SPATIAL EPIDEMIOLOGY OF TUBERCULOSIS IN SIAYA AND KISUMU COUNTIES, WESTERN KENYA, 2013

BY

SIFUNA M. PETER

A THESIS SUBMITTED IN PARTIAL FULFILMENT FOR THE REQUIREMENTS FOR THE DEGREE OF MASTER OF PUBLIC HEALTH

SCHOOL OF PUBLIC HEALTH AND COMMUNITY DEVELOPMENT

MASENO UNIVERSITY

DECLARATION

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

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

Peter Mwangala Sifuna

(PG/MPH/071/2012)

This thesis has been submitted for examination with our approval as supervisors:

Prof. Collins Ouma, (PhD)

School of Public Health and Community Development,

Department of Biomedical Sciences and Technology,

Maseno University.

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

Dr. Harrysone Atieli, (PhD)

School of Public Health and Community Development,

Maseno University.

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

ACKNOWLEDGEMENTS

I acknowledge the blessings of The Almighty God in my life. Throughout this period, He granted me courage, hope and health amidst obstacles and disappointments.

I thank my academic advisors, Professor Collins Ouma and Dr. Harrysone Atieli for accepting to guide me through my research, for their guidance, critical comments, valuable input and correction of the thesis.

I value the contribution of Dr. Kevin Cain (CDC). His comments and insightful thoughts throughout my research were thought-provoking and they helped me focus my ideas. Special thanks go to Dr. Jessica Cowden (KEMRI/WRP) for allowing me the time and space to pursue my master's degree within the work environment. I am sincerely grateful to her.

I would like to acknowledge my family who supported me during the course of my study. I thank my Dad, Mr. Vincent Watenya and my Mom, Rosemary Kwatindi for their constant encouragement. I thank my siblings Irene Ngila, Edwin Sifuna, Emily Nekesa, and Reuben Buronya for their constant love and support.

DEDICATION

To the 1.5 million people who lost their lives to TB in 2013.

ABSTRACT

Kenya is among the 22 high burden countries that contribute 80% of the world's TB and is ranked 15th globally, while in Africa, it is ranked number five. Strategies for reduction of TB rely on knowledge of where, when and to what degree the disease is present. Information on the geographical occurrence of TB in the high burden regions of Kisumu and Siaya Counties, Western Kenya remains scanty. As such, the main objective of this study was to analyze the spatial distribution of TB and identify geographic factors associated with its occurrence in Kisumu and Siaya Counties, Western Kenya. The study area has a high TB prevalence with a mixture of urban and rural areas, with Kisumu County being the most urbanized. The specific objectives were to determine the presence of TB clusters and hotspots in Kisumu and Siaya Counties, Western Kenya and to correlate the occurrence of TB and residency (urban or rural), population density and slum dwelling in Kisumu and Siaya Counties. In a cross-sectional study, data on the notified TB cases for the year 2013 were extracted from the Kenyan National TB Control Program database. Since November of 2012, Kenya has been implementing an electronic TB register system and complete baseline data was available for the year 2013. The TB cases were stripped off personal identifiers and geocoded at the County Ward level based on the physical address provided in the register data. The ArcGIS® software was used to visualize and generate choropleth maps which helped to identify geographical patterns of TB. A spatial autocorrelation indicator, Global Moran's I index was used describe the pattern of TB. The Getis-Ord Gi* static was used to identify TB hotspots and cold spots while the Local Moran's I statistic was used to identify statistically significant TB cluster and TB outliers. A simple histogram was used to test for data normality. TB rates per 100,000 populations were computed for each county Ward grouped by level of urbanization and population density. Computed means were used to compare TB occurrence in the urban and rural areas. Spearman's test was used to correlate TB rate and population density. A total of 5,568 TB cases were abstracted from 237 TB clinics in Kisumu and Siaya Counties. Of these, 5,063 (91.0%) were linked to the 65 Wards within the study area. The notified TB rate of 278 cases per 100,000 populations with variations at the Ward level (76 per 100, 000 population) in South Uyoma Ward to (813 per 100, 000 population) in Nyalenda A Ward. Moran's indices were positive (Moran's I=0.423, p<0.001) indicating a clustered characteristic of the distribution. The study revealed distinct TB hotspots regions (with z-score> 2.58) and 7 County Wards high-high relationship with its neighbors (clusters). There was a positive correlation between population density and the rate of TB, which was statistically significant (rs=0.5739, p=0.0001). The study also found higher TB mean (3.9 and 2.2) in the urban Wards and rural Wards, respectively (Wilcoxon rank-sum, p=0.0001). In conclusion, it is clear that TB occurrence in the high burden Counties of Siaya and Kisumu varies geographically at the small area level. The study revealed that the distribution of TB in Siaya and Kisumu Counties is nonrandom and clustered with the significant TB clusters and hotspots identified for the year 2013. Urbanization and high population density areas were found to be positively-correlated with TB occurrence. The current study has added knowledge that is being utilized currently in the allocation of TB resources in the identified high TB regions in Siaya and Kisumu Counties. The identification of TB clusters and hotspots can now be used by TB programs in targeting prevention strategies in the identified areas. The correlation of TB with urban population and population density will be utilized in designing of specific interventions that would target these areas. In general, these findings will help inform TB prevention and control strategies and in allocation of resources towards these efforts in Western Kenya.

TABLE OF CONTENTS

OPERATIONAL DEFINITIONS

ArcGIS- Proprietary GIS software by Environmental Systems Research Institute (ESRI)

Disease Cluster-An excess of cases above some background rate bounded in time and space

Cluster analysis/Clustering - Is the task of grouping a set of objects in such a way that objects in the same group (called a cluster) are more similar (in some sense or another) to each other than to those in other groups (clusters).

Choropleth map - Is a thematic map in which areas are shaded or patterned in proportion to the measurement of the statistical variable being displayed on the map.

Cold spot –Statistically significant spatial clusters of low values

Extra pulmonary tuberculosis – This is the type of tuberculosis that affects other organs of the body other than the lungs. These include; bones, kidney, female reproductive organs, abdominal cavity, Joints etc

GIS - Set of tools (hardware and software) that allow the acquisition, manipulation, analysis and display of spatial data (geographically referenced data).

Hot spot – Statistically significant spatial clusters of high values

Prevalence - The ratio (for a given time period) of the number of occurrences of a disease or event to the number of units at risk in the population.

Risk factors – It is something that increases ones chances of getting a disease.

Spatial - Space and time (relating to, occupying, or having the character of space).

Spatial data - It is the data or information that identifies the geographic location of features and boundaries on earth, such as natural or constructed features.

Spatial autocorrelation – A measure of the degree to which a set of spatial features and their associated data values tend to be clustered together in space (positive spatial autocorrelation) or dispersed (negative spatial autocorrelation)

Spatial epidemiology- Description and analysis of geographic variations in disease with respect to demographic, host characteristics and geographical factors

Spatial statistics - Involves statistical methods utilizing location and distance in inference.

Shapefiles – Is a set of files that work together to create a map with all its features.

Thematic map - A thematic map is a type of map or chart especially designed to show a particular theme connected with a specific Geographic area

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

CHAPTER 1:INTRODUCTION

1.1 Background Information

Tuberculosis (TB) kills nearly 200 people per day globally (WHO, 2013), and the widespread emergence of drug resistance has made the situation worse. Kenya is among the 22 high burden countries that contribute 80% of the global burden and is ranked number 15 globally while in Africa, it is ranked number five(WHO, 2015). The occurrence of TB in Kenya is characterized by geographical variations across the country. Nyanza region (encompassing Kisumu and Siaya Counties) in Western Kenya has the highest TB burden in Kenya, with an estimated prevalence of 500-600 cases per 100,000, and an HIV prevalence of 70% in patients with TB (Van't Hoog *et al*., 2011).

The success of TB interventions depends on understanding of the geographical distribution of the disease as well as the underlying factors that influence its occurrence. Disease mapping using geographical information systems (GIS) is an activity widely used to identify geographical patterns of diseases and to develop new ideas about the causation factors of disease (Newsom, 2006). GIS refers to a computerized database management system for the analysis and display of geographically referenced data (Clarke *et al*., 1996).

Disease maps provide a visual summary of geographical disparities in occurrence and risk. Disease mapping can take the form of point patterns, where the precise location of actual cases of disease are mapped or aggregate data where case counts or risk, rates or prevalence of disease for different areas are mapped (Elliott and Wartenberg, 2004). Maps are used variously for descriptive purposes such as to generate hypotheses as to etiology, for surveillance to highlight areas at apparently high risk, and to aid policy formation and resource allocation.

Disease mapping is often traced back to the work of John Snow and his analysis of cholera in London in 1854(Snow, 1988). Mapping cases in relation to the location of a potential pollution source, in this case water pumps, showed that cases were clustered around a pump on Broad Street. Snow's work exemplified the use of mapping in providing insights into communicable-disease etiology and intervention, helping disease control efforts.

GIS has been used increasingly in public health leading to some notable etiologic insights in diseases such as in malaria (Shirayama *et al*., 2009), HIV/AIDs (Ali *et al*., 2002; Cuadros *et al*., 2013), cholera (Ali *et al*., 2002), cancer (Jacquez and Greiling, 2003; Kitron and Kazmierczak, 1997) and Lyme Disease (Kitron and Kazmierczak, 1997). The geographical distribution of TB within the high burden Counties of Siaya and Kisumu was unclear Kenya, making it necessary to carry out this study.

One key public health information derived from disease mapping is knowledge on whether the occurrence of disease is random or clustered or regular. A disease cluster is defined as a limited area within the general study area with a significant increase in the incidence of a disease (Wakefield, 2001). Identification of TB clusters and TB hotspots can be a useful tool designing TB control and prevention programs and for generating etiologic hypothesis.

In recent years, GIS techniques have been used to highlight TB clusters and hotspots in India (Tiwari *et al*., 2006), China (Wang Tao *et al*., 2012) as well as in South Africa(Munch *et al*., 2003) and Gambia, West Africa (Touray *et al*., 2010). Whether or not TB occurrence in the high burden Counties of Siaya and Kisumu was random, regular or clustered was unclear. As such, this study sought to establish the pattern of TB and identify TB clusters and hotspots in the study area.

Within the high burden regions, underlying differences in demographic, host characteristics or socioeconomic factors may influence the distribution of disease at the small area level. A variety of environment related factors (urban residence, population density, concurrent HIV epidemic and proximity and persistence of contacts) have been shown to play a role in the distribution of TB (Lopez De Fede *et al*., 2008; Moonan *et al*., 2004; Muniyandi and Ramachandran, 2008). The two Counties selected both fall within the larger Nyanza region, which has a high TB prevalence. Kisumu County was the administrative capital of Nyanza Province under the old constitution and as such more urbanized as compared to the rest of the Counties. In terms of population density (number of people per square kilometer), Siaya County has the highest population density of the remaining rural Counties in the former Nyanza Province. The inclusion of the two Counties allowed for the comparison of urban and rural residence as well as population density.

The interest in geographic correlation lies in determining whether a risk factor is associated with the geographic variation of a disease. Although geographic correlation cannot reveal any causal relationships, it often provides important clues for associations that may be worth more rigorous studies, for example in a case-control setup (Rezaeian *et al*., 2007). Whether or not TB occurrence in Siaya and Kisumu Counties was correlated with environmental factors was unclear. As such, this study sought to evaluate the correlation of TB and a number of environmental factors in Kisumu and Siaya Counties using GIS techniques.

1.2 Problem Statement

TB prevention programs in Kenya have mostly focused on detecting and treating index cases. Because TB is primarily spread through close contacts with infected individuals (mostly household contacts), programs need to identify specific regions with the highest TB burden so as to employ aggressive prevention measures (such as intensive case finding and contact tracing) in order to interrupt ongoing disease transmission. Available data suggest that to speed the decline in TB incidence will require two synergistic strategies: 1) Improve TB case detection (find more cases of TB, find them faster, and treat them effectively; and 2) Address the large burden of latent TB infection with strategies to prevent the development of active TB disease in these patients (Dye, 2008). Implementation of these strategies is hampered by inadequate supporting information on where the disease is present, more so within the high burden regions. Technologies, such as geographical information systems (GIS), may be useful in this process. Despite the demonstrated utility of GIS in TB prevention and control programs, there has been no application within the highly burdensome regions in Kenya at a small area level. Due to underlying differences in demographic, host characteristics or socioeconomic factors, the distribution of TB may vary at the small area level even within the high burdensome regions. In order to fill this crucial gap, this study examined the geographical distribution of TB, investigated TB patterns and analyzed the correlation of TB with the various geographic factors in Siaya and Kisumu Counties.

1.3 Objectives

1.3.1 General Objective

To analyze the spatial distribution of TB and identify geographic factors associated with TB occurrence in Kisumu and Siaya, Western Kenya.

1.3.2 Specific Objectives

i) To analyze the spatial distribution of TB in Kisumu and Siaya Counties, Western Kenya.

ii) To determine the presence of TB clusters and hotspots in Kisumu and Siaya Counties, Western Kenya.

iii) To correlate the occurrence of TB and residency (urban or rural), population density and type of settlement (slum dwelling or presence of informal settlement) in Kisumu and Siaya Counties.

1.3.3 Research Questions

i. What is the spatial variability of TB disease in Kisumu and Siaya Counties, Western Kenya?

ii. What are the TB clusters and hotspot areas within Kisumu and Siaya Counties, Western Kenya?

iii. What is the correlation of TB and residency (urban or rural), population density and type of settlement (slum dwelling or presence of informal settlement) within Kisumu and Siaya Counties?

1.4 Significance of the Study

Mapping the geographical occurrence of TB will enable TB programs to identify specific regions with the highest disease burden within the larger high burden region. This information is critical in assisting programs targeted prevention and control strategies (such as intensive case finding and contact tracing) with the goal of reducing disease incidence. Targeting of interventions ensure the efficient use of the otherwise limited resources and also allows programs to monitor the effectiveness of the interventions employed. Results from this study has also served as a geographic early warning disease surveillance system, especially when similar data is mapped in real time. The study also gives important clues as to etiology from the observed differences in TB occurrence, which form the basis for further studies at individual-level studies such as cohort–control or case–control. This study was conducted Kisumu and Siaya Counties, western Kenya, two of the high TB burden areas in Kenya. The geographical distribution of TB at a small area level within a high burden population was unclear. Further, the presence of TB clusters/hotspots as well as the correlation of TB with the various geographic factors within Kisumu and Siaya Counties was unclear. As such, this study analyzed the spatial distribution of TB in Kisumu and Siaya Counties, investigated the presence of TB clusters and hotspots and examined the correlation of TB and residency (urban, rural), population density and slum dwelling. Results from this study, is in the process of guiding TB programs in targeting interventions, resource allocation to these areas and monitoring the effectiveness of the employed interventions.

CHAPTER 2: LITERATURE REVIEW

2.1 The Burden of TB

Tuberculosis (TB) remains one of the world's deadliest communicable diseases (WHO, 2015). Globally, TB accounts for over 9 million cases and 1.5 million deaths each year. TB ranks as the second leading cause of death from an infectious disease worldwide, after HIV. The incidence of TB is estimated at 1280 per 1 million population per year globally.

Africa carries the most severe burden, with 280 cases per 100 000 population compared with a global average of 133 cases per 100 000 population (WHO, 2015). It is estimated that if TB is left unchecked, it will Kill a 35 million people in a span of 20 years (Partnership, 2015). The WHO estimates thatmore than 80% of all TB patients live in sub-Saharan Africa and Asia.

Kenya is among the 22 high burden countries that contribute 80% of the global burden and is ranked number 15 globally, while in Africa, it is ranked number five. It is estimated that approximately 120,000 incident TB cases occur in Kenya each year. This equates to a TB incidence rate of 2980 per 1 million population per year (WHO, 2010).

In Kenya, Nyanza region, located in Western part of the country on the shores of Lake Victoria, has the highest TB burden in Kenya. Province-wide, approximately 4500 TB cases are reported per 1 million population per year. The major factor responsible for the large TB disease burden is the concurrent HIV epidemic. Other factors that have contributed to this large TB disease burden include poverty and social deprivation that has led to a mushrooming of peri-urban slums, congestion in prisons and limited access to general health care services (NTLLD, 2015b).

2.2 The concept of spatial epidemiology

Spatial epidemiology concerns describing and understanding geographical variations in health, especially in small area level (Elliott and Wartenberg, 2004). Spatial epidemiology is often traced back to the work of John Snow and his analysis of cholera in London 1854 (Snow, 1988). Mapping cases in relation to the location of a potential pollution source, in this case water pumps, showed that cases were clustered around a pump on Broad Street. Snow's work exemplified the use of mapping in providing insights into communicable-disease etiology and intervention, helping disease control efforts. There are three categories in spatial epidemiologic inquiry;

i. Disease mapping.

ii. Cluster detection.

iii. Geographic correlation studies.

2.2.1 Disease Mapping

Disease maps provide a visual summary of geographical disparities in occurrence and risk. Disease mapping can take the form of point patterns, where the precise location of actual cases of disease are mapped or aggregate data where case counts or risk, rates or prevalence of disease for different areas are mapped (Elliott and Wartenberg, 2004). Over the last two decades, there has been an increased application of GIS mapping in the context of public health and research (Eisinger and Thulke, 2008). According to Eisinger, this is because transmission of infectious diseases is closely linked to the concepts of geographical proximity and hence transmission is more likely to occur if the at-risk individuals are close in a spatial and a temporal sense.

In china for example, Wang and colleagues used GIS techniques to map and identify TB patterns in Linyi City. The study revealed distinct geographic regions with elevated TB occurrence illustrating that TB distribution in Linyi City was nonrandom and clustered (Wang *et al*., 2012). Such information is useful for disease prevention and control as it provides an opportunity for targeted TB interventions in the identified high-risk areas. In Almora district of India, similar techniques were used to analyze and map TB occurrence revealing distinct patterns of TB occurrence (Tiwari *et al*., 2006).

Disease maps are used variously for descriptive purposes such as to generate hypotheses as to etiology, for surveillance to highlight areas at apparently high risk, and to aid policy formation and resource allocation. Couceiro and colleagues used GIS techniques to map out the risk of pulmonary TB in Portugal. The study found that some areas were at higher risk of PTB than others because of high incidence of HIV/AIDS and sub-standard accommodation (Couceiro *et al*., 2011).

In South Africa, GIS and spatial analysis were utilized to generate maps that revealed transmission patterns of the disease in a high-incidence area (Munch *et al*., 2003). In an urban West Africa, GIS techniques were employed to visualize and investigate TB patterns in Greater Banjul region in Gambia resulting in the identification of distinctive TB patterns (Touray *et al*., 2010). Comparable GIS mapping techniques have been applied in Madagascar (Randremanana *et al*., 2009) and in India (Tiwari *et al*., 2010) in TB analysis and visualization.

No study looking at small area mapping with the high burden regions in Kenya have been published in Kenya. A study by Musika used GIS to analyze the distribution of TB in one of the low burden Counties in Kenya using secondary data obtained from TB registers. Visualization and analysis in the study was achieved by converting the physical addresses of the facilities and the participants into GIS coordinates. The study found higher TB rates in major towns within the County, along the major roads and densely populated areas (Musika, 2013). In another study in Kenya, GIS techniques were used to describe the spatial temporal distribution patterns of notified pediatric cases in urban Counties in Kenya. The study revealed a number of Counties with a high-high relationship with each other a likely indication of clustering (Kibet, 2014). Despite the demonstrated utility of GIS in TB prevention and control programs, there has been limited application within the highly burdensome regions in Kenya at a small area level. In order to fill this crucial gap, this study examined the geographical distribution of TB in Siaya and Kisumu Counties using GIS techniques.

2.2.2 Cluster detection

Fundamental to public health is the investigation of underlying disease clusters. Cluster analysis provides opportunities for the epidemiologist to understand possible associations between demographic and environmental exposures and the spatial distribution of diseases (Kulldorff and Nagarwalla, 1995).

In recent years, GIS with spatial statistics including cluster analysis has been applied to analyze and visualize the spatial patterns of TB. For example GIS spatial analysis techniques were used to detect the TB hotspots in Almora district of India and found significant high-rate spatial clusters in three areas of the district (Tiwari *et al*., 2006). The analysis tools were also used to investigate TB patterns in Linyi City, China revealing significantly high-rate clusters in different districts (Wang Tao *et al*., 2012).

In china, Zhao and colleagues studied TB notification rate including new sputum smear-positive (SS+) notifications by using GIS and spatial statistics and found some areas with significant TB clusters (Zhao Fei *et al*., 2013). In South Africa, GIS and spatial analysis were utilized to study transmission patterns of the disease in a high-incidence area (Munch *et al*., 2003). In an urban West Africa, GIS techniques were used to investigate TB clusters in Greater Banjul region in Gambia. Significant primary and secondary TB clusters were identified settlements in Greater Banjul (Touray *et al*., 2010).

Research findings from these studies provided valuable information that may assist health departments to develop a better preventive strategy and increase the public health intervention's effectiveness in the study area (Wang *et al*., 2012). Whether or not TB occurrence in the high burden Counties of Siaya and Kisumu was random, regular or clustered was unclear. As such, this study sought to describe the pattern of TB and identify TB clusters and hotspots areas using GIS techniques.

2.2.3 Spatial correlation

The interest in geographic correlation lies in determining whether a risk factor is associated with the geographic variation of a disease. Although geographic correlation cannot reveal any causal relationships, it often provides important clues for associations that may be worth more rigorous studies, for example in a case-control setup (Rezaeian *et al*., 2007). A variety of environment related factors (poverty, urban residence, population density, HIV epidemic and proximity and persistence of contacts) have been shown to play a role in the geographical distribution of TB (Lopez De Fede *et al*., 2008; Moonan *et al*., 2004; Muniyandi and Ramachandran, 2008).

For example, a study in the US using GIS techniques found that TB incidence rates were significantly higher in high poverty/high social deprivation areas ($P < 0.0001$) (Lopez De Fede *et al*., 2008) compared to low social deprivation areas. These findings are supported by another study in South Africa which also found significant association of TB notifications with high social deprivation areas (Munch *et al*., 2003) and in Madagascar where TB was found to be associated was associated with socio-economic (Randremanana *et al*., 2009).

The concurrent HIV epidemic is thought to influence the geographical distribution of TB. This is because the risk of developing TB is said to be 20 times higher in a HIV infected person than in a person not infected with HIV (Selwyn *et al*., 1989; Van Asten *et al*., 2003). This is because the weak immune system of HIV-infected persons often results in the reactivation of latent TB infection. A study by Cicero and colleagues in Portugal found that some areas were at higher risk of PTB than others because of high incidence of HIV and AIDS (Couceiro *et al*., 2011).

High population density areas like towns, slums, market centers and prison facility potentially increases exposure of susceptible individuals to infectious TB cases(Canadian Tuberculosis, 2007). It is for this reason that occurrence of TB has majorly been linked to urban areas than rural areas due to the high population densities in urban areas. In the US, Moonan and colleagues used GIS technology to identify geographical risk factors for TB transmission. Population density, poverty and overcrowding were found to be to be major factors for TB disease transmission within the study setting (Moonan *et al*., 2004). Information generated from such studies is useful in identifying high-risk neighborhoods for targeted TB prevention efforts.

Whether or not TB occurrence in Siaya and Kisumu Counties was correlated with environmental factors was unclear. As such, this study sought to evaluate the correlation of TB and residency (urban or rural), population density and type of settlement (slum dwelling or presence of informal settlement) in Kisumu and Siaya Counties.

2.3 Conceptual Framework

Global and country level data reveal significant variations in TB occurrence. These variations can be attributable to underlying differences in geographic, characteristics of the population such as urban residency, high population density and slum dwelling. Urban residence has been associated with greater risk of TB compared to rural residence. Crowding (which is indicated by high population densities) and informal settlement (such as slums), that are associated with poor ventilation are proximate factors that influence the distribution of TB.

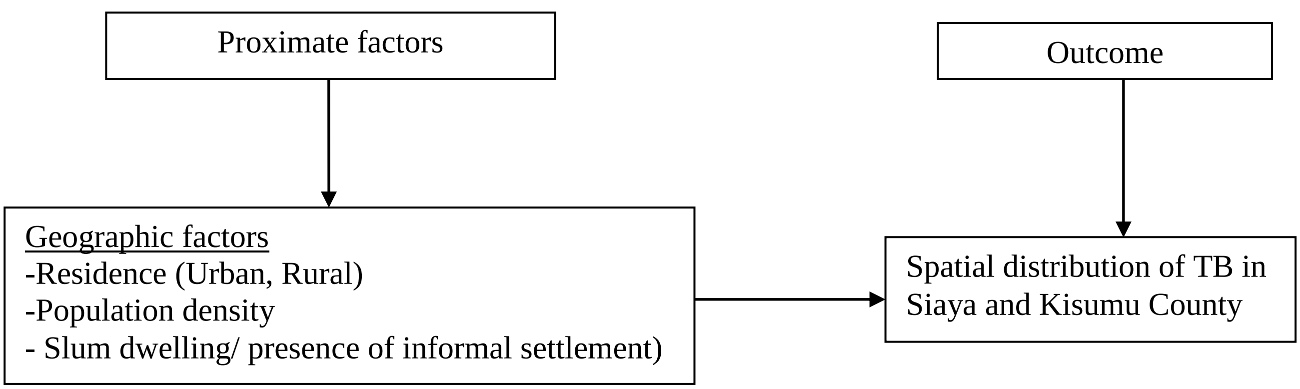

Figure 1: Conceptual Framework (Source: from the literature Review).

CHAPTER 3: METHODOLOGY

3.1 Study Site

This study was done in Kisumu and Siaya Counties, which are two of the 47 devolved Counties of Kenya. The Counties are headquartered in Kisumu City and Siaya town, respectively. The Counties are made up of 12 Sub-Counties (Constituencies) and covering a combined area of 5,301 km^2 along the shores of Lake Victoria. The population of Kisumu and Siaya Counties is estimated at 968 909 and 842, 304, respectively (according to the 2009 Kenya national census). The region has a mixture of urban and rural areas with Kisumu County being the most urbanized.

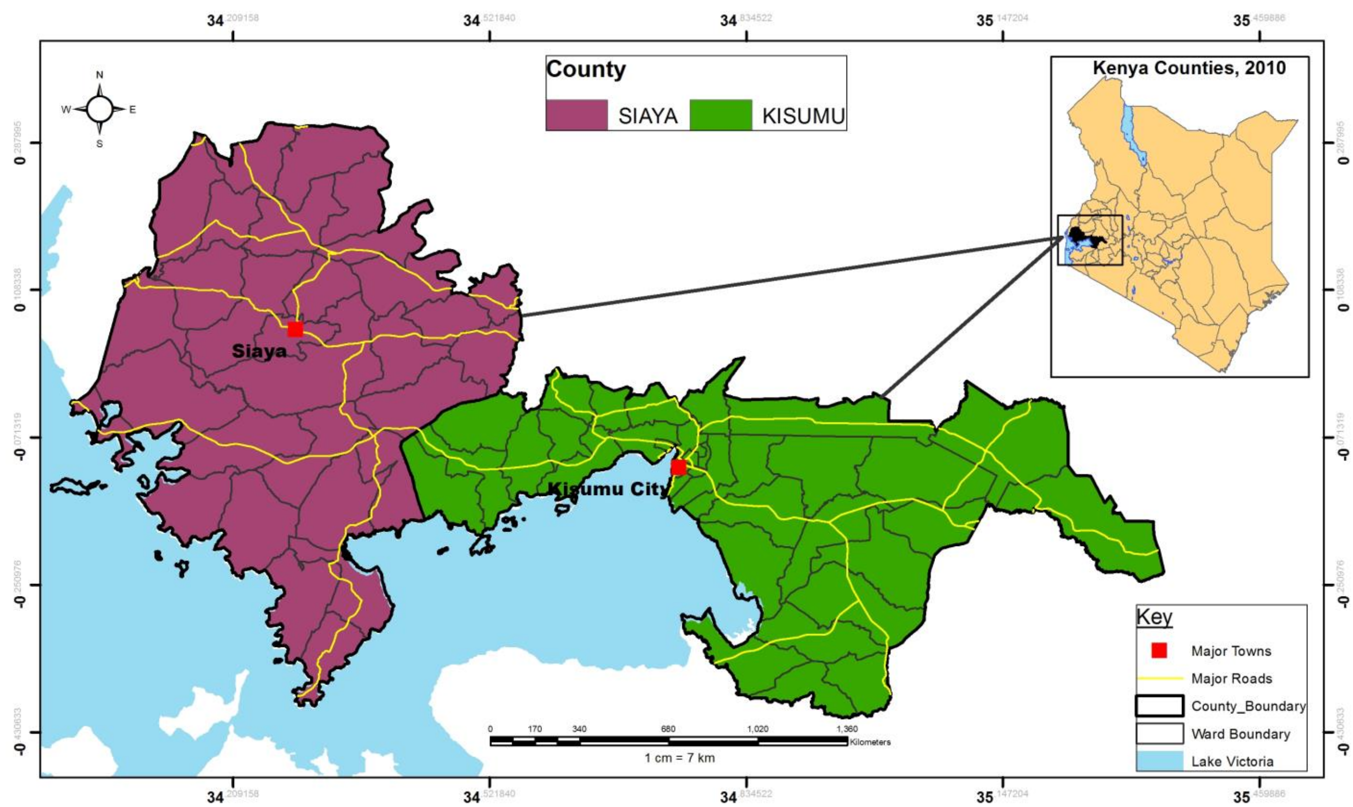

Figure 2: Map of the study area showing County Wards

3.2 Study Design

A descriptive retrospective cross-sectional study was conducted to analyze the spatial distribution of TB and to identify geographic factors associated with TB occurrence in Kisumu and Siaya Counties. Saturated sampling was used in enrolling 5,568 individuals treated with TB in the year 2013.

3.3 Study Populations

The study population consisted of all notified TB cases in health facilities within Siaya and Kisumu Counties between 1st January and 31st December 2013. As the end of 2013 there were a total of 5,568 notified TB cases from 237 health facilities.

3.3.1 Inclusion Criteria

i. Confirmed TB cases registered at the various TB clinics within Kisumu and Siaya Counties between 1st January and 31st December 2013.

ii. Confirmed TB cases at the various TB clinics in Kisumu residing within Kisumu and Siaya Counties.

3.3.2 Exclusion Criteria

i. Confirmed TB cases registered at the various TB clinics within Kisumu and Siaya Counties with missing or inconclusive physical address.

3.4 Research Variables

3.4.1 Independent Variables

The independent variables were residency (Urban or rural), population density and type of settlement (slum dwelling or presence of informal settlement).

3.4.2 Dependent Variables

The dependent variable was the geographical occurrence of TB.

3.5 Sampling technique

Saturated sampling technique was used to obtain TB patient data for Kisumu and Siaya Counties. Since November of 2012, Kenya has been implementing an electronic register system known as "TIBU" coined from a Swahili word meaning "to treat". Under TIBU, all facility level data from the registers is collected electronically with mobile computer tablets and uploaded into the central database of the Division of Leprosy, Tuberculosis and Lung Disease. The use of saturated sampling for this study was therefore appropriate and feasible as it would allow for the near complete coverage of the entire population.

3.6 Data processing and Analysis

3.6.1 TB Data

TB patient details (clinic name, patients date of registration, patients age, gender, type of TB diagnosed, HIV status, county of residence and physical address) was extracted from the TIBU registers and entered into a Microsoft excel spreadsheet. Information on TB diagnosis (how diagnosis was made) was also contained in the registers. Cases were geocoded to the lowest formal administrative unit (sub-location) based on their physical address. Individual data points were summarized into a single value to produce the total number of TB cases per Ward assembly. Aggregation at the Ward level was done to eliminate biases associated with smaller denominators (when calculating percentages and rates).

3.6.2 Population Data

The Kenya population census data for 2009 was used as the base data for estimating population for 2013. The population estimates was obtained by extrapolation of the census 2009 population using a growth rate of 2.8076% for the 2010 data, 2.8439% for the year 2011 for the 2.7% for the year 2012 and 2.69% for the year 2013. The prevalence for each county Ward was then normalized using the population sizes for each county i.e. notified TB rate = (number of cases/population)*100,000.

3.6.3 Cartographic Boundary Files

The cartographic boundary files for Sub locations tracts were downloaded from the International Livestock Research Institute website (ILRI, 2015) and used for displaying spatial patterns of TB occurrence. Information on the current Ward boundaries was obtained from the Independent Electoral and Boundaries Commission (IEBC) website (IEBC, 2015). Sub-locations that make up the county Ward were aggregated within the ArcGIS® environment to create new shape files boundaries for the Wards.

3.7 Visualization

The aggregated TB counts at the Ward level (TB counts) were exported to the GIS environment for generation of maps. The ArcGIS 10® software was then used to come up with choropleth maps as a means for visualizing this data. Choropleth maps show information by “filling” (colouring) each component area with colour, providing an indication of the magnitude of the variable of interest (Wu *et al*., 2004).

3.7 Geographical analysis and cluster detection

Geographical analysis and cluster detection plays an important role in quantifying geographic variation patterns. It is commonly used in disease surveillance to detect aggregation of disease cases, to test the occurrence of any statistically significant clusters, and for hypothesis generation to specify realistic and testable explanations for the geographic patterns. Broadly, there are two types of spatial clustering methods, global and local ones. The global method is used to identify the presence of clustering in the whole study area but does not identify location of the clusters. This shortcoming is overcome by using the local method, which can locate the clusters. The Moran's Index and Local Indicator of Spatial Association (LISA), two commonly used spatial statistical instruments, were applied in this study for the detection of global and local spatial autocorrelation respectively (Anselin, 1995).

3.7.1 Global cluster detection

Moran's Index is a classic indicator of global spatial autocorrelation. It signifies whether clustering of values exists and gives a summary value for all observations in the study area as a whole. Moran's Index is interpreted very much like a correlation coefficient. The range of possible value is +1 to -1. Index near +1 indicates a strong spatial pattern, meaning similar values are located near one and other. Index near -1 indicates a strong negative spatial pattern, meaning dissimilar values are located near one and other. Index near zero indicates the absence of spatial autocorrelation, meaning the distribution of value is random. Moran's Index is defined by

$$I = \frac{\frac{N}{S_0} \sum_i \sum_j W_{ij} Z_i Z_j}{\sum_i Z_i^2}$$

where

Z_i is the deviation of the variable of interest with respect to the mean;

W_{ij} is the matrix of weights that in some cases is equivalent to a binary matrix with ones in position i,j whenever observation i is a neighbor of observation j, and zero otherwise; and

$$S_0 = \sum_i \sum_j W_{ij} .$$

A score of zero indicates no clustering. A positive score indicates clustering of areas of similar attribute values, whereas a negative value indicates that neighboring areas tend to have dissimilar attribute values (Lai *et al*., 2008).

3.7.2 Local cluster detection

While Moran's Index, as a global measure, tells us if there exists spatial clustering in the distribution of values, LISA, as a local measure, highlights specific areas where the cluster of values are found. LISA (such as Anselin Local Moran's I, Getis-Ord Gi*) decomposes the global measure into its contribution for each spatial unit and then detects similarities or dissimilarities in values of a spatial unit in relation to its neighbourhood. Instead of a summarized value, the results of LISA are generated for specific areas within the study area (Anselin, 1995).

Equation 1: Anselin Local Moran's I formulae

$$I_i = \frac{Z_i}{m_2} \sum_j W_{ij} Z_j$$

where:

$$m_2 = \frac{\sum_i Z_i^2}{N}$$

then,

$$I = \sum_i \frac{I_i}{N}$$

I is the Moran's I measure of global autocorrelation, I_i is local, and N is the number of analysis units in the map.

The Getis-Ord Gi*Hot Spot Analysis is a spatial cluster detection method which identifies statistically significant spatial concentrations of the high and of low values associated with a set of geographic features. The current study used the The Getis-Ord Gi* tool within the ArcGIS environment to detect TB “hot spots” and “cold spots”. The Hot Spot uses the Getis- Ord Gi* algorithm which is given as follows:

Equation 2:Getis-Ord Gi* statistic

The Getis-Ord local statistic is given as:

$$G_i^* = \frac{\sum_{j=1}^{n} w_{i,j}x_j - \bar{X}\sum_{j=1}^{n} w_{i,j}}{S\sqrt{\frac{\left[n\sum_{j=1}^{n} w_{i,j}^2 - \left(\sum_{j=1}^{n} w_{i,j}\right)^2\right]}{n-1}}} \quad (1)$$

where x_j is the attribute value for feature j, $w_{i,j}$ is the spatial weight between feature i and j, n is equal to the total number of features and:

$$\bar{X} = \frac{\sum_{j=1}^{n} x_j}{n} \quad (2)$$

$$S = \sqrt{\frac{\sum_{j=1}^{n} x_j^2}{n} - \left(\bar{X}\right)^2} \quad (3)$$

The G_i^* statistic is a z-score so no further calculations are required.

The hot spot analysis tool works by looking at each feature within the context of neighboring features. If the value for a feature is high, and the values for neighboring features are also high, we've found a hot spot. The Hot Spot Analysis tool finds statistically significant clusters of either high or low values.

3.8 Geographic and Statistical Correlation

Population-based ratios were computed for strata of county Wards grouped by level of urbanization and population density. The indicator for urbanization was based on presence of urban centers as defined by the Kenyan Urban Areas and Cities Act, 2011. The Act defines an "urban area" as a Municipality or a town typically with populations >10,000 people. Population density was computed by dividing the population of each of the county Ward with the geographical area covered by the Ward (Km^2). Data was first tested for normality. Spearman's test was used to correlate TB rate and population density while the Wilcoxon rank-sum was applied to compare TB mean between the urban Wards and rural Wards. Relationship between

TB and proximity slums and informal settlements was achieved by overlaying TB rates to a map showing the location of informal settlements.

3.9 Ethical considerations

Ethical approval was granted by the Maseno University Ethical Review Committee, REF: MSUlDRPI/MUERC/000162/15 (Appendix 1). Permission to use TB register data was sought from the Division of Leprosy, Tuberculosis and Lung Disease (DLTLD). The data obtained from the TB registers was stripped of personal identifiable information names. No social risk events were anticipated and adverse events were not relevant to this study, as this was a minimal risk study. The data was stored a password protected computer with limited and restricted access. Results from this study are presented in aggregated form and no individual identifiable information appears in the final analysis or reports.

CHAPTER 4: RESULTS

4.1 Spatial distribution of TB in Siaya and Kisumu Counties

In this objective, individual cases from the registers were summarized into a single value to produce the total number of TB case counts per Ward assembly. A map of the study area in which the Wards are colored in proportion to the total counts in the area was generated. The general demographic characteristics of the TB cases included in this analysis are presented.

4.1.1 Descriptive Analysis of TB Case Notification

A total of 5,568 tuberculosis patients were abstracted from combined TB clinic registers (n=237) in Siaya and Kisumu Counties in 2013. Of these, 5,063 (91.0%) were geographically linked to residencies within the study area using physical address information from the registers. A total of 63 cases (1.1%) had missing locator information whereas 442 cases (7.9%) were linked to residences outside of the study area or had inconclusive addresses. Table 1 below, shows the general characteristics of the 5,063 TB cases included in this study. Overall, there were more females (53.8%) than males (46.2%) with the mean (SD) age of 33.3 years. Majority of the cases (82.5%) were between ages 15-54 years.

A total of 4,248 (83.9%) presented with pulmonary TB while 815 (16.1%) had extra pulmonary TB. Kisumu county had the highest percentage of notified TB cases at 61.1% (3,060/5,063) compared to Siaya county with 38.9% of the cases (1142/5063). Kisumu Central within Kisumu County recorded the highest number of cases 22.8 % (n=1, 142) accounting for 37% of all registered cases within Kisumu County. In terms of HIV status, majority of the TB cases (66.5%), were HIV positive. Only 2.2% of the TB cases did not have a known HIV status.

Table 1: Characteristics of cases collected

Characteristics	% (n/N)
Age at registration	
<14	8.3 (419/5063)
14-24	15.5 (786/5063)
25-34	34.5 (1748/5063)
35-44	21.4 (1084/5063)
45-54	11.1 (563/5063)
>=55	9.2 (463/5063)
Sex	
Male	46.2 (2,342/5063)
Female	53.8 (2,721/5063)
Type of TB	
Extra pulmonary	16.1 (815/5063)
Pulmonary	83.9 (4248/5063)
County of residence	
Siaya County	38.9 (1,970/5063)
Kisumu County	61.1 (3,093/5063)
HIV Status	
Negative	31.2 (1,578/5063)
Positive	66.5 (3,367/5063)
Unknown	2.2 (118/5063)

Key: Unknown-HIV test not done

4.1.2 Geographical distribution of notified TB in Siaya and Kisumu Counties

TB cases were found in each of the 65 Wards with distinct geographical variations observed. The average number of TB cases was 77 per Ward. The highest number of TB cases (n=230) was found in Nyalenda A Ward while the lowest number of TB case (n=15) was found in South Uyoma Ward. An isolated high count of TB cases was also observed in South East Nyakach (bottom right Ward of the study area) with 162 cases. Figure 3 shows the distribution of TB case counts per Ward assembly in Siaya and Kisumu Counties, 2013.

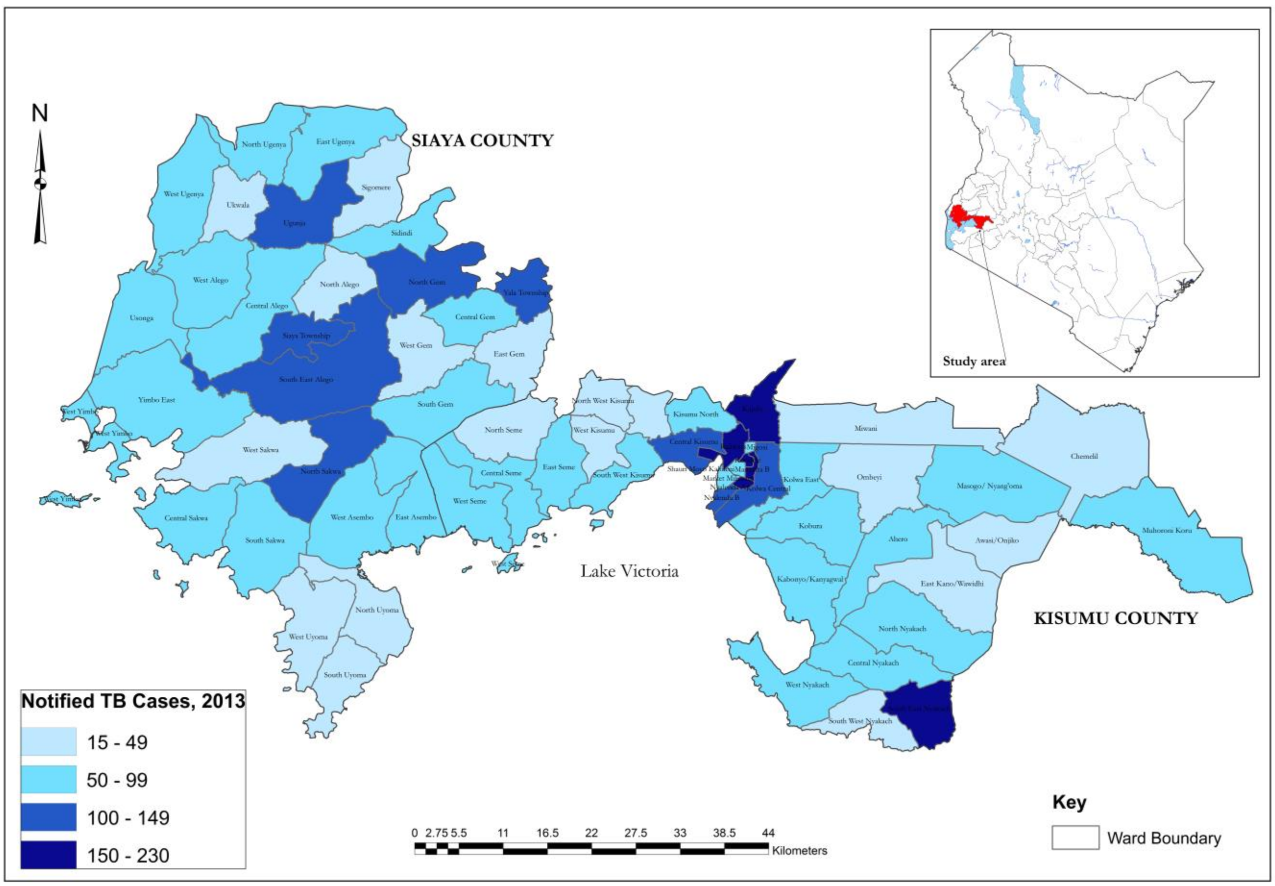

Figure 3: A Choropleth map showing the distribution of notified TB cases

When standardized by the population, the notified TB rate for the entire study area was 278 cases per 100,000 population. Areas around the major towns (Kisumu City and Siaya town) had higher TB case counts as compared to the rest of the study area. Geographically the notified TB rates varied significantly from 76 per 100, 000 in South Uyoma to 813 per 100, 000 populations in Nyalenda A. Figure 4 shows the distribution of TB per 100,000 populations at the Ward assembly level in Siaya and Kisumu Counties in 2013.

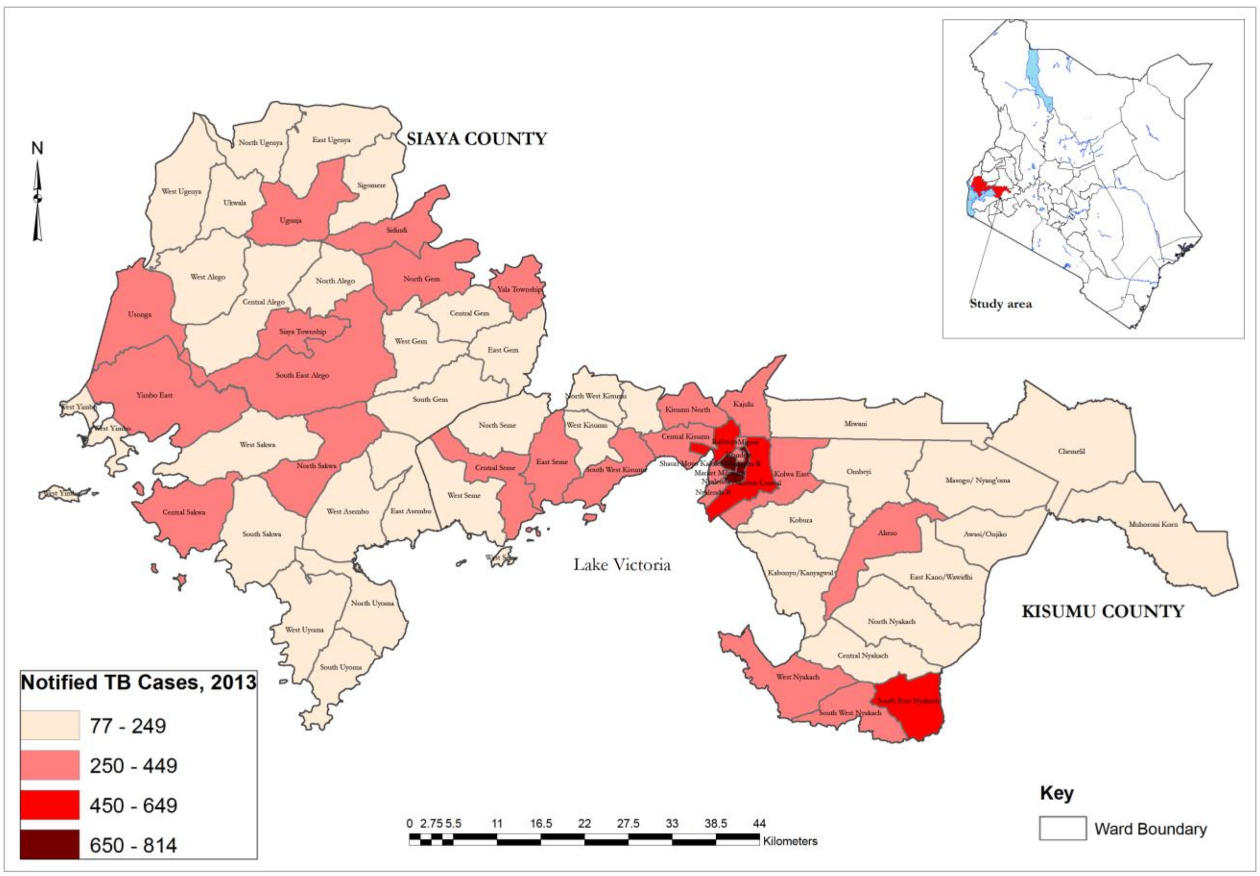

Figure 4: A Choropleth map showing the distribution of notified TB per 100,000 populations

Three of the Wards located within Kisumu City in the central part of Kisumu County (Nyalenda A, Nyalenda B and Kaloleni Wards) had TB rates exceeding 650 per 100,000 populations. The surrounding Wards (Central Kisumu, Kisumu North, Kajulu, Migosi, Kondele and Kolwa East) had equally high TB rates of over 450 per 100,000 populations. The southern part of Kisumu County, which borders the neighboring Homabay County had higher TB rates compared to the surrounding Wards in Kisumu County.

4.2 Detection of TB Clusters and Hot spots

In this objective, two types of spatial clustering methods (global cluster detection and local cluster detection) were used to identify the presence of TB clusters and hotspots in the study area. First, the global method was applied aiming to initially detect if a TB clustering pattern should be further analyzed. The Global Moran's Index is interpreted very much like a correlation coefficient. The range of possible value is +1 to -1. Index near +1 indicates a strong spatial pattern, meaning similar values are located near one another. Index near -1 indicates a strong negative spatial pattern, meaning dissimilar values are located near another. Index near zero indicates the absence of spatial autocorrelation, meaning the distribution of value is random. Local methods (Anselin Local Moran's I, Getis-Ord Gi*) were then applied to identify exact location where the clusters of values are found. The results are presented below.

4.2.1 Global cluster detection (Global Moran's *I*)

There was a positive and spatial autocorrelation of TB prevalence in the study area (Moran's I=0.423, p-value=0.000) implying spatial clustering of TB values. Table 2 provides a summary of Global Moran's I output.

Table 2: Global Moran's I Summary

Moran's Index:	Expected Index:	Variance:	z-score:	p-value:
0.4238	-0.015625	0.002185	9.400212	<0.001

The higher the absolute value of Global Moran's ***I*** is, the stronger a spatial autocorrelation exists. The interpretation of the Global Moran's I statistic is that given the z-score of 9.4, there is a less than 1% likelihood that this high-clustered pattern could be the result of random chance.

4.2.2 Local cluster detection (Anselin Local Moran's I)

While the Global Moran's calculates a single value, which applies to the entire data set, LISA techniques such as the Anselin Local Moran's I, calculates a value for each observation unit and highlights specific areas where the clusters of values are found. A High-High (HH) relationship means a Ward with high TB rate is surrounded by Wards with high TB rate. A Low-Low (LL) relationship means a Ward of low TB rate is surrounded by Wards with low TB rate. A Low-High (LH) relationship means a Ward of low TB rate is surrounded by Wards with high TB rate. A High-Low (HL) relationship means a Ward of high TB rate is surrounded by a Ward with low TB rate.

The study identified 7 Wards with a HH relationship with the surrounding Wards indicating clustering of high values. There were 2 Wards (West Uyoma and South Uyoma) on the southern tip of Siaya County with LL relationship indicating clustering of low values. Areas with LL relationship are often treated as cold spots. Figure 6 shows the location of TB clusters and outliers in Siaya and Kisumu Counties in 2013.

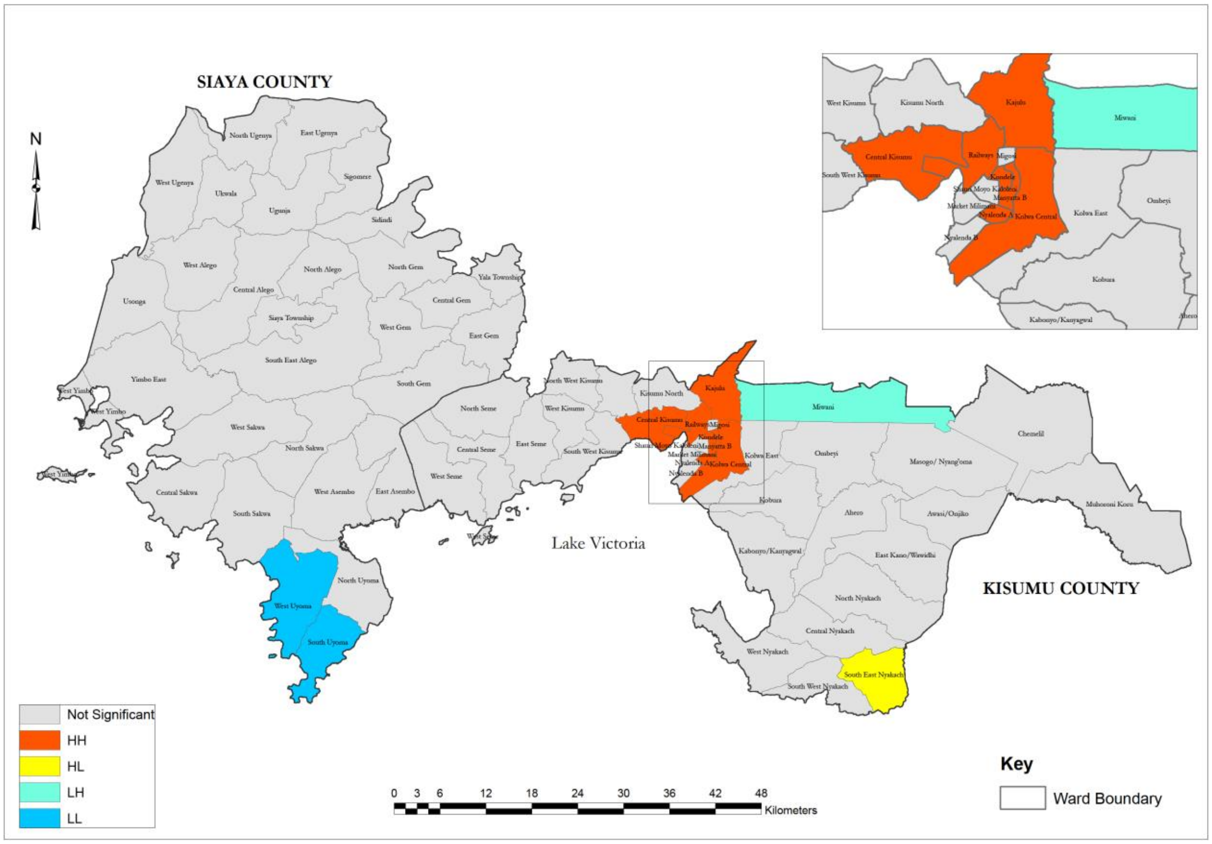

Figure 5: A map showing location of TB clusters and outliers in Siaya and Kisumu Counties, 2013.

Since Low-high clusters and High-Low clusters are usually treated as spatial outliers, investigation of spatial pattern usually focuses on the High-High cluster and Low-Low cluster, which is regarded as hotspot and cold spot, respectively.

4.2.3 TB Hot spot detection (Getis and Ord's local statistic)

The Getis-Ord Gi index identified statistically significant spatial clusters of high values (hot spots) and low values (cold spots). Wards with a z-score value of less than -1.96 and greater than 1.96 were found to be statistically significant at 0.05 significance level. Statistical significance in the negative direction implies a cold spot while in the positive direction implies hot spots. Figure 4 shows the TB hotspots and cold spot areas in Siaya and Kisumu Counties, 2013.

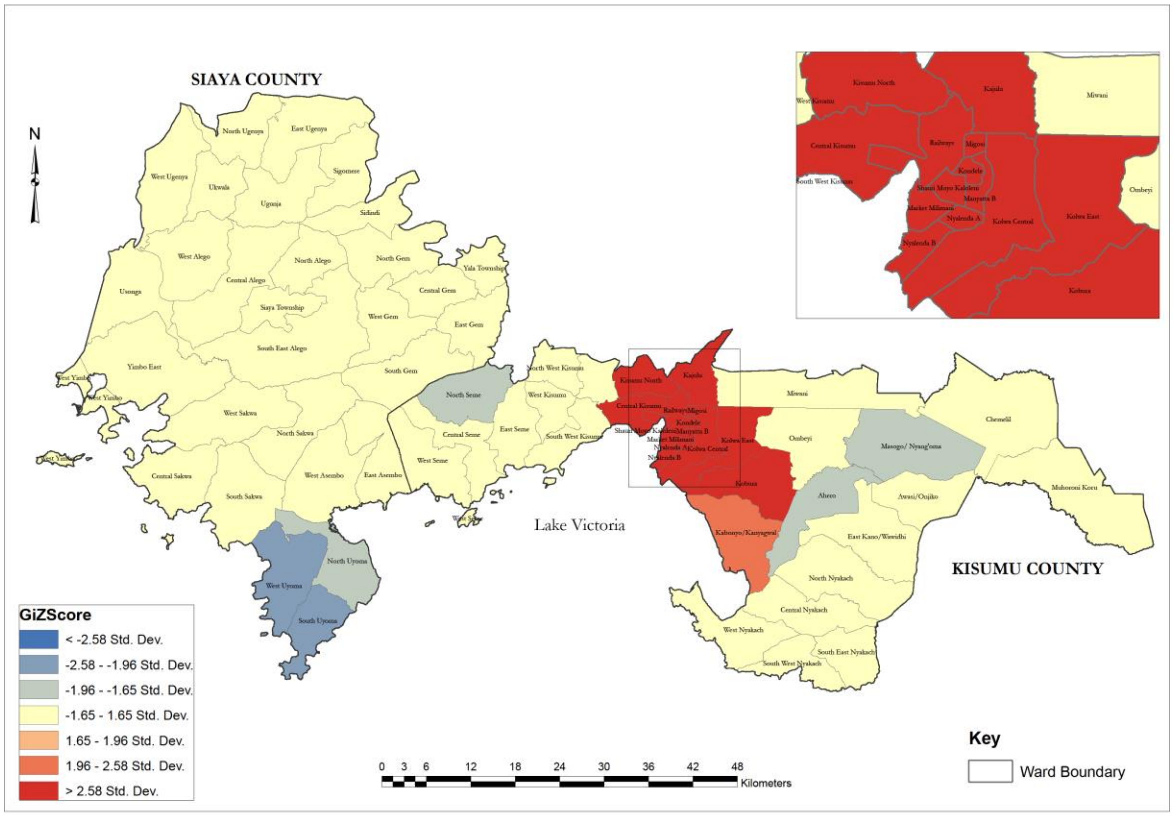

Figure 6: A map showing the hot spot areas in 2013.

Figure 6 shows varying z-scores for the different County Wards. Majority of the Counties had z-score ranging between -1.65 to 1.65. A total of 14 County Wards around Kisumu City had z-score of > 2.5 (hotspots) West Uyoma and South Uyoma has z- score of between-2.58 and -1.96 implying a statistically significant cold spot (cluster of low values).

4.3 Correlation of TB with geographic factors

In this objective, TB rates per 100,000 populations were computed for each County Ward grouped by level of urbanization and population density. A simple histogram of the data showed that the data was not normally distributed. Spearman's test was used to correlate TB rate and population density while the Wilcoxon rank-sum was applied to compare TB mean between the urban Wards and rural Wards. Results from this study showed a strong positive correlation between population density and the rate of TB, which was statistically significant ($r_s = 0.5739$, $p = 0.0001$). In terms of urbanization, the study found higher TB rates in urban areas compared to rural areas. The mean TB rate in the rural areas was 2.2, in the rural area and 3.9, within the urban area with the Wilcoxon rank-sum ($p=0.0001$).

A key limitation here was that some important risk factors were excluded from this analysis, such as, poverty and concurrent HIV epidemic at the population level mainly due to the lack of corresponding data at the Ward level. Nevertheless an attempt was made to relate the occurrence of TB to informal settlement/slum dwelling (Appendix 3). This was done by overlaying the informal settlement to the map showing the distribution of TB in the study area.

CHAPTER 5: DISCUSSION

5.1 Spatial distribution of TB in Siaya and Kisumu Counties

This research examined the geographical distribution of TB within the high burden regions of Siaya and Kisumu Counties, Western Kenya. A key strength of this study was that it was able to reveal specific geographic areas with elevated TB cases within the larger high burdensome region. Knowledge of the geographical distribution of TB disease at a small area level is important in informing prevention and control strategies and in allocation of resources towards these efforts.

The areas around the major towns of Kisumu and Siaya revealed greater concentration of TB rates. This is consistent with previous studies associating high TB occurrence with towns that have experienced increased urbanization and immigration (Lopez De Fede *et al*., 2008; Moonan *et al*., 2004; Muniyandi and Ramachandran, 2008). The southern part of Kisumu County, which borders the neighboring Homa Bay County had higher TB rates compared to the surrounding Wards in Kisumu County. This region (South East Nyakach) is located close to an industrial town (Sondu) that has cropped up due to the presence of a power plant (Sondu Miriu power station). This could potentially explain the higher TB rates in this region resulting probably from the high population of workers and traders in the town.

The observed differences in TB occurrence within Kisumu and Siaya Counties provide important information that could give clues as to etiology, which may then be further studied at individual-level studies such as cohort–control or case–control. Since the observed patterns or differences in occurrence may change over time, this study could serve as a tool for monitoring effectiveness of interventions or used to monitor change in TB patterns, when similar data is mapped in real time.

A limitation of this study is that it relies on data from official surveillance and therefore cannot exclude the possibility that some facilities may not have submitted their data to TIBU for various reasons. There is also the possibility of TB cases being missed by routine notification systems since not all people with TB seek care but remain undiagnosed, or are diagnosed by providers that do not necessarily report cases to the national system. Another limitation is that not all cases could be geocoded to the respective residences due to inconclusive or missing physical location, although this was a very small percentage. There is also a likelihood of over-estimation of cases at a given Ward given that the data is cumulative and included cases with a death outcome or transfer out. Moreover, only cases from facilities within Siaya and Kisumu Counties were used in this study. There is a high likelihood of TB cases residing in the study area but seeking care from facilities outside the area.

It is, however, important that TB programs identify communities with the highest burden so as to employ aggressive prevention measures (such as intensive case finding and contact tracing) with a goal to interrupt ongoing disease transmission.

5.2 Detection of TB Clusters and Hot spots

The current study also identified TB clusters and hotspots within the high burden region. The global Moran's Index revealed a tendency towards clustering thus allowing for further analysis of TB patterns in the study area. Two LISA methods, the local Moran's I and Getis-Ord Gi, revealed specific County Wards where the cluster of TB values are found. Similar techniques cluster detection techniques have been used in identifying TB clusters in China (Wang Tao *et al*., 2012), South Africa, (Munch *et al*., 2003), Gambia West Africa (Touray *et al*., 2010).

In the present study, the local Moran's I identified 7 primary clusters significant spatial clusters for high incidence of TB. The 7 represent County Wards with high TB rates surrounded by Wards with equally high TB rates. The Getis-Ord Gi index identified 14 County Wards with statistically significant spatial clusters of high values (hot spots). Information on identified TB hotspots and clusters could be used for targeted screening and intensive case finding with a goal of reducing TB incidence.

The cold spot were identified in West Uyoma, South Uyoma and North Uyoma. Other areas were identified in South Uyoma Wards within Siaya County. It is still not clear the reasons for the low TB counts in the two areas given that the area has a number of fishermen population and is a transport corridor for those using the ferry to south Nyanza. One explanation that could potentially be investigated further could be the fact that the fishermen in the area seek treatment elsewhere and thus was not captured by the facilities within Siaya and Kisumu Counties. There was one Ward with a HL relationship (Miwani) and one other Ward with a LH relationship in South East Nyakach. As mentioned earlier, this region (South East Nyakach) is located close to an industrial town (Sondu) that has cropped up due to the presence of a power plant (Sondu Miriu power station). The higher TB rates in the study area could potentially be linked to the presence of the high population of workers and traders in the town.

The information generated by the study presents an opportunity for targeted interventions within the identified clusters and hotspots. This information could also be used in allocation of resources as well as identifying locations for further studies that might inform interventions aimed at reducing TB incidence.

5.3 Geographical and Statistical Correlation

This study explored the association between population density and urbanization with the observed TB patterns. A key finding of this study was the correlation of TB with urban residency. Urban County Wards had a higher TB rate as compared to rural County Wards. Another key finding was the correlation of TB and population density. The study found a strong positive correlation between population density and the rate of TB, which was statistically significant. In support of these results are the findings from other similar studies, which showed that urban residence, and high population density areas play a role in the geographical distribution of TB (Lopez De Fede *et al*., 2008; Moonan *et al*., 2004; Muniyandi and Ramachandran, 2008).

The correlation between TB and these two factors was a crucial finding since it revealed specific areas where the public health intervention is needed. Although geographic correlation cannot reveal any causal relationships, it often provides important clues for associations that may be worth pursuing in more rigorous studies such as in a case-control setup. The utility of geographic correlation lies in hypothesis generation which will usually need validation and replication at the individual level, through cohort, case–control studies (Piantadosi *et al*., 1988). Due to the lack of socio-economic data from the TB register used, relationships between socio-economic factors and TB were unable to be assessed in the current study. Additional data collection and further extensive studies need to be undertaken in order to evaluate the associations between TB and its potential factors in Siaya and Kisumu Counties.

The study, however, sought to explore the relationship between TB occurrence and presence/proximity to informal settlement. This was achieved by overlaying the location of slums to the map of aggregated TB counts within each Ward. The present study found that a total of 5 out of the 7 Wards with major slums and informal settlements had a TB prevalence of >500 cases per 100,000 populations. Although the incidence of TB in informal settlement areas throughout Kenya is not known, informal settlements are considered a high-risk setting given the ease of transmission due to overcrowding, poor lighting and poor ventilation of dwellings, and financial, geographical and social barriers to care (NTLLD, 2015a).

Even though these findings clearly do not demonstrate a causal relationship between TB infection and slum/informal dwelling, it provides opportunity for further studies and lends support to the general observation that SES does affect individual health.

CHAPTER 6: SUMMARY OF FINDINGS, CONCLUSION AND RECOMMENDATIONS

6.1 Summary of Findings

This study revealed distinct TB clusters within a high TB burden region. The distinct TB patterns as revealed by the study, provides an understanding of the high TB risk areas where surveillance and public health is needed.

6.2 Conclusions

i. TB rates in the high burden Counties of Siaya and Kisumu vary geographically at the small area (Ward) level.

ii. The distribution of TB in Siaya and Kisumu Counties was nonrandom and clustered with the significant TB clusters and hotspots identified for the year 2013.

iii. The occurrence of TB was positively correlated with urbanization and population density (crowding) and was statistically significant.

6.3 Recommendations from the current study

i. Adoption of GIS technology in the management and control of TB programs in Siaya and Kisumu Counties to assist in decision making for TB prevention and control programs as well as to monitor impact of various interventions.

ii. Priority of TB interventions should that target regions with highest disease burden for efficient resource management.

iii. Since this kind of analysis heavily relies on correctly linking cases to their respective geographic locations, we further recommend training of staff to ensure accuracy and completeness of the patient locator information.

6.4 Recommendations for further studies

i. Since the effectiveness of TB control measures in specific areas could be assessed by a longitudinal change in TB prevalence, a space-time analysis with 2013 as the baseline could be carried for program evaluation.

ii. More detailed individual level investigations are needed in the identified TB clusters and hotspots to evaluate the most important determinants of disease distribution.

REFERENCES

1. Ali, M., M. Emch, J.P. Donnay, M. Yunus, and R.B. Sack. 2002. The spatial epidemiology of cholera in an endemic area of Bangladesh. *Social Science & Medicine*. 55:1015-24.
2. Anselin, L. 1995. Local indicators of spatial association-LISA. *Geographical Analysis*. 27:93-115.
3. Canadian Tuberculosis, C. 2007. Housing conditions that serve as risk factors for tuberculosis infection and disease. An Advisory Committee Statement (ACS). *Canada Communicable Disease Report Volume 33. 2007 Oct 1;33(ACS-9):1-13.*
4. Clarke, K.C., S.L. McLafferty, and B.J. Tempalski. 1996. On epidemiology and geographic information systems: a review and discussion of future directions. *Emerging Infectious Diseases*. 2:85-92.
5. Couceiro, L., P. Santana, and C. Nunes. 2011. Pulmonary tuberculosis and risk factors in Portugal: a spatial analysis. *The International Journal of Tuberculosis and Lung Disease*. 15:1445-1455.
6. Cuadros, D., S. Awad, and L. Abu-Raddad. 2013. Mapping HIV clustering: a strategy for identifying populations at high risk of HIV infection in sub-Saharan Africa. *International Journal of Health Geographics*. 12:28-29.
7. Dye, C.W., Brian G. 2008. Eliminating human tuberculosis in the twenty-first century. *Journal of the Royal Society Interface*. 5:653-662.
8. Eisinger, D., and H.H. Thulke. 2008. Spatial pattern formation facilitates eradication of infectious diseases. *Journal of Applied Ecology*. 45:415-423.
9. Elliott, P., and D. Wartenberg. 2004. Spatial epidemiology: current approaches and future challenges. *Environmental Health Perspectives*. 112:998-1006.
10. IEBC. 2015. Independent Electoral and Boundaries Commission website. Accessed on 10 August 2015. http://www.iebc.or.ke/.
11. ILRI. 2015. International Livestock Research Institute (ILRI) website. Accessed on 10 August 2015. http://192.156.137.110/gis/search.asp.
12. Jacquez, G., and D. Greiling. 2003. Local clustering in breast, lung and colorectal cancer in Long Island, New York. *International Journal of Health Geographics*. 2:3-15.
13. Kibet, K.R. 2014. Spatial temporal analysis of the distribution of pediatric tuberculosis patterns in Kenya. Kenyatta University.

14. Kitron, U., and J.J. Kazmierczak. 1997. Spatial Analysis of the Distribution of Lyme Disease in Wisconsin. *American Journal of Epidemiology*. 145:558-566.
15. Kulldorff, M., and N. Nagarwalla. 1995. Spatial disease clusters: detection and inference. *Statistics in Medicine*. 14:799-810.
16. Lai, P. C., F. M. So, and K. W. Chan. 2008. Spatial epidemiological approaches in disease mapping and analysis. CRC Press.
17. Lopez De Fede, A., J.E. Stewart, M.J. Harris, and K. Mayfield-Smith. 2008. Tuberculosis in socio-economically deprived neighborhoods: missed opportunities for prevention. *The International Journal of Tuberculosis and Lung Disease*. 12:1425-1430.
18. Moonan, P.K., M. Bayona, T.N. Quitugua, J. Oppong, D. Dunbar, K.C. Jost, G. Burgess, K.P. Singh, and S.E. Weis. 2004. Using GIS technology to identify areas of tuberculosis transmission and incidence. *International Journal of Health Geographics*. 3:23.
19. Munch, Z., S.W.P. Van Lill, C.N. Booysen, H.L. Zietsman, D.A. Enarson, and N. Beyers. 2003. Tuberculosis transmission patterns in a high-incidence area: a spatial analysis. *The International Journal of Tuberculosis and Lung Disease*. 7:271-277.
20. Muniyandi, M., and R. Ramachandran. 2008. Socioeconomic inequalities of tuberculosis in India. *Expert Opinion on Pharmacotherapy*. 9:1623-1628.
21. Musika, F. 2013. Spatial Visualization and Analysis of Tuberculosis Infection Case of Kitui County. *In* Proceedings of the1st Esri Eastern Africa Education User Conference, 17 – 18 September, 2013; Kenyatta University, Nairobi, Kenya.
22. Newsom, S.W.B. 2006. Pioneers in infection control: John Snow, Henry Whitehead, the Broad Street pump, and the beginnings of geographical epidemiology. *The Journal of Hospital Infection*. 64:210-216.
23. NTLLD. 2015a. Communication Strategy For The Curefor TB 2015.
24. NTLLD. 2015b. National Tuberculosis, Leprosy and Lung Disease Program Website: accessed on 14 September 2015. http://www.nltp.co.ke/
25. Partnership, S.T. 2015. World TB Day 2015: "Reach the 3 Million: Reach, Treat, Cure Everyone" http://www.stoptb.org/events/world_tb_day/2015/ September 14, 2015.

26. Piantadosi, S., D.P. Byar, and S.B. Green. 1988. The ecological fallacy. *American Journal of Epidemiology*. 127:893-904.
27. Randremanana, R.V., P. Sabatier, F. Rakotomanana, A. Randriamanantena, and V. Richard. 2009. Spatial clustering of pulmonary tuberculosis and impact of the care factors in Antananarivo City. *Tropical Medicine & International Health*. 14:429-437.
28. Rezaeian, M., G. Dunn, S. St Leger, and L. Appleby. 2007. Geographical epidemiology, spatial analysis and geographical information systems: a multidisciplinary glossary. *Journal of Epidemiology and Community Health*. 61:98-102.
29. Selwyn, P.A., D. Hartel, V.A. Lewis, E.E. Schoenbaum, S.H. Vermund, R.S. Klein, A.T. Walker, and G.H. Friedland. 1989. A prospective study of the risk of tuberculosis among intravenous drug users with human immunodeficiency virus infection. *New England Journal of Medicine*. 320:545-550.
30. Shirayama, Y., S. Phompida, and K. Shibuya. 2009. Geographic information system (GIS) maps and malaria control monitoring: intervention coverage and health outcome in distal villages of Khammouane province, Laos. *Malaria Journal C7 - 217*. 8:1-8.
31. Snow, J. 1988. On the mode of communication of cholera. *The Challenge of Epidemiology: Issues and Selected Readings. Washington, DC: World Health Organization*:42-45.
32. Tiwari, N., C.M.S. Adhikari, A. Tewari, and V. Kandpal. 2006. Investigation of geo-spatial hotspots for the occurrence of tuberculosis in Almora district, India, using GIS and spatial scan statistic. *International Journal of Health Geographics*. 5:1.
33. Tiwari, N., V. Kandpal, A. Tewari, K.R.M. Rao, and V.S. Tolia. 2010. Investigation of tuberculosis clusters in Dehradun city of India. *Asian Pacific Journal of Tropical Medicine*. 3:486-490.
34. Touray, K., I.M. Adetifa, A. Jallow, J. Rigby, D. Jeffries, Y.B. Cheung, S. Donkor, R.A. Adegbola, and P.C. Hill. 2010. Spatial analysis of tuberculosis in an Urban West African setting: is there evidence of clustering? *Tropical Medicine & International Health*. 15:664-672.
35. Van't Hoog, A.H., K.F. Laserson, W.A. Githui, H.K. Meme, J.A. Agaya, L.O. Odeny, B.G. Muchiri, B.J. Marston, K.M. DeCock, and M.W. Borgdorff. 2011. High prevalence of

pulmonary tuberculosis and inadequate case finding in rural western Kenya. *American Journal of Respiratory and Critical Care Medicine*. 183:1245-1253.

36. Van Asten, L., M. Langendam, R. Zangerle, I. Hernandez Aguado, F. Boufassa, V. Schiffer, R.P. Brettle, J.R. Robertson, A. Fontanet, R.A. Coutinho, and M. Prins. 2003. Tuberculosis risk varies with the duration of HIV infection: a prospective study of European drug users with known date of HIV seroconversion. *AIDS. 2003 May 23;17(8):1201-8.*

37. Wakefield, J. 2001. Statistical Methods in Spatial Epidemiology: AB Lawson. Chichester, UK: Wiley, 2001, pp.277. ISBN: 0471975729. *International Journal of Epidemiology*. 30:1504-1505.

38. Wang Tao, Xue Fuzhong, Chen Yongjin, Ma Yunbo, and Liu Yanxun. 2012. The spatial epidemiology of tuberculosis in Linyi City, China, 2005-2010. *BMC Public Health*. 12:885-935.

39. WHO. 2010. Global tuberculosis control: WHO report 2010. World Health Organization, Geneva, Switzerland; 2011.

40. WHO. 2013. Fact Sheets. Vol. 2013.

41. WHO. 2015. Tuberculosis database, Global Tuberculosis Program. Available online: http://www.who.int/tb/country/data/download/en/ Accessed on March 19, 2015.

42. Wu, Ji-lei, Wang, Jin-feng, Zheng, and Xiao-ying. 2004. Spatial exploratory data analysis of birth defect risk factors' identification. *Journal of Environment and Health*. 6:003.

43. Zhao Fei, Cheng, Shiming He, Guangxue Huang, Fei Zhang, Hui Xu, Biao Murimwa, T. C., J. Cheng, D. Hu, and L. Wang. 2013. Space-Time Clustering Characteristics of Tuberculosis in China, 2005-2011. *PLoS One*. 8:e83605.

APPENDICES

Appendix 1: Letter of approval from MUERC

MASENO UNIVERSITY
UNIVERSITY OF EXCELLENCE

MASENO UNIVERSITY ETHICS REVIEW COMMITTEE

Tel: +254 057 351 622 Ext: 3050
Fax: +254 057 351 221

Private Bag – 40105, Maseno, Kenya
Email: muerc-secretariate@maseno.ac.ke

FROM: Secretary - MUERC

DATE: 19th June, 2015

TO: Peter Mwangala Sifuna
PG/MPH/00071/2012
Department of Public Health
School of Public Health and Community Development
Maseno University, P. O. Box Private Bag, Maseno, Kenya

REF: MSU/DRPI/MUERC/000162/15

RE: Spatial Epidemiology of Tuberculosis in Kisumu and Siaya Counties, Western Kenya. Proposal Reference Number MSU/DRPI/MUERC/000162/15

This is to inform you that the Maseno University Ethics Review Committee (MUERC) determined that the ethics issues raised at the initial review were adequately addressed in the revised proposal. Consequently, the study is granted approval for implementation effective this 19th day of June, 2015 for a period of one (1) year.

Please note that authorization to conduct this study will automatically expire on 18th June, 2016. If you plan to continue with the study beyond this date, please submit an application for continuation approval to the MUERC Secretariat by 16th May, 2016.

Approval for continuation of the study will be subject to successful submission of an annual progress report that is to reach the MUERC Secretariat by 16th May, 2016.

Please note that any unanticipated problems resulting from the conduct of this study must be reported to MUERC. You are required to submit any proposed changes to this study to MUERC for review and approval prior to initiation. Please advice MUERC when the study is completed or discontinued.

Thank you.

Yours faithfully,

DIRECTORATE OF RESEARCH
PUBLICATION & CONSULTANCIES
19 JUN 2015
MASENO UNIVERSITY

Dr. Bonuke Anyona,
Secretary,
Maseno University Ethics Review Committee.

Cc: Chairman,
Maseno University Ethics Review Committee.

MASENO UNIVERSITY IS ISO 9001:2008 CERTIFIED

Appendix 2: Treatment outcomes of TB cases in Siaya and Kisumu Counties

Characteristics			Treatment Outcome									
County	Total		Success [F]		Dead		Failure		LTFU*		TO	
Kisumu	3,093	61%	2,507	81%	289	9%	7	0.2%	150	5%	140	5%
Siaya	1,970	39%	1,603	81%	206	10%	10	0.5%	95	5%	56	3%
Sex												
Female	2,342	57%	1,964	84%	189	8%	6	0.3%	96	4%	87	4%
Male	2,721	66%	2,146	79%	306	11%	11	0.4%	149	5%	109	4%
Age												
<14	419	8%	352	84%	33	8%	0	0.0%	19	5%	15	4%
15-24	786	16%	647	82%	49	6%	2	0.3%	40	5%	48	6%
25-34	1,748	35%	1,398	80%	174	10%	5	0.3%	97	6%	74	4%
35-44	1,084	21%	883	81%	110	10%	5	0.5%	53	5%	33	3%
45-54	563	11%	468	83%	62	11%	1	0.2%	17	3%	15	3%
>=55	463	9%	362	78%	67	14%	4	0.9%	19	4%	11	2%
HIV Status*												
Negative	1,578	32%	1,325	84%	90	6%	7	0.4%	93	6%	63	4%
Positive	3,367	68%	2,690	80%	398	12%	10	0.3%	145	4%	124	4%

Success [F] = cured and 2,548completed treatment

LTFU=Lost to follow up /Out of Control (Defaulted)*

**12 TB cases declined HIV testing, 106 had missing HIV status*

Appendix 3:Map showing location of informal settlements and distribution of TB