-of-moments of the Bayes shrinking factor

r_i – the crude of the event of interest (DR-TB rates)

Global Spatial Autocorrelation: The global spatial autocorrelation evaluate the average measure of spatial association across the entire country. It produce a single value which applies to the entire data set to indicate the presence or absence of a stable pattern of spatial autocorrelation for the whole country.

Moran I Statistics (I) will be used to estimate the global spatial autocorrelation.

$$\mathbf{I} = \frac{N}{\sum i \sum j w_{ij}} \frac{\sum i \sum j w_{ij}(X_i - \bar{X})(X_j - \bar{X})}{\sum i (X_i - \bar{X})^2}$$

Where;

N-number of spatial units (number of counties),

X-variable of interest (county DR-TB rates)

$\bar{X}$ - mean of X

W_{ij} - spatial weight assigned to county i and j (element of spatial weight matrix).

Local Indicators of Spatial Association (LISA): The LISA based on local Moran statistics is an effective way of evaluating local spatial autocorrelation. It allow for the decomposition of global indicators, such as Moran's I, into the contribution of each observation (measure of spatial association for each county). The LISA statistics serve two purposes. On one hand, they

may be interpreted as indicators of local pockets of non-stationarity, or hot spots. On the other hand, they may be used to evaluate the influence of individual locations on the magnitude of the global Moran's I and to identify "outliers" (Anselin 1995) (Anselin et al. 2003). Series of tools namely; LISA statistics, LISA Cluster map, Moran Scatterplot and Moran significance map were employed to detect counties or set of contiguous counties for which the LISA statistics is significant. A random permutations as high as 9999 was generated to determine the significance and sensitivity of the local Moran statistic.

LISA Statistic (I_i)

$$\mathbf{I_i} = \frac{(\mathbf{Xi}-\bar{X})}{\sum \mathbf{i}(\mathbf{Xi}-\bar{X})\mathbf{2}} \sum \mathrm{jwij}(\mathrm{Xj}-\bar{X})$$

LISA Cluster map: The LISA cluster map also called ***Moran scatterplot map*** provide a classification of counties base on their varying pattern of DR-TB rate. The map incorporates information from the individual counties, neighbouring counties, and the global average of the outcome measure (rate of DR-TB). This is an especially effective way of identifying outliers.

Moran Scatterplot: A Moran I scatterplot provides more information on the LISA cluster map. Its main contribution is classification of the type of spatial autocorrelation into two categories, called spatial clusters and spatial outliers. Conventionally, the scatterplot has the variable of interest (rate of DR-TB) on the X-axis and the spatially lagged values of variable of interest on the Y-axis. The four quadrant of the Moran scatterplot corresponds to different type of spatial correlation; low–low, low–high, high–low, and high–high. The global Moran's I (Moran's I coefficient) is the slope which can be used to indicate degree of fit, presence of outliers and leverage points. Generally Moran scatterplot may help get insight on the following aspect of spatial association (Associates, 2002) (Anselin, 1996).

i. Decomposing spatial association into four component: low–low and high–high as positive association and low–high and high–low as negative association.
ii. Identifying observations that are outliers relative to the global measure of spatial autocorrelation given by Moran's I
iii. Finding observations that exact a large influence (leverage) on the Moran coefficient
iv. Suggesting problem with the specification of the spatial weight matrix.

Moran significance map: The Moran significance map incorporates information on statistically significant clusters and is builds on Moran scatterplot. Up to 9999 random permutations were generated to determine the significance of the local Moran statistic.

3.8.3 Bayesian spatial temporal model

The outputs of the Bayesian modelling are in form of probability distributions, termed predictive posterior distributions, which represent the probability of a parameter of interest taking values from within a plausible range of values. This inferential framework has significant important practical implications in risk mapping because posterior predictive distributions can be derived for both the parameters (which include the spatial autocorrelation parameters and the coefficients of covariates) and the epidemiological outcome of interest (e.g. prevalence or intensity of infection) at un-sampled locations.

The Besag-York-Mollie (BYM) model, a computationally efficient CAR-model that incorporate both the spatially structured and spatially unstructured random effects to produce results in a smooth risk surface and prediction variance that changes mainly as a function of the predicted risk was used (Besag et. al. 1991).

The Bayesian inference of latent Gaussian models is highly computational. For this reason two approaches namely Markov chain Monte Carlo simulation (MCMC) and the Integrated Nested

Laplace Approximation (INLA) are used for Bayesian inference. The MCMC methods involve more computationally and time intensive simulations to obtain the posterior distribution for the parameters. To overcome this challenge, the INLA approach has recently been developed as efficient and a reliable estimations method with lower computational time than MCMC (Rue et al., 2009). Thus, the model parameter for this study was estimated using R-INLA.

The BYM model can be expressed as;

$$\boldsymbol{Log(\mu_{it}) = Log\,(E_{it}) + \beta_0 + \sum_{j=1}^{j=k} \beta_j X_{ijt} + u_i + v_i + \omega_t}$$

$\boldsymbol{u_i}$ - spatially unstructured random effect

$\boldsymbol{v_i}$ - spatially structured random effect

$\boldsymbol{\omega_t}$ - temporal random effect

$\mathbf{X_{ijt}}$ - $\mathbf{j^{th}}$ covariate for county i at time t

$\boldsymbol{\beta_j}$ - parameter vector of the covariates $\boldsymbol{X_{ijt}}$

$\boldsymbol{\beta_0}$ - model intercept

$\boldsymbol{E_{it}}$ is the expected number of drug-resistant TB cases in county ***i*** at time ***t, i*** ranges from 1 to 47 whereas t = 1,2,3,4, 5 representing the years 2012, 2013, 2014, 2015 and 2016 respectively.

A non-informative prior for the covariate parameter vector i.e. $\beta j \sim (-\infty,+\infty)$

Spatially unstructured random effects assumed to be normally distributed i.e. $ui \sim Normal\,(0,2$) whereas spatially structured random effects will be assigned a conditional autoregressive prior i.e. $vi \sim CAR(\delta v2)$, and the corresponding precision parameters given non-informative gamma distributed priors.

The time trend will be of first order random walk i.e. $\omega t \sim Norm(\omega t-1,\delta\omega 2)$ and $\omega 1 \sim Normal(0,\delta\omega 12)$.

3.9 Software and tools

The statistical analyses was carried out using the latest version of the following software;

i. *R statistical software:* Data cleaning and EDA and Spatial modelling
ii. *GeoDa:* ESDA

3.10 Confidentiality and data security

The study sought ethical approval from Kenyatta National Hospital-University of Nairobi ethical review committee (KNH-UoN ERC). Data confidentiality and disposal of information was guided in accordance with the data ownership policy.

3.11 Ethical considerations

The study was conducted in compliance with the ethical principles and guidelines for the protection of human subjects of research. Data acquisition was initiated after ethical approval obtained. The ethical clearance for the study was sought from the KNH-UoN ethics review committee and the Kenya national tuberculosis, leprosy and lung diseases program. Considering the fact that the study used a secondary data, a waiver of consent from the patients was sought from the ERC. Confidentiality of patient's information was maintained, no patient identifier information extracted from the database.

3.12 Study results dissemination plan

The final copy of this document will be submitted to University of Nairobi Institute of Tropical and Infectious diseases (UNITID) and a manuscript will be send to one of the peer review journal for review and publication.

3.13 Data collection tool

The study used a secondary data from the Kenya national TB surveillance database also known as "TIBU". The data was extracted and stored in MS Excel 2013, hence, study doesn't require

a primary data collection tool since the standard surveillance tools where used to capture the data electronically.

CHAPTER FOUR: STUDY RESULT

DR-TB patient characteristics and distribution: A total of 1677 cases of DR-TB were notified and captured in the national DR-TB surveillance database over the period of five years (2012-2016). There has been a significant increase in the number of notified cases of DR-TB from 2012 to 2016 with a marginal decrease in 2016. The number of new cases has risen significantly since 2014. Figure 3 presents the trend of DR-TB cases over years.

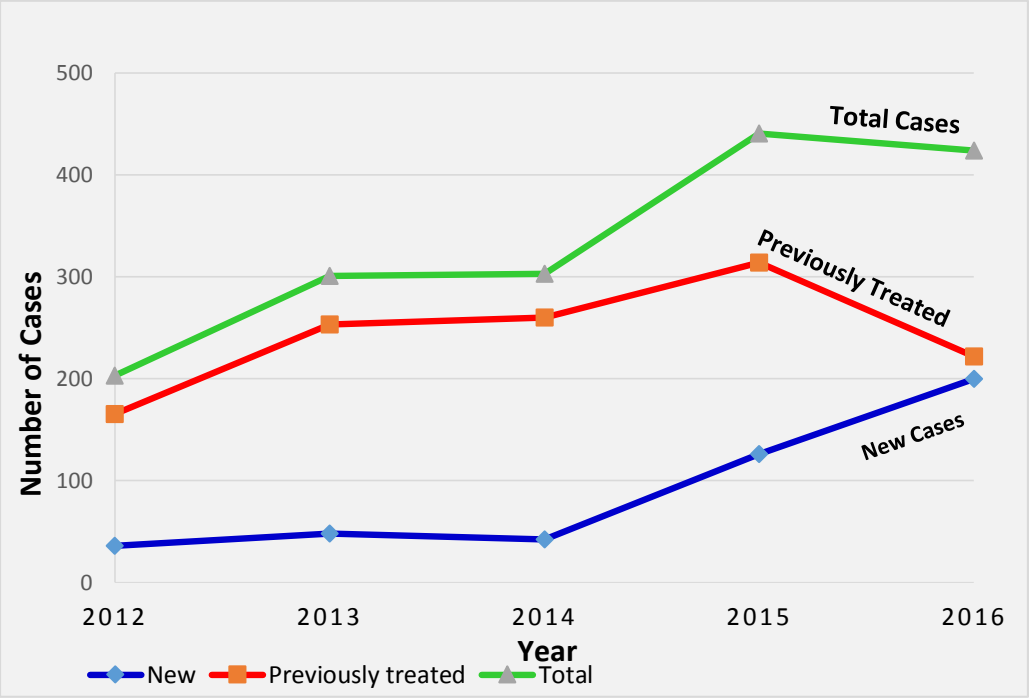


Figure 3: Distribution of New and Previously Treated case of DR-TB 2012-2016

The cases were unevenly distributed over the 47 Kenyan counties with the exception of Mandera and Wajir that reported no cases of DR-TB over the five years period. The patients cut across all age groups from 8month infant to 99 year old, with majority from productive age groups. Five cases aged less than 3 years were excluded from the analysis. Figure 4-5 and table 1 presents the summary of age distribution. More male (62.5%) were notified than female (37.5%). Figure 6 and 7 shows the patient's sex.

Table 1: Age distribution

	Minimum	Median	Mean	Maximum
Age (Years)	3	33	34.09	99

Figure 4: Patient age distribution

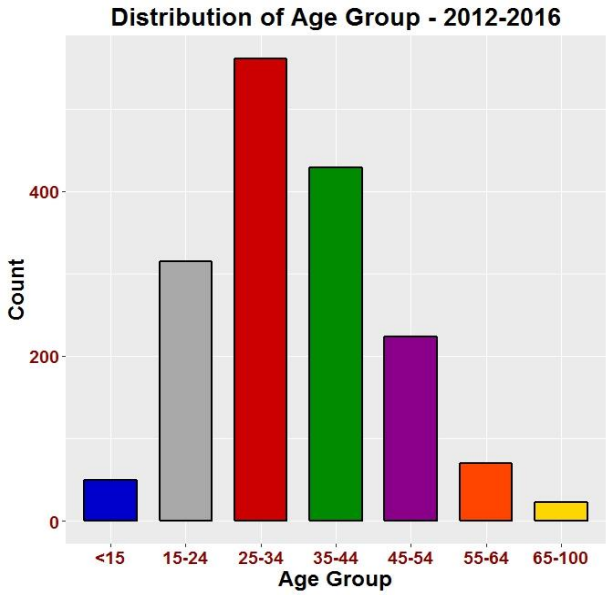


Figure 5: DR-TB distribution by age group

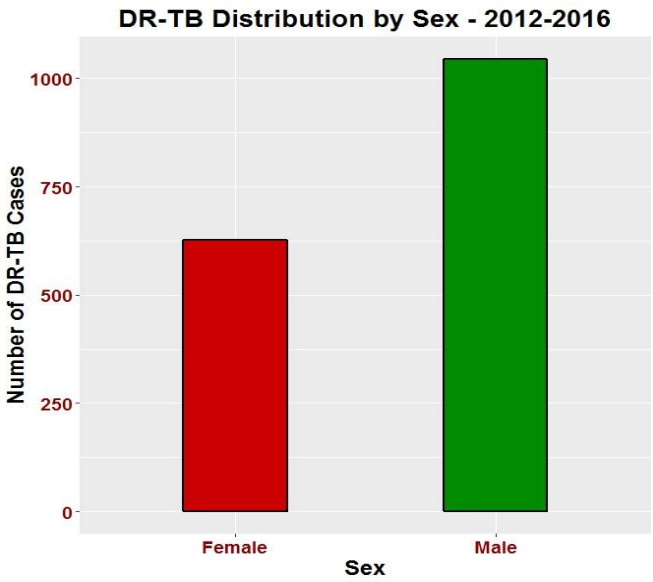


Figure 6: DR-TB distribution by sex 2012-2016

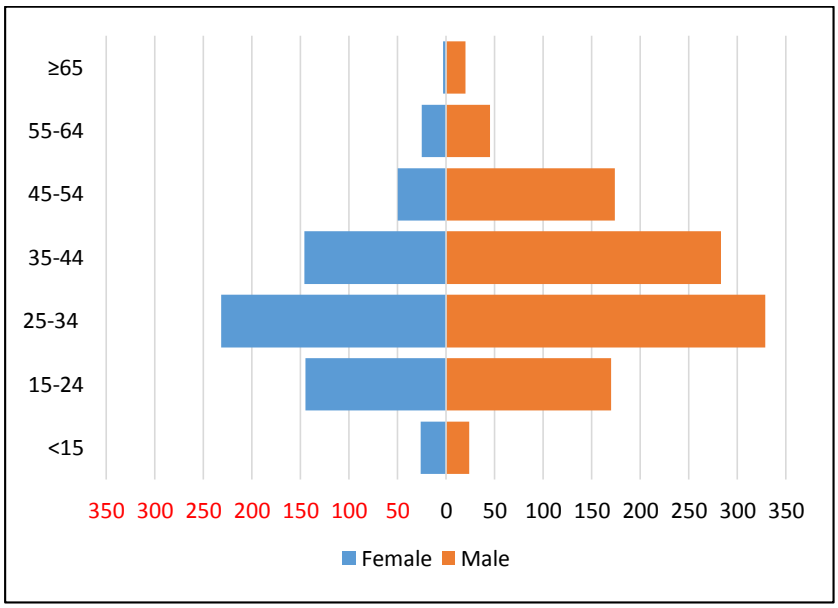


Figure 7: Notified DR-TB cases by age group and sex 2012-2016

DR-TB/ HIV: One third of the patients (37.2%) has HIV co-infection. The mortality among DR-TB/HIV patients is 19.9% compared to 6.2% among the non HIV patients. Table 2 and figure 8 summarises the treatment outcome among HIV negative and positive patients.

Table 2: Treatment outcome for HIV positive and HIV negative DR-TB cases

HIV status	Treatment Outcome						Total
	Cured	*Died*	*Defaulted*	*Tx completed*	*Not evaluated*	*Failed*	
HIV Positive	112***(18%)***	124***(19.9%)***	27***(4.3%)***	91***(14.6%)***	268***(43.1%)***	0	**622**
HIV Negative	406***(38.8%)***	65***(6.2%)***	42***(4%)***	175***(16.7%)***	356***(34%)***	2***(0.2%)***	**1046**

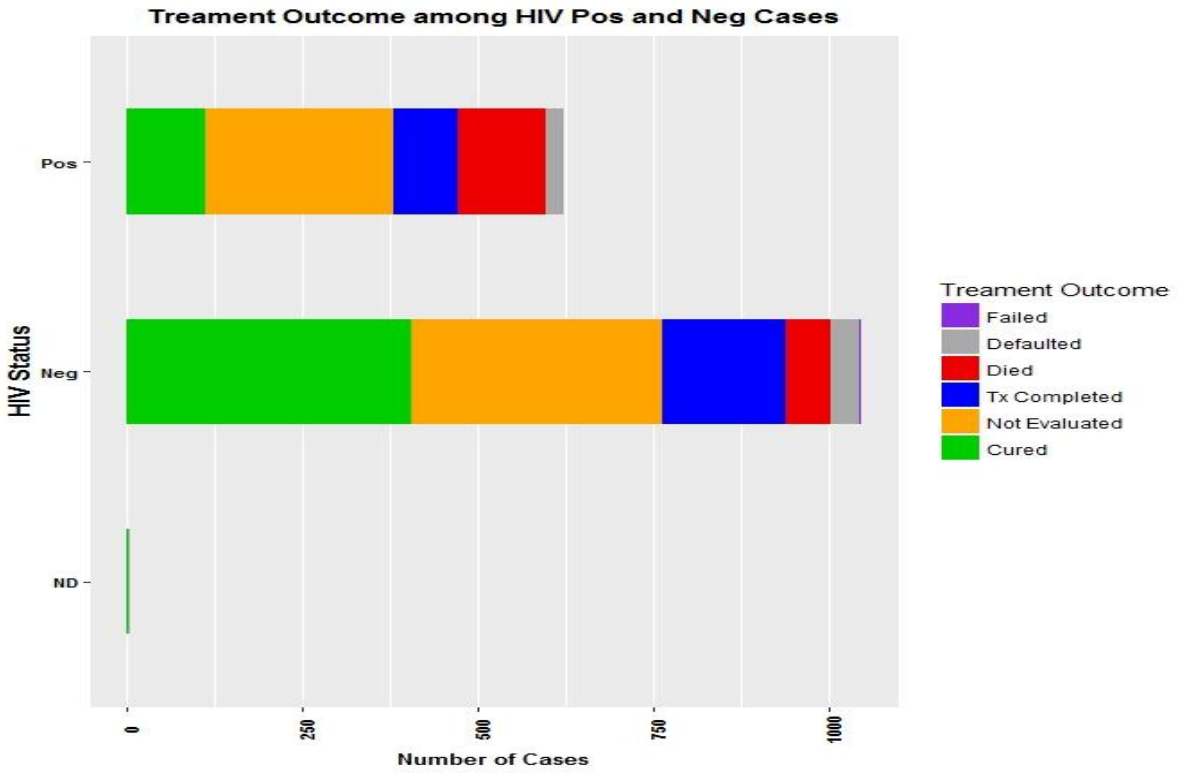


Figure 8: Treatment outcome among HIV positive and HIV negative DR-TB cases

Distribution of DR-TB notification rates from 2012-2016: Over the last five years, there has been a remarkable change in the distribution EBS notification rate and excess risk of DR-TB. The number of counties with marginal increase in EBSR has increased during this period, Table 4. For instance, twelve (25.5%) counties has EBSR ≥ 1.14 per 100,000 population in 2016 compared to five (10.6%) and nine (19%) counties in 2014 and 2015 respectively. On the other hand, number of counties with EBS notified rate < 0.5 has decrease significantly. The map of the Empirical Bayes smoothed notification rate revealed a remarkable temporal pattern between 2012 and 2016. Figure 9.

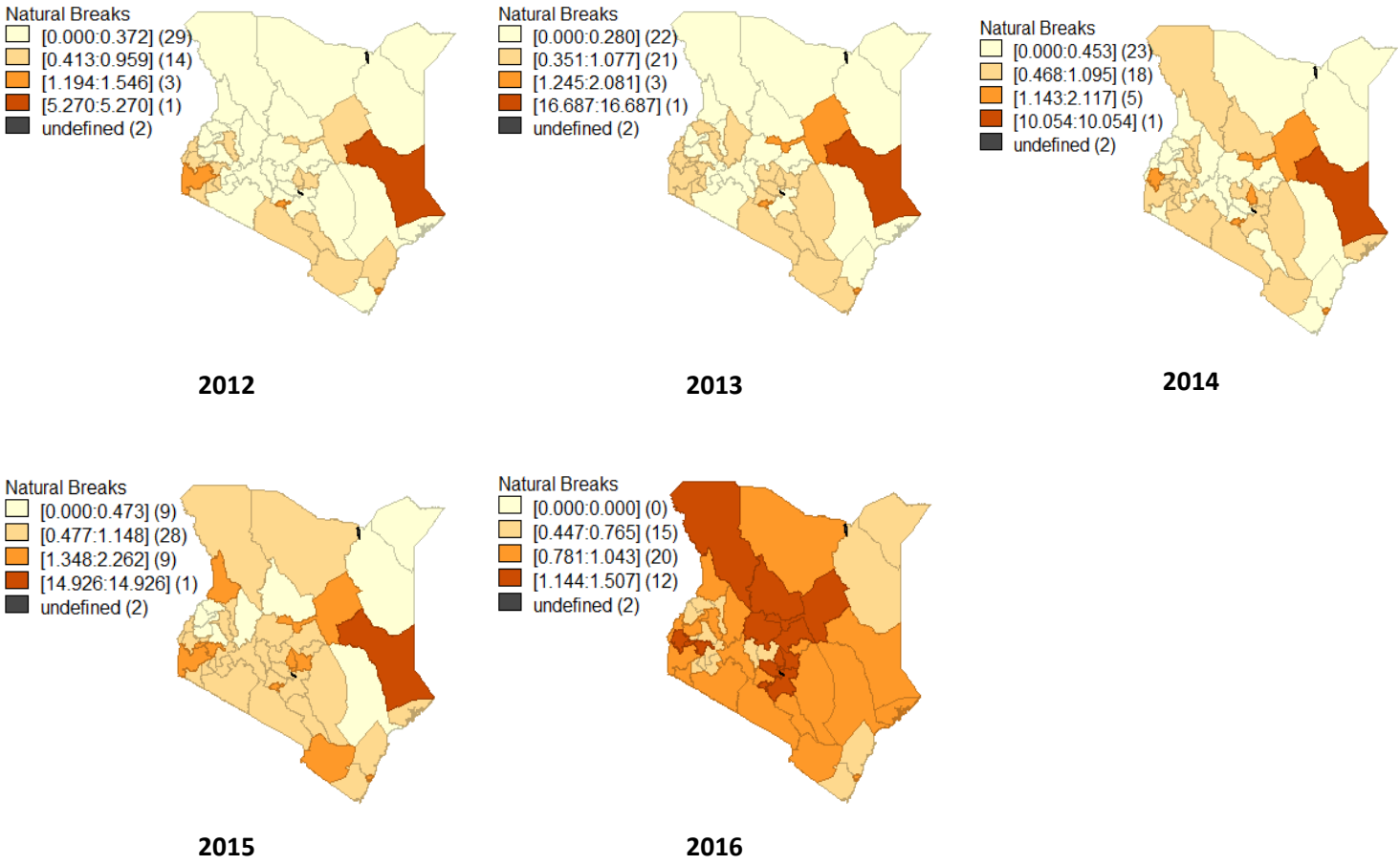


Figure 9: Empirical Bayes smoothed notification rate map 2012-2016

The number of counties with DR-TB risk greater than the national average (excess risk >1) has significantly increased from sixteen (34%) counties in 2012 to twenty two (46%) counties in 2016, Table 3. Garissa County maintained a four-fold higher than the national average until 2016 when the excess risk fell below it. Nairobi, Mombasa, Homa bay, Kirinyaga, Isiolo and Siaya are other counties with persistently above national average risk of DR-TB over the five years period. Figure 11, presents the county excess risk map.

Table 3: summary of County excess risk 2012 – 2016

Year	County excess risk group			
	ER 0 - 1	ER 1 -2	ER 2 -4	ER > 4
2012	31	11	4	1
2013	40	5	1	1
2014	33	10	3	1
2015	34	10	2	1
2016	25	20	2	0

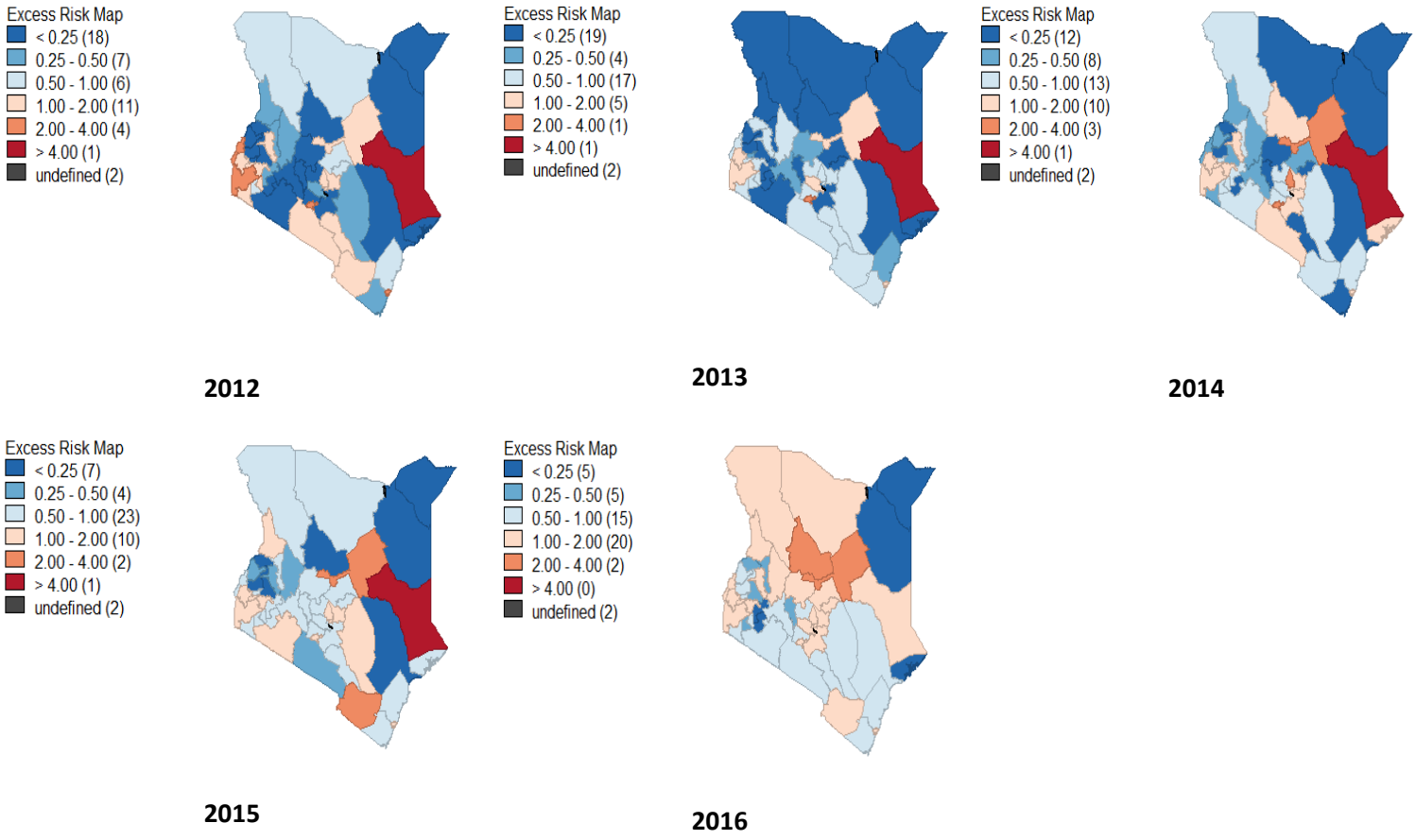


Figure 10: Excess risk map

Uneven distribution of DR-TB. A significant change in the distribution of EBS notification rate and excess risk of DR-TB has been observed between 2012 and 2016. However, the only 2016 shows a significant evidence of global spatial autocorrelation, with Moran's I statistics of 0.3099 and p-value of 0.001.The statistics summary of global spatial autocorrelation is presented in table 4.

Table 4: Global Moran's I Statistics

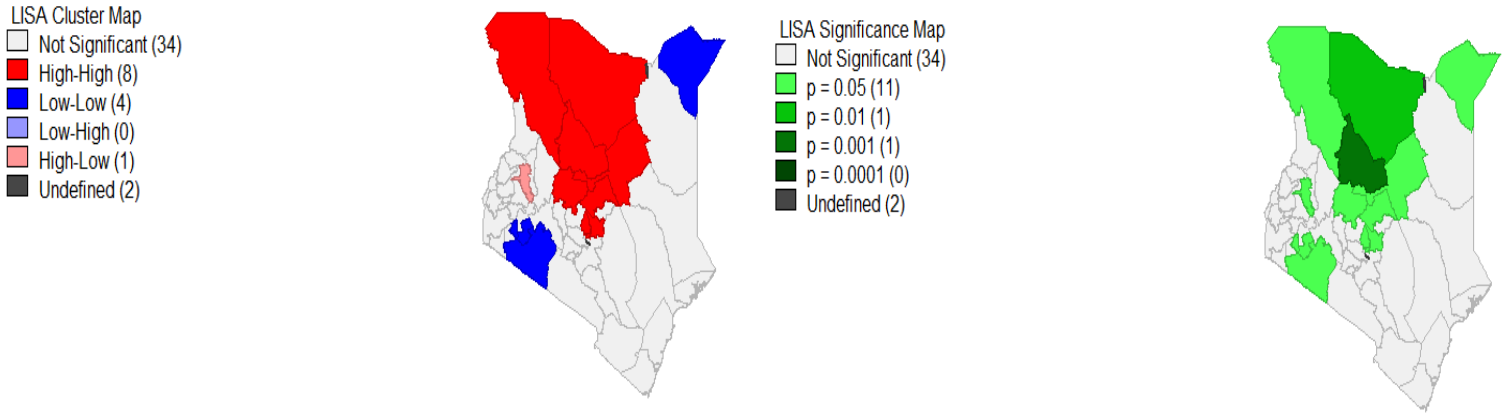
Year	Moran's I	Z – value	P - value
2012	-0.0656	-0.8434	0.188000
2013	-0.0377	-0.6874	0.229000
2014	0.0127	0.8254	0.172000
2015	-0.0272	-0.2424	0.420000
2016	0.3099	3.3754	0.001000

Being the only year with significant global spatial autocorrelation, 2016 was therefore, used to determine the local spatial autocorrelation. Thirteen of the 47 counties shows a significant evidence of local spatial autocorrelation. The LISA statistics summary, EBS notification rate and excess risk of the 13 counties is presented in table 5. Bomet, Mandera, Narok, and Nyamira were identified as counties with *below average value* of EBS notification rate surrounded by counties *with below average value* of EBS notification rate **(clusters).** On the other hand, Embu, Isiolo, Kirinyaga, Laikipia, Meru, Turkana, Marsabit and Samburu were identified as counties with *above average value* of EBS notification rate surrounded by counties with *above average value* of EBS notification rate **(hotspot).** Uasin Gishu is the only county with *above average value* of EBS notification rate surrounded by counties with *below average value* of EBS notification rate **(outlier)**.

Table 5. Summary of LISA statistics, EBSR and ERR for 13 counties with significant local Moran's I

Serial No.	Province	County	Excess Risk Ratio	EB Smoothed Rate	Category	LISA Statistics	P - value
1	Rift Valley	Bomet	0.234	0.516	L-L	1.23383	0.0108
2	Eastern	Embu	1.911	1.328	H-H	0.98169	0.0342
3	Eastern	Isiolo	3.441	1.374	H-H	1.283735	0.0486
4	Central	Kirinyaga	1.584	1.198	H-H	0.825823	0.0361
5	Rift Valley	Laikipia	1.693	1.218	H-H	0.833426	0.0187
6	North Eastern	Mandera	0	0.447	L-L	2.445716	0.0213
7	Eastern	Meru	1.600	1.324	H-H	0.680319	0.0305
8	Rift Valley	Norok	0.794	0.815	L-L	0.330107	0.0466
9	Nyanza	Nyamira	0.612	0.747	L-L	0.629272	0.0103
10	Rift Valley	Turkana	1.876	1.446	H-H	1.422116	0.0320
11	Rift Valley	Uasin Gishu	1.039	0.958	H-L	-0.048920	0.0239
12	Eastern	Marsabit	1.355	1.043	H-H	0.586449	0.0054
13	Rift Valley	Samburu	3.016	1.507	H-H	2.9911516	0.001

The *LISA cluster map* shows a clear clustering of counties with like value of EBS notification rate. The eight-(8) *high-high* counties are neighbouring one another and lies between the Eastern province (Embu, Isiolo, Meru and Marsabit), Rift valley province (Laikipia, Turkana and Samburu) and central province (Kirinyaga). Figure 11 and 12 presents LISA cluster map and LISA significant map.



LISA Cluster map 2016

LISA Significance map 2016

Figure 11: LISA cluster and significance map for 2016

Bayesian model for DR-TB cases: The expected cases of DR-TB was computed for the year 2016 (been the only year with significant spatial autocorrelation). Two models for count event namely "Poisson" and "negative binomial" were fitted. The deviance information criterion (DIC) was used in choosing the most fitting model. The Poisson model has a DIC = 262.09 while the Negative binomial model has DIC = 358.19. The Poisson model having the least DIC was used to model the count of DR-TB cases (i.e. DIC Poisson model < DIC Negative binomial model)

Three (3) covariate; the proportion of DR-TB patient with HIV co-infection, proportion of smear positive DR-TB cases, Proportion of malnourished cases were used for model. Table 6

presents the posterior mean, standard deviation (SD) and the 95% credible interval (i.e. the fixed effect) of the fitted model.

Deviance Information Criterion (DIC) = 262.09
Effective number of parameters = 43.59

Table 6: Model Fixed effects statistics summary

	Mean	SD	0.025 Quantile	0.5 Quantile	0.975 Quantile	Mode
Intercept	-3.069	1.002	-5.227	-3.004	-1.282	-2.878
Proportion of HIV	1.946	1.282	-0.548	1.933	4.511	1.909
Proportion of smear	2.197	1.287	-0.268	2.172	4.806	2.125
Proportion of malnourished	0.980	1.527	-1.962	0.956	4.061	0.912

Summary statistics for the model: Posterior mean, posterior standard deviation, posterior 95% credible interval and posterior mode of the model fixed effect

Table 7, presents the summary statistics of the model random effect (spatially structured and spatially unstructured effect). According to the summary, the variability in the expected number of cases of DR-TB between counties unexplained by the covariates (the fixed effect of the model) is largely explained by the heterogeneity of space (the spatially unstructured random effect with a precision mean 0.469 with a 95% credible interval (0.277-0.734), while the little remaining variability can be explained by the spatially structured random effect with a precision mean of 1867 and a credible interval (1265-6729).

Table 7: Model random effects statistics summary

Model hyper parameter	Mean	SD	0.025 Quantile	0.5 Quantile	0.975 Quantile	Mode
Precision for structured random effect	1867	1840	1265	1323	6729	3456
Precision for unstructured random effect	0.469	0.117	0.277	0.458	0.734	0.435

Summary statistics for the model: Posterior mean, posterior standard deviation, posterior 95% credible interval and posterior mode of the random effect.

The exponentiated value of the posterior mean, SD, mode and 95% credible interval of the model fixed effect is presented in table 8. According to the statistics summary, the proportion

of smear positive DR-TB cases appeared to have the highest effect on increasing the expected number of DR-TB cases for a given county than the other 2 covariates. For every one percent (1%) increase in the smear positive DR-TB cases for a given county, the expected number of DR-TB cases will increase by 8.994 with a 95% credible interval (0.765 - 122.266).

The proportion of HIV positive people is second most important predictor of spatial distribution of DR-TB cases. For every one percent (1%) increase in the HIV positive people for a given county, the expected number of DR-TB cases will increase by 6.998 with a 95% credible interval (0.578 - 91.047).

Lastly, for every one percent increase in the proportion of malnourished people for a given county, the expected number of DR-TB cases will increase by 2.665 with a 95% credible interval (0.141 - 58.040).

Table 8: Exponentiated model fixed effects statistics summary

	Mean	SD	0.025 Quantile	0.5 Quantile	0.975 Quantile	Mode
Intercept	0.046	2.724	0.005	0.049	0.278	0.0563
Proportion of HIV	6.998	3.605	0.578	6.909	91.047	6.747
Proportion of smear (+)	8.994	3.621	0.765	8.775	122.266	8.371
Proportion of malnourished	2.665	4.603	0.141	2.602	58.040	2.489

Finally, the relative risk (RR) of DR-TB for the counties was estimated using the expected number of DR-TB obtained by the fitted Poisson model. Table 8, presents the RR of the 47 Counties, while figure 12 shows the RR map.

Table 9: Expected relative risk for counties

S.No	County	Province	RR	S.No	County	Province	RR
1	**Baringo**	Rift Valley	0.651	25	**Marsabit**	Eastern	0.548
2	**Bomet**	Rift Valley	0.225	26	**Meru**	Eastern	4.485
3	**Bungoma**	Western	1.977	27	**Migori**	Nyanza	13.615
4	**Busia**	Western	0.654	28	**Mombasa**	Coast	3.673
5	**Elgeyo Marakwet**	Rift Valley	0.297	29	**Murang'a**	Central	10.489
6	**Embu**	Eastern	0.915	30	**Nairobi**	Nairobi	6.126
7	**Garissa**	North Eastern	0.562	31	**Nakuru**	Rift Valley	2.128
8	**Homa Bay**	Nyanza	4.594	32	**Nandi**	Rift Valley	0.296
9	**Isiolo**	Eastern	0.846	33	**Narok**	Rift Valley	7.098
10	**Kajiado**	Rift Valley	0.574	34	**Nyamira**	Nyanza	0.368
11	**Kakamega**	Western	2.634	35	**Nyandarua**	Central	0.231
12	**Kericho**	Rift Valley	0.218	36	**Nyeri**	Central	0.403
13	**Kiambu**	Central	2.542	37	**Samburu**	Rift Valley	1.111
14	**Kilifi**	Coast	2.017	38	**Siaya**	Nyanza	0.880
15	**Kirinyaga**	Central	0.793	39	**Taita Tateva**	Coast	0.493
16	**Kisii**	Nyanza	1.071	40	**Tana River**	Coast	0.383
17	**Kisumu**	Nyanza	6.137	41	**Tharaka Nithi**	Eastern	0.504
18	**Kitui**	Eastern	4.229	42	**Trans Nzoia**	Rift Valley	8.451
19	**Kwale**	Coast	0.344	43	**Turkana**	Rift Valley	12.856
20	**Laikipia**	Rift Valley	0.740	44	**Uasin Gishu**	Rift Valley	3.448
21	**Lamu**	Coast	0.122	45	**Vihiga**	Western	0.368
22	**Machakos**	Eastern	3.885	46	**Wajir**	North Eastern	0.051
23	**Makueni**	Eastern	0.4699	47	**West Pokot**	Rift Valley	0.613
24	**Mandera**	North Eastern	0.043				

CHAPTER FIVE: DISCUSSION

In this paper, ESDA methods were used to assess the spatial patterns of EB smoothed notification rates of DR-TB in the 47 counties of Kenya from 2012 to 2016. A Bayesian model was used to evaluate the effect of 3 covariate on the spatial distribution of DR-TB. We found a remarkable temporal pattern in the spread of DR-TB between 2012 and 2016, with more counties having increasing values of EB smoothed notification rate and excess risk.

The year 2016 shows a significant evidence of spatial autocorrelation in the distribution of EB smoothed notification rates. Also a significant clustering of counties with higher values of EB smoothed notification rates surrounded by counties with higher values of EB smoothed notification rate has been identified (hot spot). Since TB is an aerosol transmissible disease, the spatial clusters are likely due diffusion of cases from neighbouring counties. Further researches are needed to explore the local driver of the spread, especially, among other high risk group such population of immigrant and IDPs. Other factors such as socio-economics also need be explored.

Recognising the tendency of every 1% increase in the smear positive DR-TB cases to rise county expected number of DR-TB by 8.994 with a 95% credible interval (0.765-122.266), measures like intensive health education and awareness campaign on cough etiquette and good housing ventilation can help reduce spread of the disease. Intensified case finding, especially, among high risk groups and improve nutrition can have significant impact in reducing the burden of the disease.

To the best of our knowledge this study is the first that applied EB smoothing technique, ESDA methods and Bayesian model to analyse spatial distribution of DR-TB in Kenya. We believe that conclusions drive from this methods provide a reliable results, especially recognizing the fact that the 3 methods complement one another.

CHAPTER SIX: CONCLUSION AND RECOMMENDATION

6.1 Conclusion

This study provide a good insight on the spatial patterns of DR-TB notification rates in Kenya. The results show the effectiveness of ESDA and Bayesian modelling in mapping DR-TB risk, and might help in policy decision of allocating resources to high-risk areas. It also revealed the uneven distribution of DR-TB and this can help the national TB, and leprosy control program to focus more on locations with the greatest risk. Special intention is needed especially, in the regions with significant clustering of high-high counties (hotspots).

6.2 Study recommendation

More researches are needed to fully understand the local drivers responsible for the uneven distribution of DR-TB in Kenya. Contribution of Socio-economic factors, immigration and IDP population in the spread of DR-TB need to be evaluated.

6.3 Study limitations

Like any other, this study was faced with some limitations. Firstly, there is limited control over the quality of secondary data. Secondly, the number of reported cases of drug-resistant TB obtained from the Kenyan national TB surveillance database may not be the actual representative of all the cases of drug-resistant TB across Kenya. This is due to the fact that tuberculosis is often underreported or misdiagnosed. Lastly, the data is assumed to meet all the required model assumptions.

The study intends to model both the spatial and temporal risk of DR-TB, however, having only the year 2016 with a significant spatial autocorrelation has limit the analysis to spatial only. Lastly, the study was not able to incorporate socio-economic and environmental covariate into the model due to unavailability of such data.

REFERENCE

1. Álvarez-Hernández, G.; Lara-Valencia, F.; Reyes-Castro, P. A.; Rascón-Pacheco, R.A., 2010. An analysis of spatial and socio-economic determinants of tuberculosis in Hermosillo, Mexico, 2000–2006. *The International Journal of Tuberculosis and Lung Disease*, Volume 14, p.708–713(6). Available at: http://www.ingentaconnect.com/content/iuatld/ijtld/2010/00000014/00000006/art00008.
2. Anselin, L., 1995. Local Indicators of Spatial Association-LISA. , 27(2).
3. Anselin, L., Kim, Y.W. & Syabri, I., 2003. Web-Based Analytical Tools for the Exploration of Spatial Data ∗. , pp.1–24.
4. Associates, C., 2002. Exploratory and Confirmatory Spatial Data Analysis Approaches to Studying the Correlates of Juvenile Violent Crimes, Volume II Final Report.
5. Auchincloss, A.H. et al., 2012. A Review of Spatial Methods in Epidemiology, 2000–2010. *Annual review of public health*, 33, pp.107–122. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3638991/.
6. Beyers., M.L.B.Z.E., Tuberculosis transmission patterns in a high-incidence area: a spatial analysis. *The International Journal of Tuberculosis and Lung Disease*, 7(Number 3), pp.271–277. Available at: http://www.ingentaconnect.com/contentone/iuatld/ijtld/2003/00000007/00000003/art00015.
7. Bivand;, R.S. & Gómez-Rubio, V., 2008. *Applied Spatial Data Analysis with R* R. K. H. G. P. Gentleman, ed., Springer.
8. Carroll, L.N. et al., 2014. Visualization and analytics tools for infectious disease epidemiology: A systematic review. *Journal of Biomedical Informatics*, 51, pp.287–298. Available at: http://www.sciencedirect.com/science/article/pii/S1532046414000914.
9. Casal M, Vaquero M, Rinder H, Tortoli E, Grosset J, Rüsch-Gerdes S, Gutiérrez J, J.V.., 2005. A Case-Control Study for Multidrug-Resistant Tuberculosis: Risk Factors in Four European Countries. , 11, pp.62–67.
10. Chang, K.C. & Yew, W.W., 2013. Management of difficult multidrug-resistant tuberculosis and extensively drug-resistant tuberculosis: Update 2012. *Respirology*, 18(1), pp.8–21.
11. Clarke, K.C., McLafferty, S.L. & Tempalski, B.J., 1996. On epidemiology and geographic information systems: a review and discussion of future directions. *Emerging Infectious Diseases*, 2(2), pp.85–92. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2639830/.
12. Couceiro, L.; Santana, P.; Nunes, C., 2011. Pulmonary tuberculosis and risk factors in Portugal: a spatial analysis. *The International Journal of Tuberculosis and Lung Disease*, Volume 15, p.1445–1455(11). Available at: http://www.ingentaconnect.com/content/iuatld/ijtld/2011/00000015/00000011/art00008.
13. Cressie, N., 1992. Smoothing Regional Maps Using Empirical Bayes Predictors. , 24(1).
14. Faustini, A., Hall, A.J. & Perucci, C.A., 2006. Risk factors for multidrug resistant tuberculosis in Europe: a systematic review. *Thorax*, 61. Available at: https://doi.org/10.1136/thx.2005.045963.
15. Fulcher, C. & Kaukinen, C., 2005. Mapping and visualizing the location HIV service providers: An exploratory spatial analysis of Toronto neighborhoods. *AIDS Care*,

17(3), pp.386–396. Available at: http://dx.doi.org/10.1080/09540120512331314312.
16. Gotway, L.A.W.C.A., 2004. *Applied Spatial Statistics for Public Health Data*, John Wiley & Sons, Inc.
17. Grandjean, L. et al., 2015. Transmission of Multidrug-Resistant and Drug-Susceptible Tuberculosis within Households: A Prospective Cohort Study. *PLoS Medicine*. Available at: http://journals.plos.org/plosmedicine/article/file?id=10.1371/journal.pmed.1001843&type=printable [Accessed March 14, 2017].
18. Harling, G. & Castro, M.C., 2014. A spatial analysis of social and economic determinants of tuberculosis in Brazil. *Health & Place*, 25, pp.56–67. Available at: http://www.sciencedirect.com/science/article/pii/S135382921300141X.
19. Hendricks, Brian; Mark-Carew, Miguella P; Parker, D., 2016. An Exploratory Spatial Data Analysis Approach to Lyme Surveillance and Characterization of Border Bias in an Eastern United States Region. Available at: https://cste.confex.com/cste/2016/webprogram/Paper6559.html.
20. Jenkins, H.E. et al., 2013. Assessing spatial heterogeneity of multidrug-resistant tuberculosis in a high-burden country. *European Respiratory Journal*, 42(5), p.1291 LP-1301. Available at: http://erj.ersjournals.com/content/42/5/1291.abstract.
21. Kipruto, H. et al., 2015. Spatial Temporal Modelling of Tuberculosis in Kenya Using Small Area Estimation Spatial Temporal Modelling of Tuberculosis in Kenya Using Small Area Estimation. *International Journal of Science and Research*, (November 2016). Available at: https://www.researchgate.net/publication/282914116%0ASpatial.
22. Latouche, A. et al., 2007. Robustness of the BYM model in absence of spatial variation in the residuals. *International Journal of Health Geographics*, 6(1), p.39. Available at: https://doi.org/10.1186/1476-072X-6-39.
23. LIN, H. et al., 2011. Assessing spatiotemporal patterns of multidrug-resistant and drug-sensitive tuberculosis in a South American setting. *Epidemiology and Infection*, 139(11), pp.1784–1793. Available at: https://www.cambridge.org/core/article/assessing-spatiotemporal-patterns-of-multidrugresistant-and-drugsensitive-tuberculosis-in-a-south-american-setting/46E47465E27C230F5CD1DA69B00217AA.
24. Liu, Y. et al., 2011. Spatial epidemiology and spatial ecology study of worldwide drug-resistant tuberculosis. *International Journal of Health Geographics*, 10(1), p.50. Available at: https://doi.org/10.1186/1476-072X-10-50.
25. Lönnroth, K. et al., 2009. Drivers of tuberculosis epidemics: The role of risk factors and social determinants. *Social Science & Medicine*, 68(12), pp.2240–2246. Available at: http://www.sciencedirect.com/science/article/pii/S0277953609002111.
26. Ma, B., Lawson, A.B. & Liu, Y., 2007. Evaluation of Bayesian models for focused clustering in health data BT - Environmetrics.
27. Messner, S.F. et al., 1999. The Spatial Patterning of County Homicide Rates: An Application of Exploratory Spatial Data Analysis. *Journal of Quantitative Criminology*, 15(4), pp.423–450. Available at: https://doi.org/10.1023/A:1007544208712.
28. NTLD-Program, 2014. *The Kenya National Strategic Plan on Tuberculosis, Leprosy and Lung Diseases 2015 – 2018*,
29. Nunes, C., 2007. Tuberculosis incidence in Portugal: spatiotemporal clustering. *International Journal of Health Geographics*, 6(1), p.30. Available at: https://doi.org/10.1186/1476-072X-6-30.
30. Pascutto C; Wakefield JC, B.N.R.S.B.L.S.A.E.P., 2000. Statistical issues in the analysis of disease mapping data. *NCBI*, 19, pp.17–18. Available at:

https://www.ncbi.nlm.nih.gov/pubmed/10960868.
31. PMDT-Kenya, 2014. Kenya Programmatic Management of Drug resistant TB Guideline. *WHO*. Available at: http://apps.who.int/medicinedocs/documents/s21976en/s21976en.pdf.
32. Roberts, G., 1996. Markov chain concepts related to sampling algorithms BT - Markov chain Monte Carlo in Practice. In W. Gilks, S. Richardson, & D. Spiegelhalter, eds. -: Chapman and Hall.
33. Shrestha, S.S. et al., 2012. Spatial Clusters of County-Level Diagnosed Diabetes and Associated Risk Factors in the United States. *The Open Diabetes Journal*, 7, pp.29–37. Available at: https://benthamopen.com/ABSTRACT/TODIAJ-5-29.
34. Sitienei, J.K., 2014. *The spatial temporal distribution of smear positive tuberculosis in Nairobi , Nyanza & Rift valley provinces of Kenya*. Jomo Kenyatta University of Agriculture and Technology.
35. Sitienei J., 2016. Kenya's adoption of an electronic tuberculosis surveillance system. *AMR Control*. Available at: http://resistancecontrol.info/2016/amr-and-tuberculosis/kenyas-adoption-of-an-electronic-tuberculosis-surveillance-system/ [Accessed August 30, 2017].
36. Smith, T.E., Spatial weight matrices.
37. Tian, Y.Q. et al., 2002. An exploratory spatial analysis to identify the appropriate scale and potential risk factors for understanding obesity distributions.
38. Tiwari, N. et al., 2006. Investigation of geo-spatial hotspots for the occurrence of tuberculosis in Almora district, India, using GIS and spatial scan statistics. *International Journal of Health Geographics*, 5. Available at: https://doi.org/10.1186/1476-072X-5-33.
39. Tu, W., Tedders, S. & Tian, J., 2012. An exploratory spatial data analysis of low birth weight prevalence in Georgia. *Applied Geography*, 32(2), pp.195–207. Available at: http://www.sciencedirect.com/science/article/pii/S0143622811000993.
40. Uthman O. A., 2008. Spatial and temporal variations in incidence of tuberculosis in Africa, 1991 to 2005. *PubMed*. Available at: https://www.ncbi.nlm.nih.gov/pubmed/18978457.
41. WHO, 2014. *Companion handbook to the WHO guidelines for the programmatic management of drug-resistant tuberculosis*, Geneva: WHO. Available at: http://apps.who.int/iris/bitstream/10665/130918/1/9789241548809_eng.pdf?ua=1&ua=1 [Accessed March 18, 2017].
42. WHO, 2016a. *Global Tuberculosis Report* 21st ed., Geneva: WHO. Available at: http://apps.who.int/iris/bitstream/10665/250441/1/9789241565394-eng.pdf?ua=1 [Accessed March 15, 2017].
43. WHO, 2015. *Guidelines for surveillance of drug resistance in tuberculosis 5 th Edition* 5th ed., Geneva: WHO. Available at: http://apps.who.int/iris/bitstream/10665/174897/1/9789241549134_eng.pdf?ua=1 [Accessed March 15, 2017].
44. WHO, 2010. *Multidrug and extensively drug-resistant TB (M/XDR-TB) 2010 Global report on surveillance and response*, Geneva: WHO. Available at: http://apps.who.int/iris/bitstream/10665/44286/1/9789241599191_eng.pdf?ua=1&ua=1 [Accessed March 18, 2017].
45. WHO, 2016b. *WHO treatment guidelines for drug- resistant tuberculosis October 2016 Revision*, Geneva: WHO. Available at: http://apps.who.int/iris/bitstream/10665/250125/1/9789241549639-eng.pdf?ua=1 [Accessed March 18, 2017].
46. Wise, S., Haining, R. & Ma, J., 1997. Regionalisation Tools for the Exploratory

Spatial Analysis of Health Data BT - Recent Developments in Spatial Analysis: Spatial Statistics, Behavioural Modelling, and Computational Intelligence. In M. M. Fischer & A. Getis, eds. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 83–100. Available at: https://doi.org/10.1007/978-3-662-03499-6_5.

47. Zenk, S.N. et al., 2005. Neighborhood Racial Composition, Neighborhood Poverty, and the Spatial Accessibility of Supermarkets in Metropolitan Detroit. *American Journal of Public Health*, 95(4), pp.660–667. Available at: http://dx.doi.org/10.2105/AJPH.2004.042150.
48. Zhang, J. et al., 2014. Spatial analysis on human brucellosis incidence in mainland China : 2004 – 2010. , pp.2004–2010.
49. Zumla, A. et al., 2012. Drug-Resistant tuberculosis-current dilemmas, unanswered questions, challenges, and priority needs. *Journal of Infectious Diseases*, 205(SUPPL. 2).