UNIVERSITY OF NAIROBI

School of Computing and Informatics

SPARTIAL VISUALIZATION AND ANALYSIS OF TUBERCULOSIS INFECTION: CASE OF KITUI COUNTY

BY

MUSIKA M. FREDRICK - P56/8128/2006

SUPERVISOR

C. A. MOTURI

AUGUST, 2014

Submitted in Partial fulfillment of the Requirements of the Master of Science Degree in Information Systems of the University of Nairobi

DECLARATION

I, **Musika M Fredrick,** confirm that this research project and the work presented in it is my original work and has not been presented for any other University award.

Signature: __________________________ Date: _______________________________

Musika M Fredrick

P56/8128/2006

This project has been submitted in partial fulfillment of the requirement of Masters of Science Degree in Information Systems of the University of Nairobi with my approval as the university supervisor.

Signature: _________________________ Date: _______________________________

Christopher A. Moturi

School of Computing and Informatics

University of Nairobi

DEDICATION

To

My family,

Supervisor,

Lecturers and colleagues,

For your valuable support throughout the whole process

I sincerely cherish you all

May the Almighty God bless you.

ACKNOWLEDGEMENT

First and foremost, I would like to thank the Almighty God for good health energy and wisdom that enabled me to come this far.

Special thanks to my supervisor, Christopher A. Moturi for his tireless support, guidance, patience, understanding and dedication throughout this whole period.

I wish to acknowledge the support received from the project assessment panel members. Your guidance, advice, perspectives and criticism were immensely helpful in shaping this project.

This project would not have been complete without the help of Dr. Felix Mutua and his assistance with the ArcGIS, Mr. Andrew Sila for his support and guidance in R-Geocoding/R-GIS and all the TB–Management Unit, Managers of Kitui County.

I wish to thank my wife Beatrice, sons Carton and Victor for their patient support and sacrifice throughout the period of this project.

May God bless you all.

ABSTRACT

The importance of using technology such as Geographic Information System (GIS) in the study of spread of diseases and its support decision making, health planning and management has brought a paradigm shift from traditional social medicine to a new technology empowered public health care. The growth of medical informatics, decreasing cost in computer systems and availability of medical geo-referenced data can be harnessed and integrated with GIS technology to help study Tuberculosis (TB) Epidemiology. This project was informed by these emerging new developments and the need of information of interest to epidemiologist to capture, store, analyze, visualize and manage TB in Kitui County using a user friendly GIS prototype.

The study models data from five TB management units. The data comprised mainly of secondary data obtained from registers and quarterly reports. The physical addresses of the facilities and the participants were converted into GIS coordinates and exported to ArcGIS/R-GIS for visualization and analysis.

The study showed evidence of high prevalence of TB in major towns, along the major road, and densely populated areas. High death rate was registered not only in high populated areas, but also in areas with poor facility coverage. It also showed correlation between the rate of TB infection and HIV; with 4 out every 10 patient being HIV positive, evident was correlation between high death rate (8%), HIV (29%) and low literacy. Further, there was evidence in lack of criteria in allocation of TB diagnostic and treatment centers and related services.

Keywords: Tuberculosis, clustering, spatial analysis, Geographic Information Systems, Kitui County.

TABLE OF CONTENTS

TABLE OF FIGURES

LIST OF TABLES

ACRONYMS/ABREVIATIONS

CDC	Center for Disease Control
USAID	United States Agency for International Development
TB	Tuberculosis
WHO	World Health Organization
MDR	Multidrug Resistance
DOT	Direct Observed Treatment
DLTLD	Kenya's National Division of Leprosy, TB & Lung Disease
GIS	Geographic Information System
DGPS	Differential Geographic Positioning Systems
IIBRC	Interim Independent Boundaries Review Commission of Kenya
DHRO	District Health Records Officer
TBMU	TB Management Unit
NTP	National TB Progam

DEFINATION OF TERMS AND CONCEPTS

Choropleth Mapping*:* Shows data value for a predefined area such as a census tract (or watershed).

Census block*:* The smallest area for which the Census Bureau collects and tabulates decennial census data.

Census tract*:* A small, relatively permanent statistical subdivision of a county in a metropolitan area, or a selected non-metropolitan county, for presenting decennial census data.

Geo-coding*:* Digital procedure for finding map co-ordinates that correspond to data attributes of features.

Epidemiology*:* This is the study of the distribution and determinants of health-related states or events (including disease), and the application of this study to the control of diseases and other health problems.

Clustering: A part of the topology validation process in which vertices that fall within a specified distance (cluster tolerance) of each other are snapped together.

Spatial: A type of geographical analysis which seeks to explain patterns of human behavior and its spatial expression in terms of mathematics and geometry

Visualization: Visualization is used to organize spatial data and related information into layers that can be analyzed or displayed as maps, three-dimensional scenes, summary charts, tables, time-based views, and schematics.

CHAPTER ONE

INTRODUCTION

1.0 Background

In 1993, the World Health Organization (WHO) declared Tuberculosis (TB) a global emergency. The control of the disease is based on effective diagnosis, treatment and monitoring of TB cases hence, WHO recommended the introduction of directly observed therapy (DOT) WHO (1997).

According to the Center for Disease Control (CDC, 2006), two billion people are infected with Mycobacterium tuberculosis (TB). About two million die from the disease every year (Maher and Raviglione, 2005). TB kills more people in the African continent than anywhere else. It is estimated that TB kills approximately 1,500 lives daily (WHO, 2005). Kenya is ranked 13th in a list of 22 high burden tuberculosis countries in the world and it is the 5th highest burden in Africa. Kenya had approximately more than 132,000 new TB cases per 100,000 populations (WHO, 2009).

A Geographic Information System (GIS) is a computer system for capturing, managing, integrating, manipulating, analyzing and displaying data that is spatially referenced to earth (McDonnell and Kemp, 1995). At first glance, GIS may seem a mare mapping program but in actual sense, it is a complicated integration of database management systems, graphics display technology and an analysis tool. The integration of tabular data and spatial data can be very useful in decision making, policy formulation and monitoring in TB testing and treatment centers (McDonnell and Kemp 1995). Disease mapping has a long history according to Howe (1989). It was the method of descriptive analysis that was first used as an attempt to identify sources of infections and to describe rate of spread. The description of spatial/temporal patterns of disease incidence and mortality can be defined as geographical epidemiology. This can be considered as a part of descriptive epidemiology, which is more concerned with describing the occurrence of diseases with respect to demographic characteristics, place and time (Elliot, 1992).

All stakeholders in the health sector, especially the ones involved in TB testing and treatment need to understand, evaluate, analyze data and continuously update information that is referenced to specific locations. GIS can be used for this because of its ability to display spatial data and also enable decision makers and users in the health fraternity to study epidemiology (study of health-event patterns in a society). It can also help them to plan and map health care facilities in the affected areas, continuously analyze data and share the information to support decision and policy making in the constituency.

Kitui County is one of the 13 counties in Eastern Kenya, covering approximately 20,402 km². It has a population of approximately 1,012,709 as per 2009 population and housing census. With around 92 health facilities owned by government, private investor and the church (mission hospitals), the average distance to the nearest facility is 5 Km. There has been a documented rise of TB prevalence in the county with approximately 2, 000 new TB infection every year. This is according to Kitui District Health Records Officer (DHRO) report, 2010.

1.1 Problem Definition

Over the years, Kitui County has been ravaged by TB infection due to its social economic status, ignorance, lifestyle, HIV among others and more so, lack of information on geospatial spread of TB.
With lack of computerized Information Systems, the current distribution of TB diagnostic, treatment centers and TB health workers is not based on spread of TB, patient location or any other substantial guidelines, but on availability of facilities. Although data is collected from all the testing and treatment centers, the data is then recorded in hard book registers and is not in a format that can be used in decision making. The main purpose of this project is to come up with a spatial model to visually analyze TB cases and thereby assist TB heath stakeholders get up-to-date data, which when analyzed can help them make timely and reliable decisions.

1.2 Main Objective

The main objective of this project is to use GIS application to study TB epidemiology in Kitui County and provide users with abilities to add, spatially visualized and analyze data in a user friendly environment, without much need for training of the users.

1.3 Specific Objectives

The specific objectives of this project were:

- To review relevant literature on GIS applications in the study of TB epidemiology;
- To study the Geographic spread of TB and spatially visualize it using GIS application;
- Use GIS to analyze TB distribution in the county;
- Demonstrate how GIS can be used for epidemiology study in Kenya.

1.4 Research Questions

- Is there relevance in using GIS application in the study of TB epidemiology?
- What is the geographical point of distribution for all reported Tuberculosis cases in Kitui County between 2010 and 2012?
- Is there evidence of patterns, clusters and hotspots of Tuberculosis in Kitui County between 2010 and 2012?
- What is the policy in allocation of facilities, resources and services in the County?

1.5 Project Justification

With the growth of medical informatics in public health, decreasing cost in computer systems and availability of medical geo-referenced data that can be used by GIS, development of GIS Health based Information System is a new and exciting field of interest. This project came up with a user friendly GIS application model to capture, store, analyze and visualize TB data.

This with the aim of providing doctors, clinical officers both in government and private medical institutions with new opportunities to visualize, explore and understand health indicators in the County.

It will also assist policy and decision makers in the County with information they can effectively use to make viable decisions concerning distribution of healthcare resources, such as, health care services, healthcare facilities and other resources. It will also assist stakeholders in evaluating performance of the services and the resources.

1.6 Scope of the Study

The project covered all health facilities both in Government and Mission hospitals in Kitui County which had been identified as TB testing and treatment centers. The system comprises personal computers loaded with ArcGIS (GIS software produced by ESRI co.), to capture, update and analyze the geo-database.

The following information was integrated in the GIS:

- TB occurrence and geo-reference to the nearest primary school, waterhole, church or town;
- TB management unit catchment areas;
- Health facilities and their status;
- Any other essential data to help mangers determine which ones can be used as diagnostics and treatment centers.

1.7 Project Constraints

The researcher encountered the following limitations:

1. Problems in obtaining accurate geographical location and addresses;
2. Difficulty in accessing medical records due to confidentiality issues and government bureaucracy;
3. Unreliability and incompleteness of government records;
4. Challenges posed by the vastness of the county.

CHAPTER TWO

LITERATURE REVIEW

2.0 Introduction

Decision making based on geography is basic human thinking. Where we go, speculation on how it shall be like and plans on what we should do when we get there is applied to the simple event of going to the store or to the major event of launching a bathysphere into the ocean's depths. By understanding geography and people's relationship to location, we can make informed decisions about the way we live on our planet. A Geographic Information System (GIS) is a technological tool for comprehending geography and making intelligent decisions (Getis, 2008).

According to DeMers (2005), GIS is a system that is designed for input, storage, editing, analyzing and output of geographic data and information. It is technology that combines a powerful database with the unique ability to display the database information on a map. This ability to visualize information on a map and allow for quick analysis of the information makes GIS a very valuable tool to public health.

Notably, Getis (2008) noted that Geographical Information System is computer software that links geographic information (where things are) with descriptive information (what things are). Unlike a flat paper map which presents a single layer, GIS present many layers with different information. To use a paper map, all you do is unfold it and what you see is what you get. That is representation of cities and roads, mountains and rivers, rail, roads and political boundaries. The cities are represented by little dots or circles, the roads by black lines, the mountain peaks by tiny triangles and the lakes by small blue areas similar to the real lakes.

A digital map is easier to use than a paper map. For the paper map, features on the map such as cities are represented with dots or points, roads as lines and lakes as small areas. All this information where the point is located, how long the road is and even how many square miles a lake occupies are stored as layers in digital (raster) format as a pattern of ones and zeros in a computer (Getis, 2008).

Nonetheless, when compared with a static map, DeMers (2005) noted that GIS has a major strength and advantage as this technology is the dynamic linkage between tabular and spatial data. For example, when tabular records are selected, corresponding features in a linked map view are also selected and vice versa, enabling easy identification of records or features of interest.

Another GIS capability is “Spatial overlay,” that is, it allows map layers drawn from different GIS databases to be superimposed and displayed as a composite map image. Information contained within different GIS databases can be spatially joined by merging information contained in each database with its geographic area or location. For example, a specific home address can be assigned to its corresponding land parcel, census tract, or zip code (Reissman, *et. al*, 2001).

In addition, Mesgari and Masoomi (2008) noted that GIS is normally associated with maps and there are three way in ways in which it can work with geographic data. That is database view (geo-database), map view (a set of intelligent maps and other views that show features and feature relationships on the earth's surface) and model view (a set of information transformation tools that derive new geographic datasets from existing datasets). Nonetheless, Murayama and Thapa (2011) opined that the use of geography in epidemiologic studies is not a new phenomenon; GIS software is used extensively throughout the medical fraternity to study disease patterns, look at the distribution of health care facilities and map any system that is visual or spatial including the inside a patient's body.

Spatial analysis and mapping epidemiology have a long history. An advance in geographical information and mapping technologies in addition to increased awareness creates new gaps and opportunities. This encourages the health care fraternity to use and integrate GIS with their work.

2.1 Brief History of GIS

Since 1960s, GIS has emerged as a discipline in its own right. This according to DeMers (2005) is based on the premise of its origin in land use applications in Canada to the all-pervasive technology used today in applications as diverse as car-navigation, retail stores site location, customer targeting, risk management, construction, weather forecasting, utility management and military planning. GIS has become ubiquitous in modern life. Its technology development can be traced from a chance meeting between Roger Tomlinson, the Director of Department of Forestry and Rural Development of Canada and IBM Executive. Which resulted to a suggestion of computer technology based solution to mapping problems and hence, the first fully operational geographic information system was built in the Canada Geographic Information System (CGIS) (Chainy & Ratcliffe, 2005).

Importantly, it was noted by Mesgari and Masoomi (2008) that the infancy of GIS grew from applications such as planning for the US census population in 1970 and national mapping agencies that began to use the technology to automate cartographic draughting. However, the imagery of earth from the satellites played a very important role in the development of GIS, particularly through the military where GIS was a platform where images could be displayed and analyzed for purposes of intelligence gathering. It was not until 1980s that the reduction in prices of computer technology created conducive environment for the development of GIS software and the subsequent growth in cost effective GIS applications (Longley *et. al*, 2001). The reduction on cost of computer hardware complimented by improved operating systems, electronic storage media and development of computer software, have had a wide and significant impact on introducing GIS technologies to new areas (Ratcliffe,2004).

2.2 Geographic Information Systems and Public Health

In brief history, it was established by Melnick and Fleming (1999) and in 1840; Robert Cowan showed the relationship between fever and overcrowding in Glasgow. He went further to attribute increased mortality to excessive immigration without any corresponding increase in housing and the steady decline of the wealthy middle class. Furthermore, in 1843 also in Cowan, Robert Perry showed the six fold difference of fever prevalence in different neighborhoods. They showed extend of the epidemic by plotting the individual affected households on a map (Melnick, 2002). In addition, in 1854, John Snow, the known “father of modern epidemiology”, plotted the geographic distribution of cholera deaths in London. He demonstrated the relation between the death and a contaminated water supply. This linked the science of epidemiology and the use of geography to review the relationship between diseases and the environment (Novick, 2008).

Notably, between 1960^{s} and 1970^{s}, Melvyn Howe was using maps to show the geographic variation all-cause and specific-cause mortality in London and Glasgow. The use of computers in geographic analysis and display began in 1960 with GIS technology developing into a multidisciplinary discipline in the1970s (Melnick, 2002). Other early applications of GIS include Open Shaw’s (1988) analysis of spatial clustering of childhood leukemia in relation to nuclear facilities in England. On the other hand, Mc Masters (1988) used GIS to assess community vulnerability to hazardous materials while Wartenberg (1993) used GIS to characterize populations living near high voltage transmission lines. By early 1990, GIS was being used to study vector borne diseases and determine the association between environmental features and vector concentration (Maantay, 2011). Furthermore, in 2002, Thailand used GIS for examining effects of different factors in public health, showing disease distribution, performing specific analysis, visualization and providing information on healthcare and also helping in different decision making. This demonstrated the importance in using spatial and statistical analysis to come up with a powerful tool for decision making.

On the other hand, in 2003, Eastern Europe International Health Organization started estimation on disease as result of water pollution by means of GIS to specify pollution resources and direction of occurring diseases (Don *et al*, 1995). The main aim was to determine system requirement for managing and taking care of diseases and factors that cause them. GIS was used to store, manage display and recognize the temporal and spatial association of diseases. Moreover in 2006, a project was accomplished by environmental conservation organization to determine how to control WN virus in Pennsylvania, USA. As such, GIS was used for gathering and combining data from different resources and creating a central geo-database to provide relation between different data centers. In this study, users could determine prevalence direction, extend of spread and number of people affected (Mesgari, 2008).

On the other hand, health and health care are considered today as the most important factors in measuring the quality of life of individuals (Melnick, 2002). The development of public health and disease management play a very important role in the social, cultural and economic development of any society. The main role of any public health organization involves environmental health, disease control and health advocacy, prevention, medical and nursing for diagnostics, control and management of diseases. Many interrelated changes such as urbanization, transportation and industrial development, unsustainable agriculture development and many others cause general and complex environmental problem that threaten mankind. Notably, it was noted by Mesgari and Masoomi (2008) that dealing with these multi aspect and complex problems can be very challenging for decision makers in public health. Therefore up to date information and models are required.

For that matter, Lang (2000) asserted that the study of Medical Geography (study of how disease and health care facilities are distributed on the face of earth) is not a new area of study. According to Mesgari and Masoomi (2008), human settlements, activities and factors causing the spread of diseases have one thing in common, that is, they spread geographically. Additionally, concentration of disease in a particular area suggests unusual presence of particular factors that cause the disease. The co-occurrence of factors usually increases the prevalence of the disease in that particular area. Such correlation, calls for the study of spatial distribution and pattern of diseases. Therefore, Geographic Information System (GIS) can be effectively used to study epidemiology.

Importantly, since health is a geographical phenomenon and various factors attributed to health diagnostics and planning are geography dependent, GIS technology for health studies comes in handy as an important tool. As such, GIS can be useful for health researchers and planners because it plays a vital role in strengthening the whole process of epidemiological surveillance, information management and analysis (Murayama &Thapa, 2011). Moreover, it serves as a common platform for convergence of multi-disease surveillance activities. GIS is being used by public health administrators and professionals, including policy makers, statisticians, regional and district medical officers, (WHO, 1999).

In addition, epidemiologic principles and methods provide the foundation for public health and preventive medicine. To avoid drawing false conclusions from maps, users of GIS technology need to understand and apply these principles and methods in formulating study questions, testing hypotheses about cause-and-effect relationships (Melnick, 2002). Furthermore, they should also critically evaluate how data quality, confounding factors and bias may influence the interpretation of results. Conversely, epidemiologists need to be able to understand and critically evaluate maps prepared using modern GIS software, data and spatial statistical methods (Richards *et al*, 1999).

According to Lang (2000), GIS has been used in many public health domains. For instance, the Center for Disease Control (CDC) which is a world premier heath organization has used GIS for many decades to study the spread of diseases from one place to another and how toxic substance affect human health. Also, the Dartmouth Medical School has used GIS to study how Americans use health care services and if the amount of heath care services that they receive is depended on where they live.

Furthermore, Lang (2000) noted that Health care planners use GIS applications to study the quality of services provided by doctors to the patients at particular sites. Health care planners can also use GIS to anticipate increase in demand for any particular service by correlating the services that different populations need and how far the services are from the population. More recently, health planners have used GIS to plan marketing drives, making sitting decisions and mapping how best to allocate healthcare facilities.

Importantly, Goodchild (1992) opined that health care managers use GIS to map new location for hospital and outpatients clinics. They also use it to tell sales people which physician are likely to test new drugs or services, provide maps to their offices or send health service to people's homes using the most efficient routes. Furthermore, according to Novick (2008), marketing and planning consultants armed with clients data and other data can use also GIS to determine the exact mix of product and services that best meet the client's needs. They can determine how much a client is spending on health care and whether they are focusing their preventative efforts in the right areas.

Conversely, Novick (2008) postulated that large health care organizations recognize the need of GIS in planning and marketing operations. They have had systems such as patient registration, but the current trend is to integrate the various systems to provide easy access to data and boost corporation between different departments. Programs like ArcGIS provide a platform for integrating heterogeneous systems and making it possible to work with data geographically. This provides hospitals with more value from their data by using it in different ways. Furthermore, Murayama and Thapa (2011) note that insurance organizations have enormous warehouses containing patient's demographic data, geographic attributes about specific diseases, interventions and treatment. Knowledge of demographic information, disease prevalence in a particular area and how they can be treated successfully can assist health care providers with information which they may then use to make decisions on how much they need to run a center successfully or where to build a new facility. Practically today, everyone needs to use GIS.

Significantly, Goodchild (1992) postulates that in the past when public's health practitioners were only concerned with communicable disease control, pushpins and dots drawn on maps proved effective in helping to analyze and control outbreaks of diseases. Population based health managers, however are responsible for analyzing complex health issues in a rapidly changing and diverse environment. To work effectively with communities in improving heath, modern health care practitioners and community partners need easy, immediate access to accurate and geographically based information. GIS is making this possible (Novick, 2008).

All the same, Melnick and Fleming (1999) in their report *GIS promise & Pitfalls*, noted that the greatest potential of GIS lies in its ability to quickly, clearly and convincingly show the results of a complex analysis using the power of visualization. However, they further go on to highlight that, "Ironically the power of the GIS tool may also be its biggest pitfall. The consequence (of integrating) complex data into a visually easy to understand picture is a setup for misunderstanding and misuse. Users, including policy makers may be tempted to infer causation from correlation and make inferences about individuals from population data. Since maps sometimes may provide selective and an incomplete picture of the reality, health care personnel should be aware and careful when making some conclusions that may not be supported by epidemiologic analysis.

Notably, the use of GIS application in public health has the following advantages and disadvantages according to Melnick and Fleming in TB Richards's report. In terms of its advantages, it provides practitioners, researchers and planners with abilities to organize and link different datasets. The linkages assist health practitioners with ability to plan cost effective interventions. Additionally, GIS technology encourages formation of data partnerships and data sharing at community level. Equally important, as new GIS methods are developed, they can be added to the "toolkits" of epidemiology and health services research. Lastly, comparisons from tables, charts and maps developed using GIS technology can be an extremely effective tool to help community decision makers visualize and understand public health problems.

On the other hand, their disadvantages include the fact that community health planning and other public health applications remain a relatively underdeveloped marketplace niche for GIS technology. Moreover, there are no current, accurate, low-cost, based street maps that are essential for epidemiologic uses. Without an up-to-date based street map, for example, a public health practitioner investigating a disease outbreak may have to spend considerable extra time and effort to digitize the locations of cases or may not be able to map all case reports. Also, practitioners, planners, and researchers, and especially state and local public health department staff, need training and user support in GIS technology, data and epidemiologic methods in order to use GIS technology appropriately and effectively. Additionally, statistical and epidemiological methods need to be developed to protect individuals and household confidentiality. Lastly, the technology to prepare and display maps on the Web is still in the very early stages of development.

However, GIS has become an empowering technology providing greater awareness and responses to localized public health issues, such as spatial trends, predicting risk and exposure assessment (Croner, 1996). In respect to this, in Healthy People (2010), GIS is identified as a powerful tool for combining geography (e.g., locations), data (e.g., health, census, or environment) and computer mapping with data linkage and analysis to identify targeted geographic areas for disease surveillance and intervention. Use of geo-coding is particularly encouraged for such usages.

Since 1990, there has been a shift from traditional social medicine to "new" public health. There has been a growing interest in heath needs assessment at small scale requiring the use of census and other data to identify what services are required, where they are needed and how heath care facilities are delivered. GIS has been used in a number of ways: some use has been in boundary of localities Census based deprivation scales that have also been incorporated in health database in order to identify areas of need and in indicating vivid picture of health care facilities inequalities in a small area (Anthony & Markku, 2003). Furthermore, the representation and analysis of maps in incidence of disease data are basic tools in the analysis of regional variation in public health. However, in recent years, there has been a development on new methods of mapping diseases. This has lead to greater use of geographical or spatial statistical tools in analysis of both routine data collected for public heath purposes and in analysis of data found ecological studies of diseases relating to explanatory variables (Kandwa, 2009).

Importantly, according to Melnick and Fleming (1999), the study of geographical distribution of diseases can be classified in to three main categories: 1) *Disease mapping* – usually the object of this analysis is to provide (estimate) the true relative risk of a disease of interest across a geographical area. Application of such methods lies in health service resource allocation. 2) *Disease clustering* – This aids in public health surveillance, to decide where it may be important to assess whether a disease map is clustered and where the clusters are located. The analysis of disease incidence around a putative source of hazard is a special case of cluster detection. 3) *Ecological analysis* – this focuses on the analysis of the geographical distribution of disease in relation to explanatory covariates, usually at an aggregated spatial level (Don *et al*, 1995).

2.3 Spatial Statistical Analysis and Geographic Visualization

Although statistics may be hard to understand and therefore unattractive to decision makers, they are a powerful means of describing large datasets that would otherwise remain unintelligible. Statistics may be used to predict and fill values where holes in data may occur (Verjee, 2008). The combination of visual and numerical techniques for summarizing data properties, detecting patterns and identifying features is very essential to decision makers in public health.

2.3.1 Spatial Analysis

Spatial analysis describes 'where' things are in relationship to other environmental factors that may influence or control the distribution pattern of a phenomenon of interest (Meuhrcke, 1996). To do so, health data and related factors must possess geographic location information such as longitude and latitude co-ordinates, a valid street address, census tract locator or ZIP code (Alexander, 2003). In the narrow sense, spatial analysis has been considered as a method of analyzing spatial data, while in the actual sense it includes revealing and clarifying processes and structures of spatial phenomena on earth's surface (Murayama & Thapa, 2011).

Nonetheless, according to Fotheringham and Rogerson (2009), spatial analysis is a technique that uses locational information in order to understand the processes generating the observed attribute value. Spatial data and the processes that generate these data have several properties that distinguish them from none spatial data that is: they are not independent of each other and similar attributes tend to cluster and are different from data drawn from far away. On the other hand, spatial dependency is described as the degree to which a value of a variable at one location is influenced by neighboring locations (Chainey & Ratcliffe, 2005). This form the basis of Toblers first law of geography which states that everything is related to everything else, but nearer things are more related than distant things (Fotheringham & Rogerson, 2009).

According to Goodchild (1992), spatial data analysis is a set of techniques that are devised to support spatial perspective on data. He further notes that to distinguish it from other analysis techniques, it can be defined implicitly as a set of techniques whose results are dependent on the location of the objects or the events that are being analyzed and further, that it requires access to both location and attributes of the objects (Don *et al*, 1995).

Conversely, the geographic phenomena of spatial association and spatial dependency of geographic objects is referred to as spatial autocorrelation. The objects in natural systems have some degree of spatial clustering. The auto correlation measures how similar objects vary at a location and the surrounding (Murayama & Thapa, 2011). Nonetheless, spatial dependency not only suggests clustering of values but also the possibility of dispersion. When high values are correlated to high levels in neighboring areas, or when low levels are clustered together partially, the series is said to display positive spatial auto correlation (Chainey & Ratcliffe, 2005).

Interestingly, when high values are associated with low levels in neighbors or low areas are surrounded by high values in neighbors, the auto correlation is said to be negative. Although spatial statistics is based on the assumption that the process under study is stationary, Fortin and Dale (2005) assert that spatial patterns in a given area are a synthesis of a dynamic process operating at various spatial and temporal scales. Spatial phenomena that occur on the earth's surface change over time and if we trace the movement of the phenomena to the past, we are able to acquire clues that can be helpful to unravel today's structure and formation mechanisms (Murayama & Thapa, 2011), although there are many different techniques of reducing data.

Fotheringham and Rogerson (2009), classify spatial analysis into four broad categories. The first category is that spatial analytical techniques are aimed at reducing large data sets to smaller amount or more meaningful information. The second category is the exploratory data analysis technique which consists of methods to explore data and model outputs in order to suggest a hypothesis. Third, is the technique that examines the role of randomness in observed spatial patterns of data and testing hypothesis about those patterns. Last, is the mathematical, modeling and prediction of spatial processes. As such, spatial modeling coupled with adequate data and rapid increase in computing power would lead the society in solving many pressing issues in urban and regional areas (Fotheringham & Rogerson, 2009).

Notably, spatial statistics can further be classified as first order statistics that are aggregation, based on the abundance of data. They detect trends in data over the entire study area (mean) and indicate whether there are spatial patterns or not (Fortin & Dale, 2005). In this first order statistics, measuring geographic distribution can reveal the center and orientation of spatial features, determining the mean, median and mode for a set of points produced in centers that are easy to understand. The mean center or centroid analysis is very helpful in identifying trends of various phenomena over time. Another method of *centroid analysis* is Minimum Aggregation Travel (MAT). This method performs centroid analysis based upon iteractive examination of infinite sets of points in continuous space (Verjee, 2008).

The second order statistics is to measure the local spatial patterns (intensity) by computing the deviations of values at neighboring locations from the area of study (Fortin & Dale, 2005). Pattern analysis is an example of this category. *Pattern analysis* is an important form of inferential statistics. Knowing if points are random, clustered or dispersed can help the decision maker to understand point to point effects amongst datasets. For example, cluster analysis may reveal casual relationships between water point and an outbreak of water borne disease as the case of John Snow in 1849, in Soho. However, the ultimate purpose of spatial statistics is to understand the relationship between various geographic phenomena. *Relationship analysis* can help explain why things are occurring where they do, predict where something might occur in future, and indicate how closely certain variables are correlated (Verjee, 2008). In summing up, simple spatial statistics or first order statistics such as mean distance, mode and median may be sufficient; however in other applications second order statistics may be applied.

2.3.2 Geo Visualization

Although visualization is not purely GIS based, it has become an increasingly popular application of GIS technologies. Since the release of Google Earth in 2004, a lot of unprecedented interest has been created in geovisualization especially among none technical audiences interested in understanding places in 3D and high resolution (Verjee, 2008). The purpose of geovisualization is to reveal knowledge in data which is not detectable by current computational methods, but which can easily be identified by the visual system. The value of visualization is the fact that it can force us to notice something in the data that we never expected to see. Visual data exploration implies generation of new ideals through creation, inspection and interpretation of visual representation (Fotheringham & Rogerson, 2009).

In respect to this, the layman now has for the very first time the capacity to create, quickly view and publish vector data without much training and almost at no cost. According to John Hanke creator of Google Earth, “ten years ago the technology was exclusive to US military and single satellite image costed 14,000 USD, but now there is global high resolution imagery” (Verjee, 2008). The sheer volume and complexity of geospatial data makes it impossible to rely solely on human vision for knowledge discovery. To take full advantage, visual explorations are combined with computational exploration methods to come up with visually enabled knowledge discovery methods that can facilitate automatic processing of patterns and relation recognition in complex data and subsequent interpolation of discovered patterns and relationships (Fotheringham & Roberson, 2009). According to Jack Dangermond, founder of GIS software firm ESRI, GIS packages are making people more spatially literate, allowing them appreciate and visualize their local situations. He goes ahead to state that though GIS is functionally very limited, application like Google Earth and other rival applications play a contemporary role to future vendor of full scale GIS solutions (Verjee, 2008).

2.4 Related Work

In order to develop a good and reliable application model, we need to look at previous work done by other people in relevant discipline to get benefit and experience from their work and also to have an understanding of the opportunities and possible constraints that might face the current project. For that case, the following are some of the studies that have been undertaken in the area of spatial analysis and visualization in the health sector.

2.4.1 Spatial Analysis of Tuberculosis in an Urban West African setting

This paper investigated a case of TB clustering, using Geographic Information System (GIS) together with Spatial Scan Statistics (SatScan). Demography (age, sex, ethnicity, occupation, place of residence), clinical information (type of TB, date of diagnosis, and date of registration) and GPS coordinates for the residence of participants were recorded. The study was concerned with both the epidemiology of TB and distribution of diagnostic and treatment centers. The study confirmed significant TB clustering in Greater Banjul. The current study was concerned with studying the impact of spatial analysis on health domain with an aim of coming up with GIS solutions to assist stakeholder and policy maker with information that they can be able to use to make timely and accurate decisions.

2.4.2 GIS Applications in Public Health as a Decision Making Support System and It's Limitation in Iran

This section discusses two studies. In the first study GIS was used to investigate evidence of clusters in four types of cancer. Different methods of geo-statistical were tested and mapping was chosen as the most suitable method of detecting and monitoring spatial clusters of these types of cancer. Probability mapping was used to analyze and visualizing statistics of the deaths caused by lung cancer, stomach cancer, leukemia and skin cancer in 18 provinces of Iran. In the second study, GIS was used to come up with a geo-statistical model to look at the correlation of occurrence of skin cancer and parameter-factors that were causing effect. The spatial distribution of the disease and its parameter was the most important aspect in the model development; this justifies the use of geo-statistical modeling approach and application of GIS in all stages of data manipulation and analysis.

The study showed the following GIS limitations: First, the role of location in data gathering is not recognized in most governmental organization. The data is collected without recording the location at all or by just mentioning the name of the place. Second, the variety of data needed is derived from different organization and often, there is lack of cooperation between the organizations or lack of access to the data. The study therefore concluded that, although GIS technology is very valuable in public health, more effort is needed for proper usage of GIS and spatial data in management of public health and disease control (Mesgari, 2008). The current study is on analysis of TB and has employed the same tool used in this study to provide information for stakeholders to make decision in the public health sector in Kitui County.

2.4.3 Geographic Information System Applied to a Malaria Field Study in Western Kenya

This paper describes the use of global positioning system (GPS) in differential mode (DGPS) to obtain highly accurate longitudes, latitudes, and altitudes of 15 villages in Western Kenya. Although Simple GIS applications are available to simplify the effort of spatial analysis and epidemiology study, however accurate base maps are required. This is one of the major obstacles for researchers pursuing geographic analysis of tropical diseases (Hightower *et al*, 1998). In the study, it is identified that the sketch maps available for this operation where very inaccurate and lacked coordinates system which is necessary for spatial analysis. The best maps were identified as satellite maps and remote sensed data. The study used a technology known as Differential Geographic Positioning Systems (DGPS) that was to obtain the geographic positions. Atlas GIS and Statistical Analysis Software (SAS) were used for all spatial analysis. Position information was linked to patient databases and integrated to the analysis tools to produce layers of information. Entomologic and parasitological data was used to demonstrate simple GIS analysis. It does not acknowledge the powerful features of the current GIS system. More emphasis in this study was on the SAS (Hightower *et al*, 1998).

Although the study is similar to the current study, the GIS application was mainly used for geographic positioning. SAS was used for analysis. The current study used the current GIS system which was extensively used in recording the data, geo-coding, visualization and statistical analysis.

2.4.4 Spatial Modeling of HIV Prevalence in Kenya

A clear understanding of geographic distribution of HIV infected people, maintenance of an up-to-date list and location of facilities providing HIV related facilities is very important in monitoring the spread, treatment, care and provision of heath care facilities (Montana *et al*, 2007). In this paper, spatial modeling of HIV prevalence in Kenya in relation to its spatial and behavioral determinants was examined. Data from Kenya Demographic and Health Survey 2009 was integrated into ERSI ArcMap 9.0.

The study came up with a model map of HIV prevalence in Kenya and related it to spatial and behavioral determinants. The survey linked individual HIV test results for male and female above 15 years of age with behavioral, social and demographic indicators. The indicators were used to develop a model to predict HIV prevalence and map its concentration at sub provincial level. The study reviewed that areas with high concentration of HIV prevalence had low density of HIV related services. It also revealed a sharp difference in HIV prevalence and associated risk between men and women in the same age bracket.

The current study used similar data collection methods and the same tools for visualization and analysis with similar expected outputs. The difference is that, in the current study, it was concerned with TB which is an air borne disease and which does not depend on the behaviors of human beings.

CHAPTER THREE

METHODOLOGY

3.0 Introduction

This chapter comprises two parts: the first part describes the research method and tools used, while the second illustrates the project methodology that was used in the implementation of the project prototype.

3.1 Research Methodology

This part describes the research design, study area, population, samples, sampling technique, data collection, tools used and data analysis techniques.

3.1.1 Study Area and Target Population

The study was conducted in Kitui County, one of the 13 counties in Eastern Kenya. The county is located between Longitudes 37º45´ and 39º0´ east and Latitudes 0º3.7´ and 3º0´ south, covering approximately 20,402km². It is divided into six constituencies as per the Interim Independent Boundaries Review Commission of Kenya (IIBRC), established by an Act of Parliament in 2009. It has a population of approximately 1,012,709 as per the 2009 Population and Housing Census. Around 92 health facilities are owned by government, private investor and the church (mission hospitals). It has 55 TB diagnostic centers and 67 treatment centers (all diagnostics centers are also treatment centers). The county has 5 TB management units namely: Kyuso District Hospital, Mwingi District Hospital, Kauwi Sub-District Hospital, Kitui District Hospital and Mutomo Sub-District Hospital. The County has a total of 138 Health facilities. The average distance to the nearest facility is 5 Km. The target population was about 5000 TB patients from January 2010 to December 2012. This target population is the aggregate of the cases that conform to some designated sets of specification (Nachmias & Nachmias, 2004).

3.1.2 Sampling Design

All TB cases that had been reported in all the 5 TBMU centers between January 2010 and December 2012 were included in this study. The total sample size was 500, with an average of 166 cases chosen from each year. This consists ten percent of the population. Nachmias and Nachmias (2004) insist that ten percent or 200 elements must be included in a sample. Further, simple random sampling method was used to choose cases from all TB diagnostics and testing centers for each year for the sake of accurate data and representative coverage of the whole population.

3.1.3 Data Collection

Data was collected randomly from patients between January2010and December 2012. Any TB case that had been reported was included in the study. Data mainly comprised secondary data obtained from registers at the TBMU centers with the authority of the concerned stakeholders and as per the policy of the Ministry of Public health & Sanitation. Primary data was also collected from participants by conducting both face to face interviews and telephony. In order to ensure strict confidentiality, personal identifiers were not recorded and participants had the option to refuse any question or to terminate the interview at any time. Data obtained comprised of demographics data (age, sex, ethnicity, physical address, and health facility), clinical information (type of TB, date of diagnosis, date of registration and HIV status) and mobile phone numbers.

The geographical locations of TB testing and treatment centers and the residential addresses of the participants were geo-decoded using GIS applications (R-geo-coding), Google Earth and other online geo-coding tool. Google Earth was used to visualize and coordinate to verify that its geo reference was correct.

The treatment outcome was categorized using the following parameters:-completed treatment, cured, defaulted failed treatment, transferred out and dead. The residential status of participants was also categorized as either permanent or temporary.

3.1.4 Geo-Visualization Techniques

Maps of Kitui County were downloaded from Google Maps using the Google mapmaker and integrated into ArcGIS and R-GIS. Data was captured in Excel data sheet for cleaning and exported to ArcGIS/ R-GIS attributes table for mapping, visualization and analysis. Dot maps were used to show the spatial density of the TB cases. Aggregation of data at the census blocks group or tract level choropleth maps were used to show spatial variation with distinctive grayscale values and bubble used to show different weights.

3.1.5 Spatial Analysis Technique

Population density and social economic status were mapped for census tract using ArcGIS/R-GIS. The locations in each of the major towns with the greatest population density were marked to create a central reference point. The relationship between location, population density and social economic status and health facilities were visualized to identify hot spots, investigate clustering and identify treads and gaps. The study between the periods 2010 to 2012 was divided into three and the location of TB cases in these periods compared. The measure of the geographical spread of TB cases from the town centers was determined by measuring the distance of each TB case to the nearest town center. Statistical significance was determined by appropriate scale proportionate to the density.

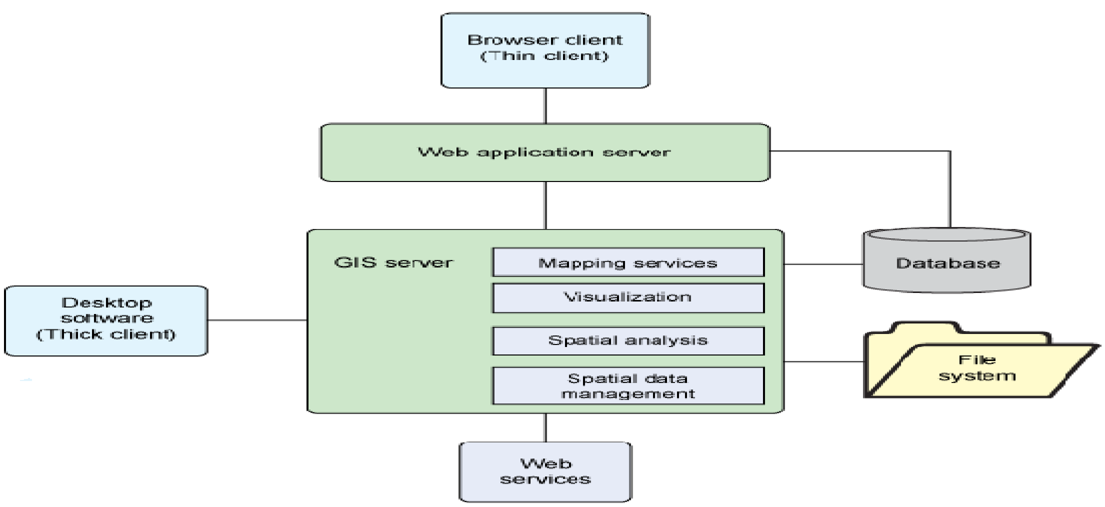
3.2 System Development Methodology

The system development methodology is a formalized approach to carry out a system. It embodies set of procedures, methods, best practices, automated tools and guidelines used to develop and maintain an application (Booch, 1986).

3.2.1 Development Methodology

This project is about creating a GIS prototype application for public health organizations to support its decision making processes. The model was implemented as in figure 3.1 below.

Figure 3.1: Overview of the Project Prototype

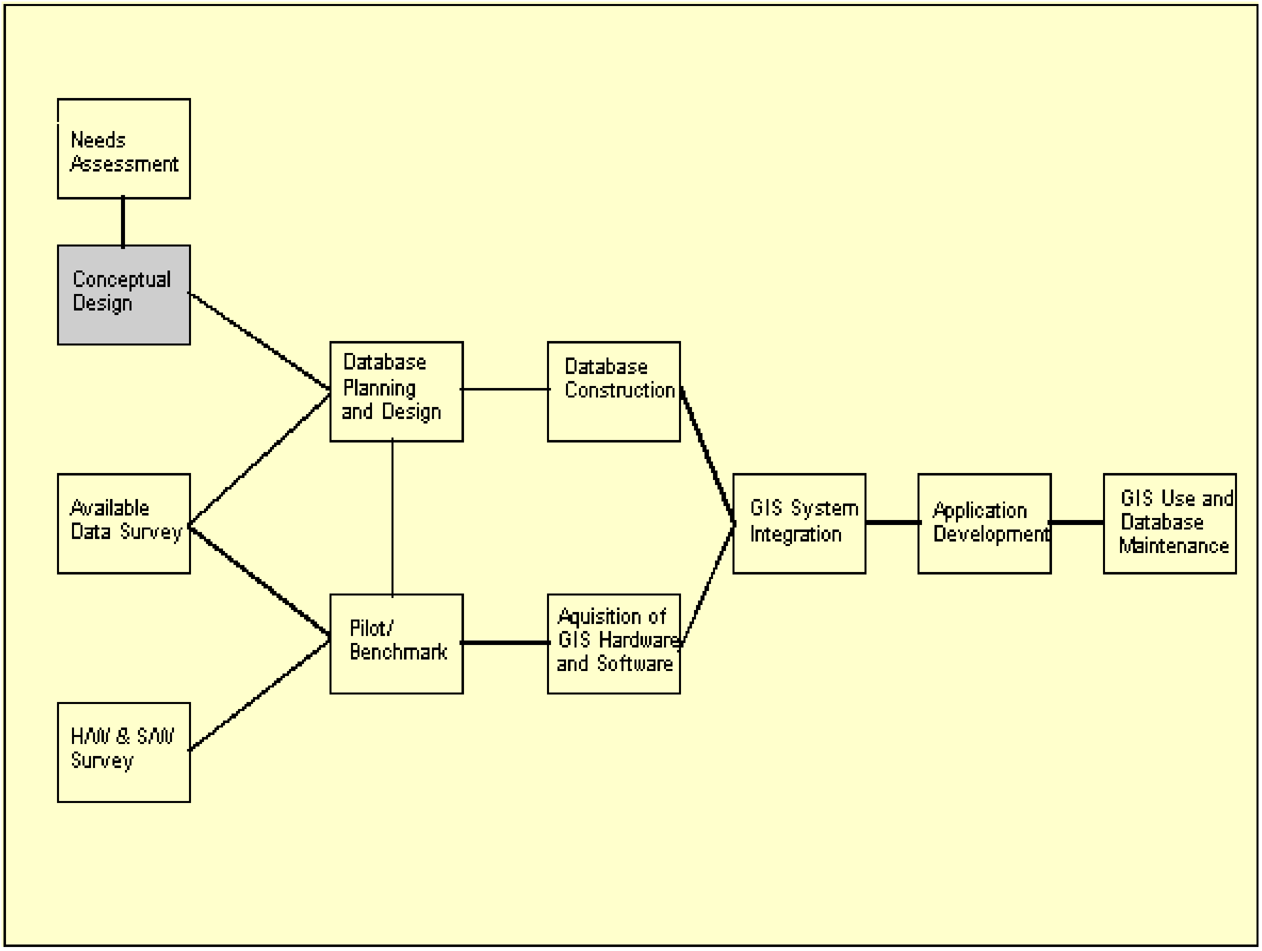


Source: Crowther S., (2008)

3.2.2 GIS Development Life Cycle

To develop the solution a GIS development life cycle was used to implement this project. The development life cycle consists of a set of eleven steps starting with the needs assessment and ending with the on-going use and maintenance of GIS system as illustrated in figure 3.2 below.

Figure 3.2: Conceptual Design of the GIS System

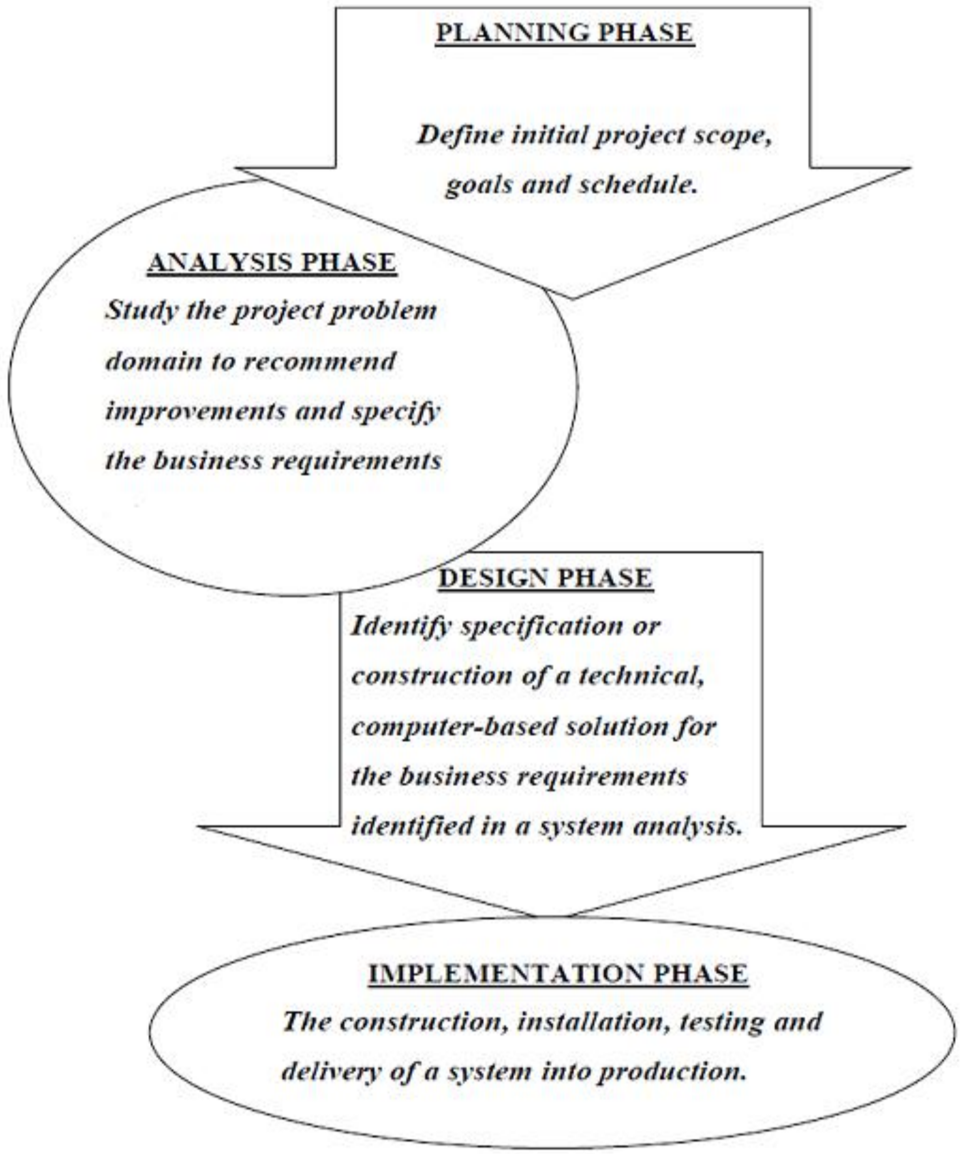


Source: [http://loi.sscc.ru/gis/sara/tblconv1.htm]

3.2.3 Development Cycle Phases

In tackling the problem as per one of the objectives of chapter one, which is prototype development, the development went through the following phases: planning, analysis & design, implementation which was repeated until the final prototype evolved. Figure 3.3 demonstrates project phases and identifies all the activities that were involved in each phase.

Figure 3.3: Project Phases



Source: Author, 2013

3.2.3.1 *Phase 1:* Planning:

This phase identifies the questions and problems that the proposal aims to resolve according to its set objectives. Consequently it define the scope of the project, the potential users of the developed system and their interactions as defined in chapter one of this proposal. It also determines the methodology that was used to accomplish this project.

3.2.3.2 *Phase 2:* Analysis:

This is one of the proposals specific objectives, which is studying the relevant literature reviews that has been handled with the purpose of identifying the similarities and possible limitations and constrains to take advantage of previous studies. Studying how GIS has been used in the study of diseases spread, it helps in understanding how they are valuable and effective in enhancing the process of making decisions regarding the distribution of services. This has been accomplished in Chapter two of this project.

3.2.3.3 *Phase 3:* Design:

This is the empirical part of the project. The design phase carries out the practical activities of the project that guide the development process. It handles the processes of identifying the system's requirements, developing the conceptual design of the project and database planning-designing-and construction.

In this phase, a conceptual design was done. This is considered as the first step in database design where the contents of the intended database is identified and described. It identifies the data content and described data at an abstract, or conceptual, level. This step is intended to describe what the GIS must do. Building GIS database was divided into two major activities:

- Creation of digital files from maps and;
- Organization of the digital files into a GIS database.

3.2.3.4 *Phase 4:* Prototype Implementation:

In this phase a quick prototype was developed. This allowed users to interact with the prototype at early stages to identify the needed changes or additional requirements. The prototype was created using the functionality provided by GIS software's such as ArcGIS, and oracle database.

The expected activities and deliverables can be summarized as is shown in the table 3.1 below.

Table 3.1: Activities and Deliverables

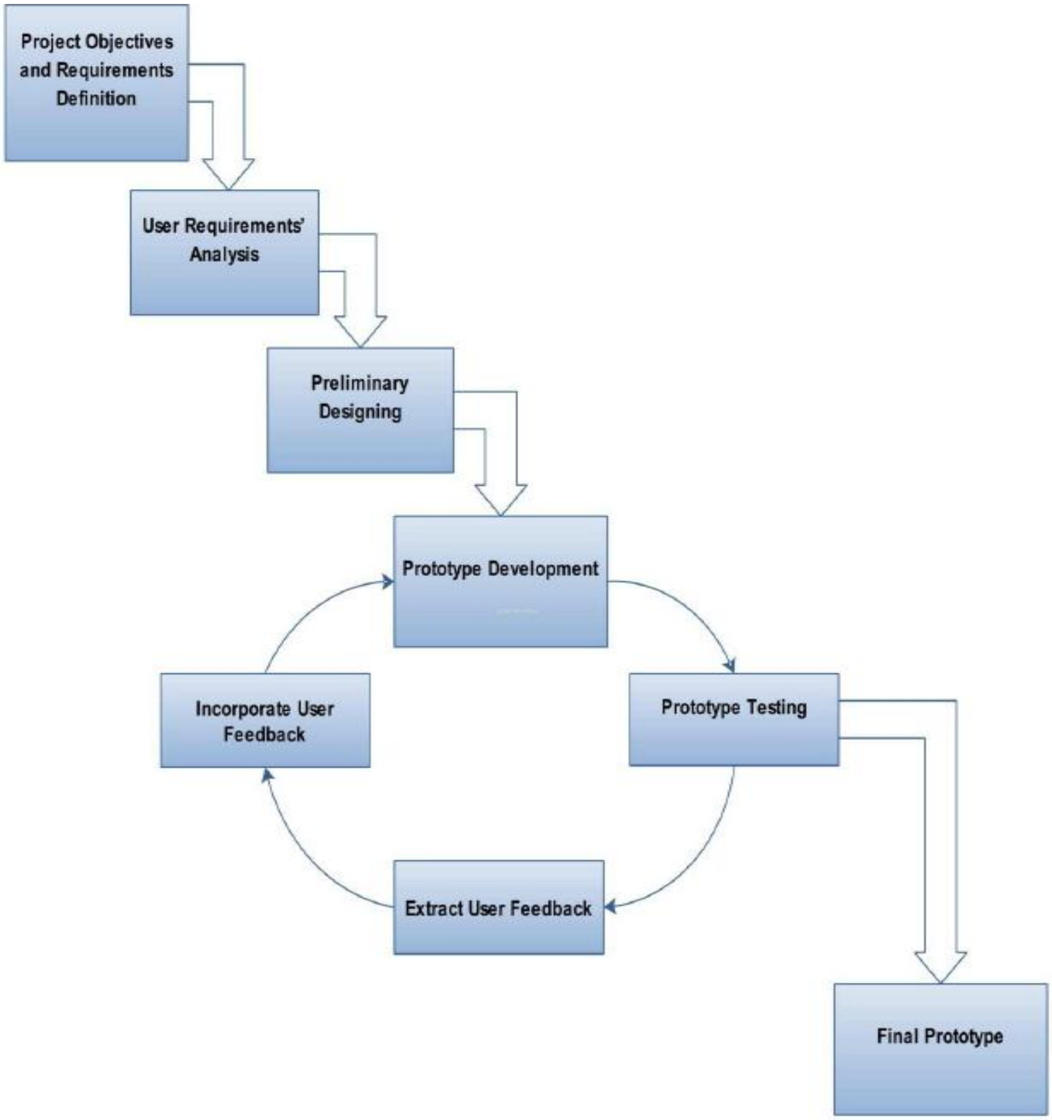
Phase	**Activities**	**Deliverables**
Planning	• Identify project objectives. • Define the project scope. • Specify project methodology.	• Project objectives list. • Scope statement. • Project methodology.
Analysis	• Study relevant literatures. • Needs assessments. • Analyze user requirements. • Gathering requirements	• Identify common features & functions, address possible limitations & constraints. • Spatial data needed in the GIS DB. • GIS functions required. • Functional & non-functional requirements. • Identifying the suitable gathering technique.
Design	• Identify system requirements. • Conceptual design. • DB planning & design. • DB construction. • Program Design.	• System models & GIS data model. • Conceptual model. • Database design (class diagram). • Logical /physical design. • Digital maps and files organized into GIS database. • Architecture Design.
Implementation	• GIS Prototype development. • Prototype evaluation. • Installation.	• System prototype. • Advantages and deficiencies of developed prototype. • Testing plan. • Conversion strategy.

3.3.1 Project Framework

Prototyping together with waterfall model was used to help demonstrate the technical feasibility of the proposal. Further, it helped to better understand all user requirements and limit the cost by carefully understanding the problem before committing all resource.

The combined model has been described in the project framework illustrated below in figure 3.4.

Figure 3.4: Project Framework



Source: Author, 2009

CHAPTER FOUR

SYSTEM DESIGN

4.0 Introduction

This chapter illustrates the analysis of the study that was performed to answer the questions who are the users, the function of the system, where and when it will be used. The analysis leads to design the proposed project. It points out the requirements gathering techniques and the functional models which describe business processes and the interaction of an information system with the environment in both current system and the to-be system.

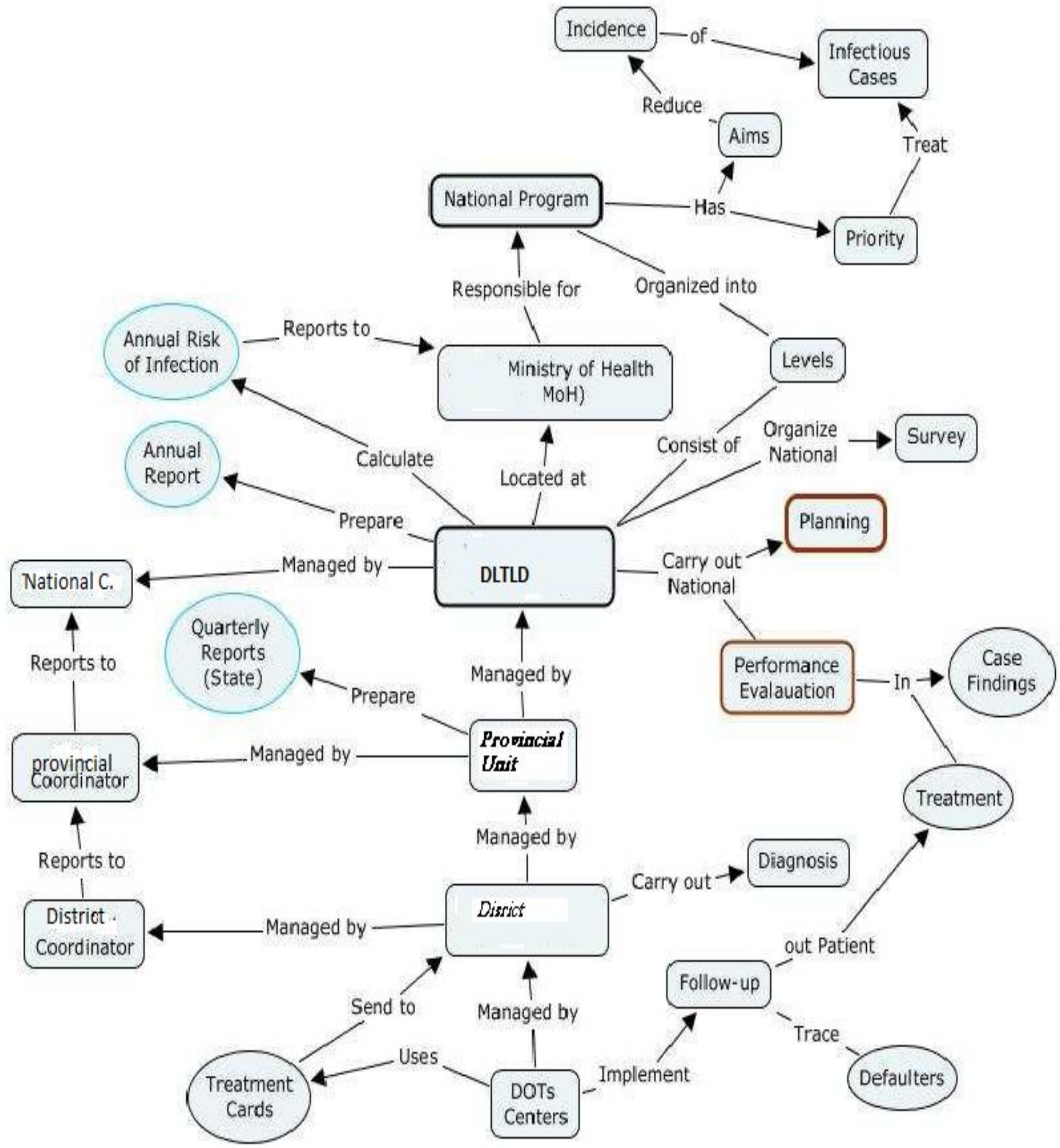
4.1 Organizational Analysis

To understand the Tuberculosis diagnostic and treatment program, its management structure, assessment and evaluation of the performance of services provided, learning of the current system, its problems, functional and non functional requirement of the staff involved, several interviews were conducted with the all county TB program coordinators to elicit and validate an understanding of the problem domain and requirements of the system, then create conceptual maps to organize this information and analyze it. Also document analysis was done to bring awareness of what kind of data, information, or reports organization needs to have.

4.1.1 Understanding the Organization

The main objective of the National Tuberculosis Program in Kenya is to ultimately eradicate the disease. The means to this end is in controlling its spread by effectively identifying and treating as many infectious cases as possible. Provision tuberculosis services are integrated into the general health service at the district level. However, special staffs of the DLTLD are responsible for coordination, supervision and technical advice in relation to management of TB at all levels. In 2010, a total of 168 District Tuberculosis/Leprosy Coordinators (DTLCs) were responsible for coordinating the delivery of TB and Leprosy services. These officers were supported by 12 Provincial Tuberculosis/Leprosy Coordinators (PTLCs). Twenty two technical officers were available at the central unit of the DLTLD to provide technical guidance for the national response to TB, Leprosy and Lung disease control.

Figure 4.1: Concept Map of National TB Program (NTP)



Source: Author, 2009

4.1.2 Management Structure

Management structure is divided into three units; the first is the central unit followed by the TB management unit and finally the treatment center where patients get their treatment.

1. DLTLD -Central Unit (CU):

This is the administrator of the Tuberculosis program; it is located in Ministry of Public Health and Sanitation in the Department of Disease Prevention and Control. Its job is planning; keeping track of the distribution of centers in the whole country and supplying these centers with drugs.

TB program covers 2,818 health units managed mainly by the Ministry of Public Health and Sanitation, Ministry of Medical Services Health (and other Ministries), NGO/FBO health units and some private institutions. Smear microscopy services were available at 1, 335 of these health units. See tables below:

Table 4.1: Provision of TB Treatment Services in 2010

	GOK	**NGO**	**Private**	**Total**
Hospital	**199**	**105**	**82**	**386**
Health Center	794	368	60	1,222
Dispensary	915	139	37	1,091
Other	8	20	53	81
Total	**1,704(60.8%)**	**382(13.5%)**	**232(8.2%)**	**2,818**

Table 4.2: Provision of TB Diagnostic Services as per 2010

	GOK	**NGO**	**Private**	**Total**
Lab	**679**	**249**	**119**	**1,047**
AFB	7853(63.8%)	241(18%)	241(18%)	1,335

2. TB management unit (TBMU) - DTLCs)

It can be any health facility, hospital or any health center. Every center must serve 100,000 persons. Every TBMU has laboratory technician who examines the patients and a District Registrar who registers all data about the patients in quarterly reports. From recording and reporting on operations to the kind of drugs needed and distribution of disease in any area specified.

3. Direct observation centers (DOTs)

It can be any health centers. TB programs follow the direct treatment of patients. There are many criteria for dots centers:

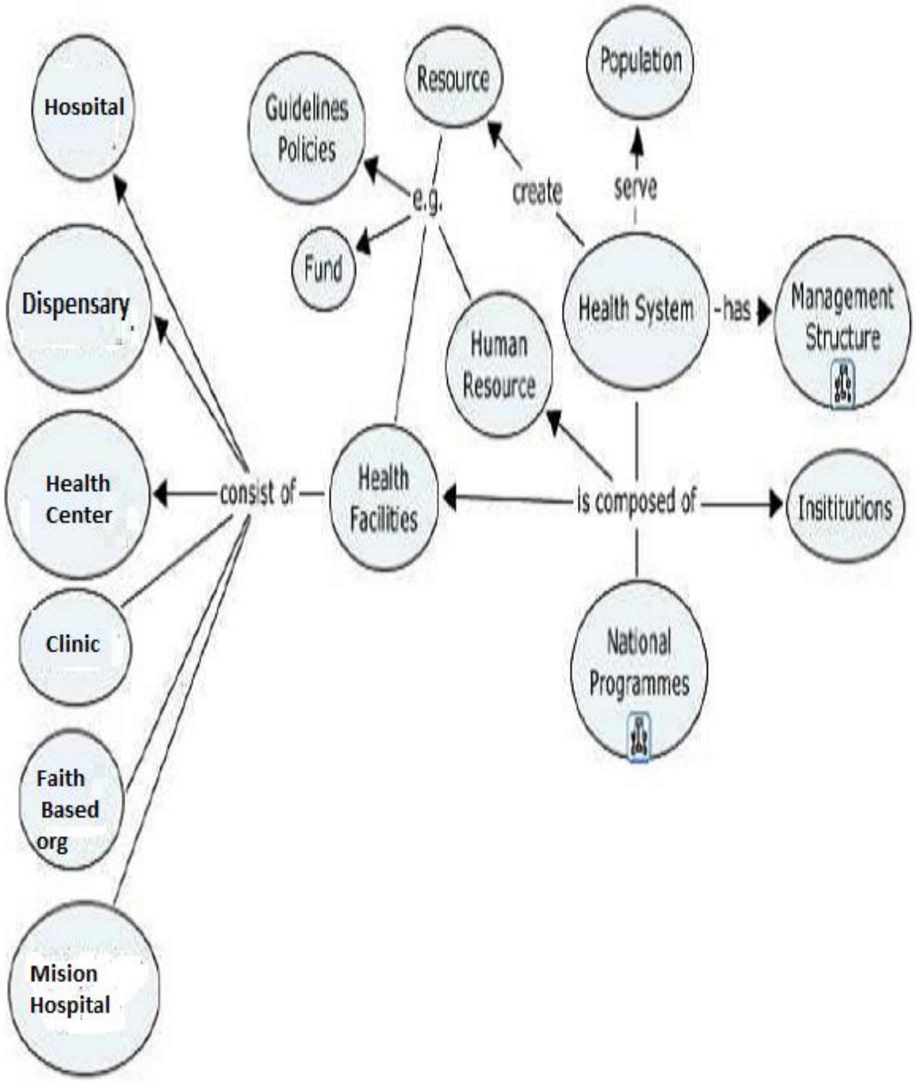
- The TBMU may be far from the patients residence;
- The patient must not walk for more than half an hour;
- The patient must not take the drug by himself.

These DOTs centers must serve 33,000 persons and with each TBMU there must be at least 3 DOTs centers. To know the distribution of disease, the TB program collects data about the number of patient and collection is done by making survey from data collection to the rate of detection cases which are then calculated. The survey can achieve in two ways:

I. Passive detection, the patient comes to the center to test for disease.

II. Active detection.

Figure 4.2: Health Systems in Kenya



Source: Author, 2009

4.1.3 Understanding Tuberculosis

Tuberculosis is second to HIV as a leading killer disease of the world population taking with it nearly two million people annually. Nearly a third of the world population is infected with the disease WHO (2011). TB is the leading killer of people living with HIV causing one quarter of all deaths. Most TB patients are never diagnosed with the active disease, because the human immune system keeps it at bay. However, weakened by other diseases such as HIV, poverty, malnutrition and aging, those who have been exposed may develop active TB. Over 95% of TB deaths occur in low-and middle-income countries, where 75% of these cases are in the most 50 economically productive age (15-54 years). TB manifests itself in two types:

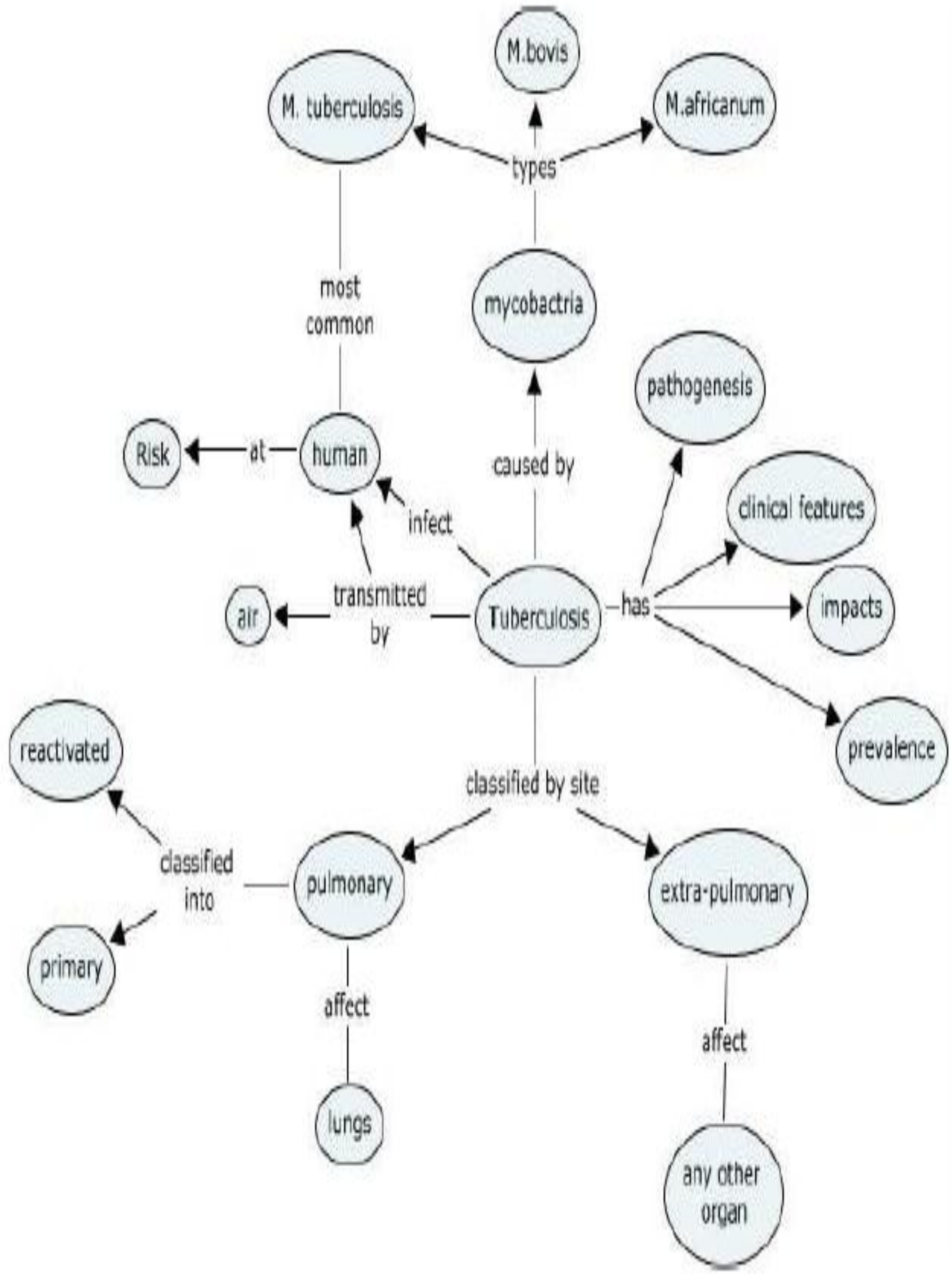
1. Pulmonary tuberculosis (PT) and;

2. Extra pulmonary tuberculosis (EP).

When the diagnoses are made it splits into two types:

1. ***Smear positive**:* The bacterium appears in the microscope.
2. ***Smear negative**:* The bacterium does not appear in the microscope.

Figure 4.3: Mapping of TB infection and Disease



Source: Author, 2009

4.1.4 TB Treatment and Its Results

TB is a treatable and curable disease. Active, drug-sensitive TB disease is treated with a standard six-month course of four antimicrobial drugs that are provided with information, supervision and support to the patient by a health worker or trained volunteer. Other categories include the (remaining type) which is a little difficult to treat. Its treatment is administered like any other but the duration varies to eight months.

4.1.5 Kenya's National Division of Leprosy, TB & Lung Disease (NDLTLD)

The Government of Kenya launched the National Leprosy and Tuberculosis Program (DLTLD) in 1980 combining the then existing tuberculosis control activities, which had been in place since 1956. As at July 2007, the National Leprosy and Tuberculosis program (NLTP) was elevated to Division of Leprosy, Tuberculosis and Lung disease (DLTLD), within the Ministry of Public Health and Sanitation in the Department of Disease Prevention and Control. Kenya's National Division of Leprosy, TB & Lung Disease (DLTLD) began to implement the WHO-recommended DOTS (Direct observed Therapy) the internationally recommended strategy for TB control strategy in 1993 with two major objectives:

1. Detect 70% of the exiting cases;
2. Achieve 85% target for treatment success rate as recommended by WHO.

For this to be achievable they need to properly manage their centers, easily monitor and evaluate their performance to understand the deficiencies and deduce mechanisms to improve the performance and control of the disease.

Quarterly reports from the District Tuberculosis/Leprosy Coordinators (DTLCs) are not enough for one to understand the current situation or make a strong decision regarding the policies of distribution of services in order to cover as large areas as possible in addition to the data in table form which makes it difficult to visualize the exact situation.

There is need to enhance their performance by decreasing the number of defaulting patients and preventing them from becoming drug resistant which is considered a big treatment problem. One of the reasons that might inhibit patients from completing their treatment can be access to direct observation treatment (i.e. patients have to come to health centers to take their dose of drugs). Geographically, there is restriction criteria regarding the distribution of services such as, the patients must not walk more than half an hour to reach the DOT center. This implies that the health center should be located near the patients' addresses, all of these cannot be monitored through tabular reports, These factors among others corroborate the need for a geographical information system that can support decision makers in their job by presenting crucial information in visual form which then show what has been and should be done and where it should be done.

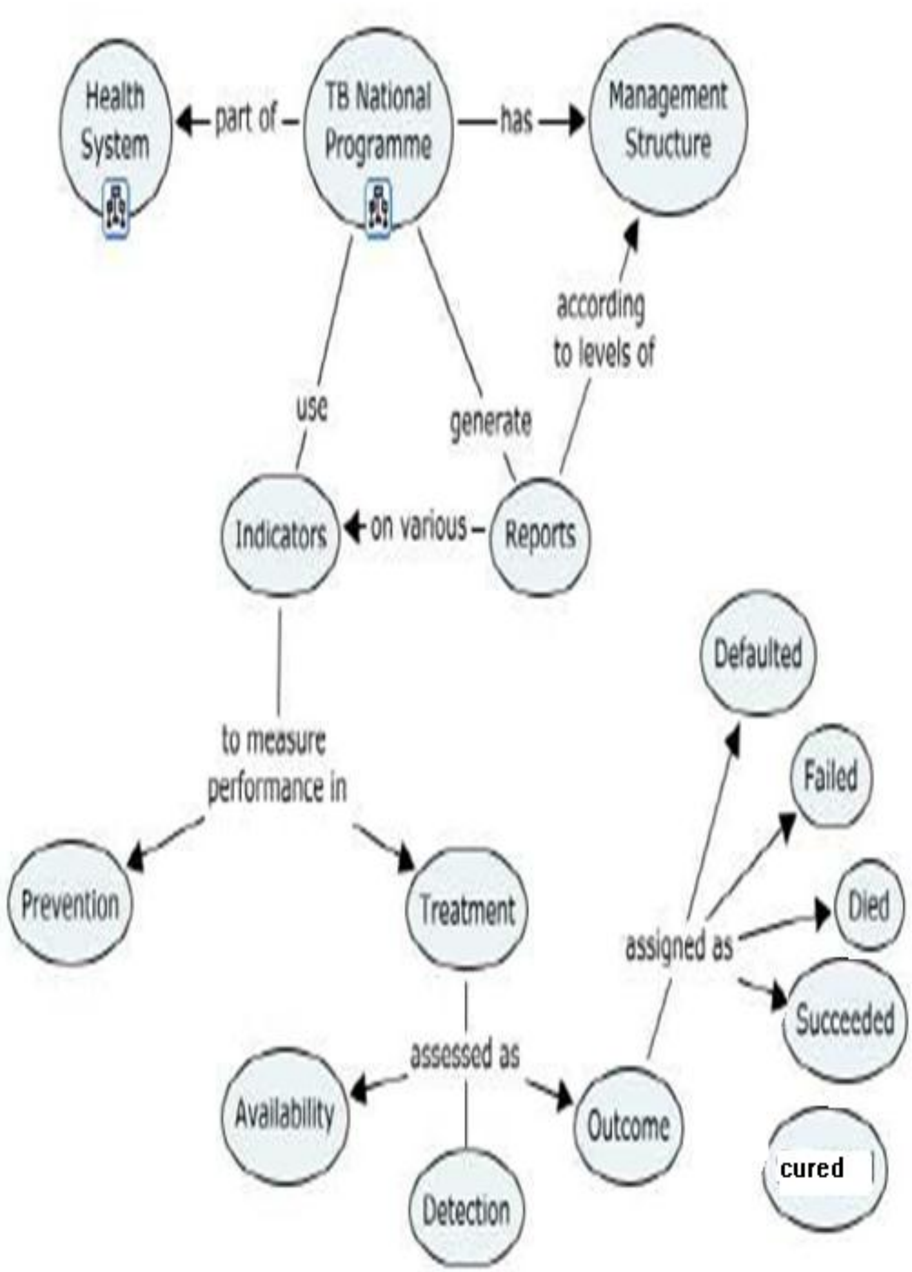
4.1.6 Current Business and Data Models

The central unit keeps track of all the centers' reports. These reports are sent after being generated in excel sheets in each TBMU every three months. The reports consists two main types. The first one is a case finding report, which determines the detection rate as compared to estimated rates. The second is the conversion rate. It represents the ratio number of patient cured (converts from smear positive to smear negative after completion of the treatment period), therefore indicating the centers performance. If the conversion rate is small, this is an indicator of problems in some of these centers.

The current system of the central unit is concerned about calculating the number of performance indicators, factors of each health center in order to achieve their objectives regarding detection and control of the spread of diseases'. These indicators are measured from the TB treatment outcomes categorized below:

1. ***Cured****:* The patient completes the treatment and the sputum test is negative. The conversion rate is calculated from the number of new cases that are cured.
2. ***Completed****:* The patient completed the treatment but doesn't make the test. The successful rate is calculated by computing the cured rate and the complete rate. Successful rate=cured rate/complete rate*100% (The successful rate must be 85%).
3. ***Died****:* The patient died either by TB or another disease. Died rate= (died/total)*100%
4. ***Failure****:* The patient still smear positive. Failure rate= (failure/total)*100%
5. ***Defaulters****:* The patient who returns to treatment. Defaulters rate = (defaulters/total)*100%. Patients who stop taking the drugs for more than a month are considered as defaulters. If we make an analysis we find that the defaulters' rate is the main influence of the TB program.
6. ***Transfer in****:* The patient who has been transferred from one TB register to another to continue treatment.

Figure 4.4: NDLTLD Indicator System



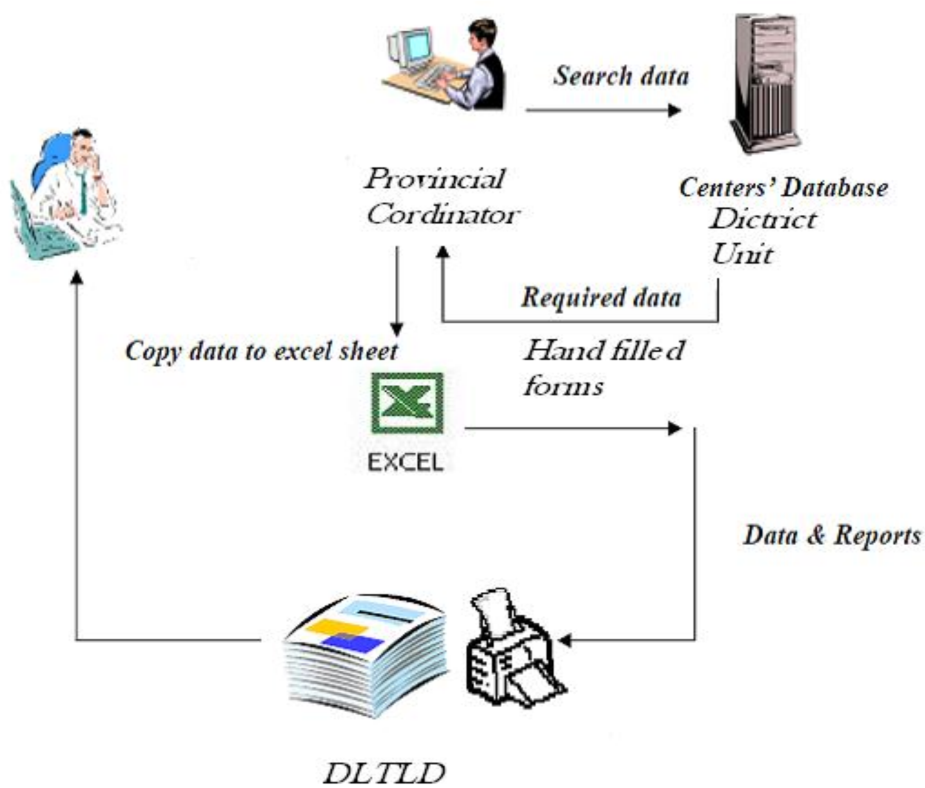
Source: Author, 2009

4.2 User Requirement

As mentioned earlier in problem definition, one of the main problem is lack of a computerized system and the current workflow of the health system processes carried by the TBMUs is insufficient in terms of helping county management and decision makers to exactly understand the current situation and accordingly assess the performance and identify of the deficiency and gaps of the health services provided.

The TBMUS coordinates all units all over the county by collecting quarterly reports from them. These reports are in a tabular manually done the quarterly reports are filled into preformatted report using pencils which are forwarded to the provincial unit which are put in excel sheet and forwarded to the central unit t his makes them less helpful to describe the how exactly each center works and exact situation as depicted in the figure below.

Figure 4.5: Current Data Flow



Source: Author, 2013

Informal interviews and document analysis are the techniques that were used to collect information with document analysis being the main tool that helped in identifying the requirements by analyzing the most helpful documents that existed; all current system documents were collected and analyzed.

Most of these documents associated with the current system are reports performed by the coordinators in Microsoft Excel format, the reports assisted in identifying what information should be kept about each treatment center, what kind of indicators need to be calculated and measures of each centers performance. In addition, when an informal interview has been conducted, the heads of district coordination centers gets a high-level perspective and a strategic view of the system in addition to understanding the big picture of the processes of the system being developed and automated. See table 4.2.

Table 4.3: Current System/System Improvement

Current System Features	**New System Features**
Reports are filled manually at the DOT and coordinating center and quarterly reports produced four times a year.	Data entered online and report can be produced any time.
The data maintained is tabular and there is no spatial dimension to it.	Data can be visualized on maps and linked to a location.
Analysis is done manually.	Analysis can be performed much quicker in addition to special analysis which can be done easily.
Performance is extracted from tabular report.	Centers' performance can be monitored And visualized.
No existing strategies for center service distribution and coverage area.	Coverage area can be visualized on maps and areas that require new services can be identified.
Only TB data is considered.	TB data can be linked to other factors.

4.2.1 Functional Requirements

After data gathering the functional requirements were designed to be handled by the proposed system as summarized below.

Table 4.4: Type of Users

Users Type	**Proposed System Functional Requirements**
Type 1	• Login to the system; • Add new records to database; • Update center's information.

Type 2	• Login to the system; • View centers' information on maps; • Generate new reports, charts and maps; • Perform queries on disease incidence & prevalence, services availability, distribution, and make spatial analysis (e.g. identify centers' catchment area).
Type 3	• Login to the system; • View centers' information on maps; • Perform quires on disease incidence & prevalence, services availability, distribution, etc.

4.2.2 None Functional Requirements

The system should be easy to use, learn, understand and able to work on different platforms including the web and mobile devices whose additional requirement are as follows:

4.2.3 Operational Requirements

- System operates on windows XP or windows Vista;
- System able to be viewed in internet explorer or fire fox browsers or on the desktop;
- Standardizations.

4.2.4 Performance Requirements

- Data consistency and reliability;
- Rapid response time;
- User friendly interface.

4.2.5 Security Requirements

- Personal (Patient) information security–considering ethical issues;
- Password is required to update the system.

4.3 Conceptual Business Process and Data Design

This is the process of constructing a model of the information used in an enterprise, independent of all physical considerations [Conn1998].

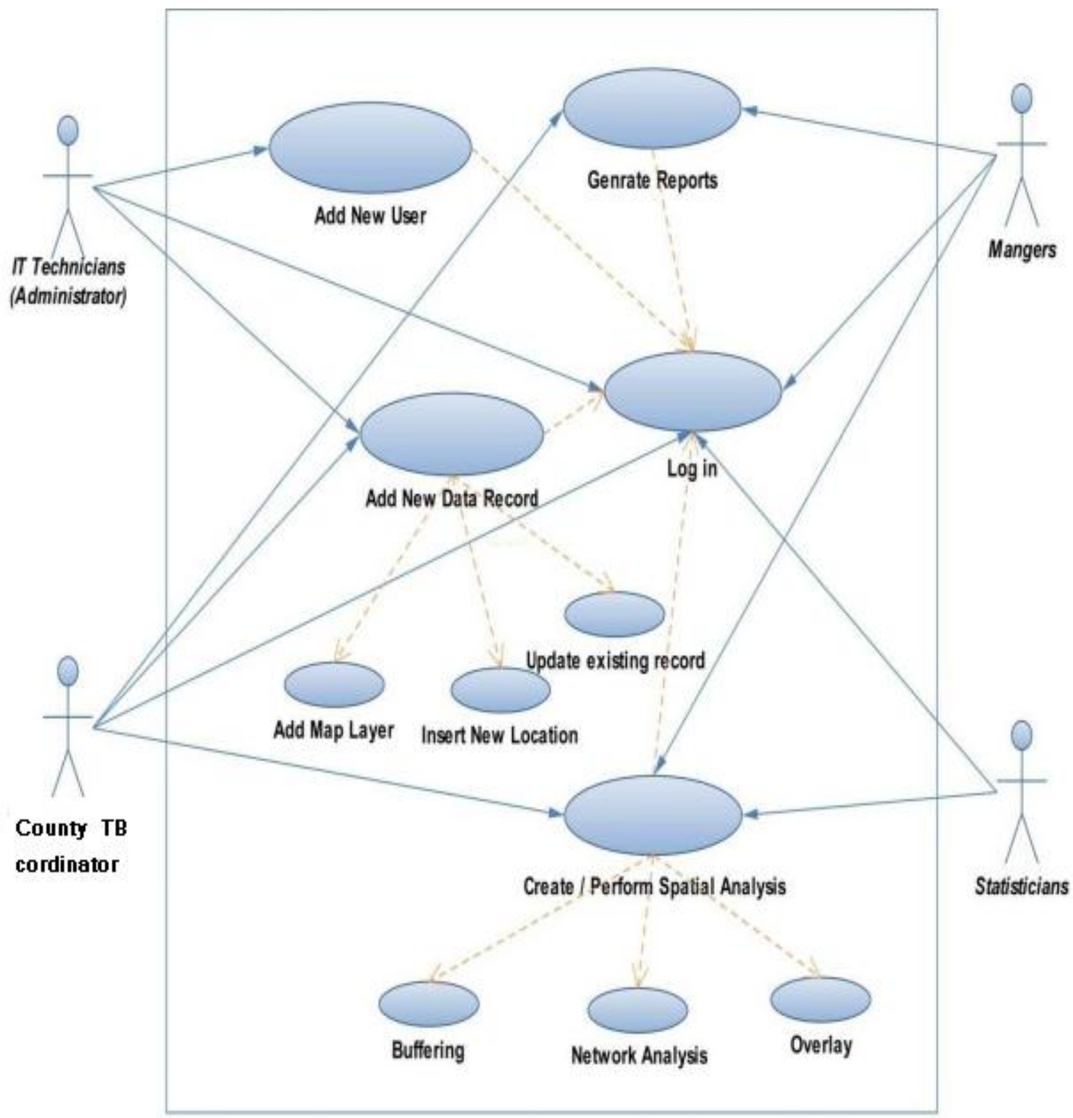
4.3.1 Use Case

Use cases are a scenario based technique for requirement elicitation which were first introduced in the Object Factory for Software Development –Objectory method (Jacobson *et al*., 1993).

They are the simplest representation of a user's interaction with the system and they depict the specifications of a use case. A use case diagram can portray the different types of users of a system and the various ways in which they interact with the system.

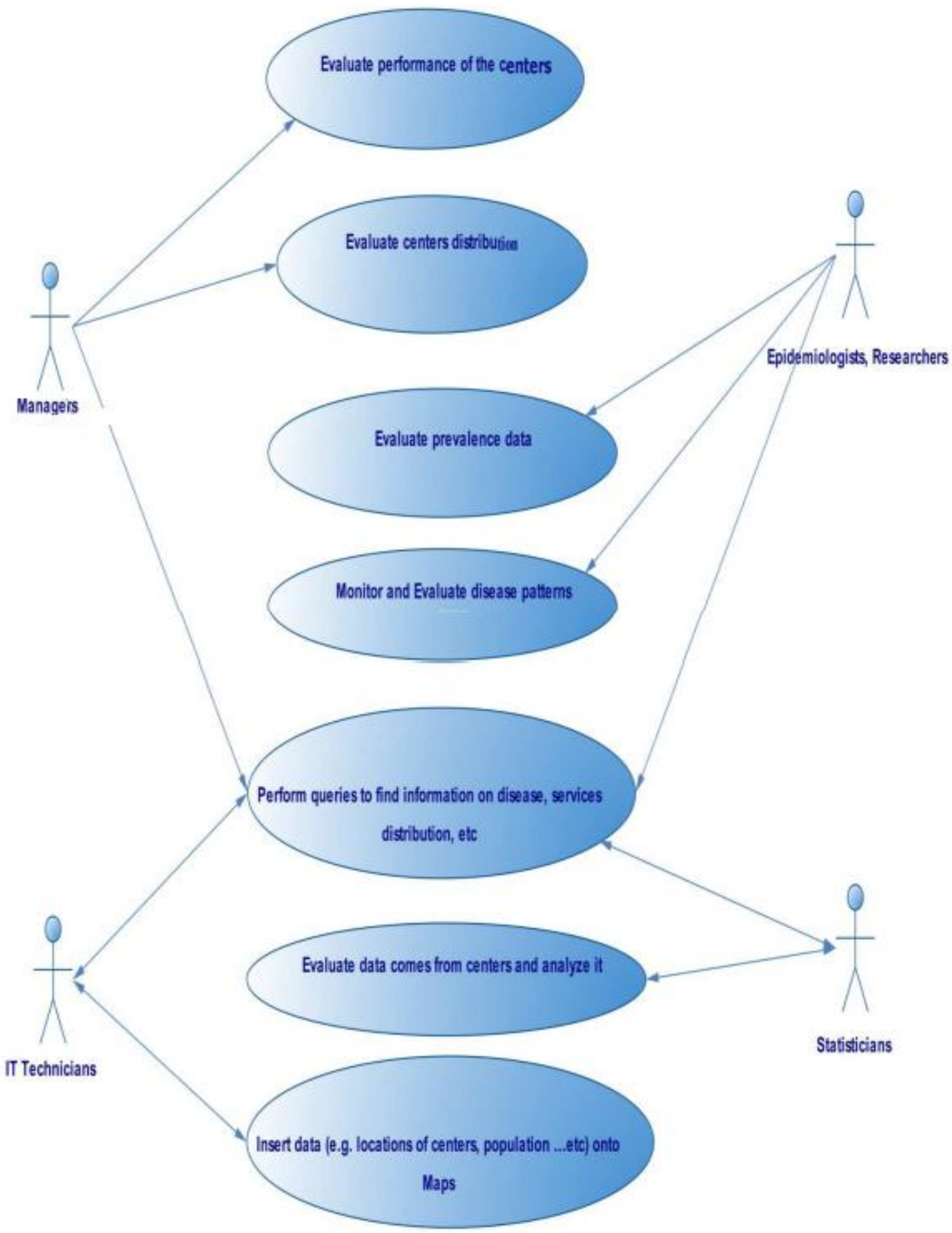
Figures 4.6 and 4.7 below illustrate the system requirements and demonstrate how each stakeholder interacts with the system.

Figure 4.6: Use case Diagram 1



Source: Author, 2013

Figure 4.7: Use Case Diagram 2



Source: Author, 2013

4.3.2 System Architecture

4.3.2.1 GIS Database-geospatial database

The geo-database is the common data storage and management framework for GIS. It combines "geo" (spatial data) with "database" (data repository) to create a central data repository for spatial data storage and management. The Geographical data representations in GIS are done by geo-database which allows users to specify how certain features were represented. GIS supports two types of data; vector and raster data. Vector data can be any of these types; points, lines, polygons etc. For example, parcels were typically represented as polygons; streets were mapped as lines, wells as points, and so on. These features are collected and put into feature classes in which each collection has a common geographic representation. Raster data sets such as digital elevation models and imagery such as Pixels, a location, value, satellite images and aerial or photos are already in this format.

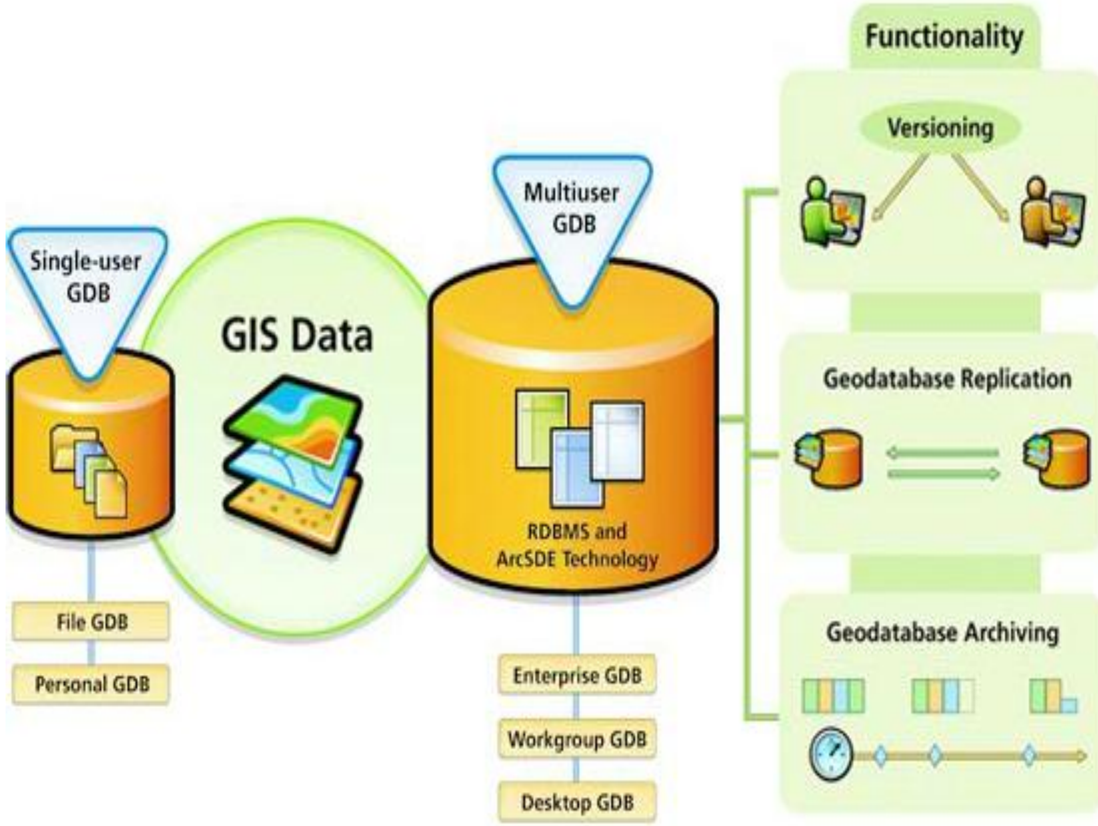
Spatial data is often categorized in two forms: filed-based and object-based models (Shekhar & Chawala, 2003; Rigaux, 2002). The first model is always associated with satellite imagery and raster data derived from grid-based collection methods, while the latter identifies discrete spatial objects.

When the map is created and opened with ArcGIS software, an attribute table is generated automatically for each layer and consists of three main columns that determine the ID number, the parameter and area of polygons. Identifying any new features or attribute such as the existence of a general health service or DOT centre, can be done by adding new field in the attribute table to include the new features. Additional tabular data can be added to this table either by inserting or importing it from other database programs. This can be achieved by using the JOIN function to create a join field in the table.

The spatial datasets included in this system are those which represent the geographic aspects of states, cities, localities, hospitals, streets and treatment centers. These could be polygons, points and lines to demonstrate the relative location and geometric form of each object in a map and to be linked with other non-spatial attributes in traditional databases that describe additional object's characteristics.

The maps generated by the system displayed the disease point of distribution according to the number of cases, shown by evidence of clustering. They also identified the catchment areas for each health facility in addition to the quality of service based on the facilities infrastructure and further the performance gauged on the treatment indicators highlighted before. All the above Information was up up-to-date information because each center's coordinator was able to add, update and view the information at anytime.

Figure 4.8: GIS Data Base



Source: ESRI, 2013

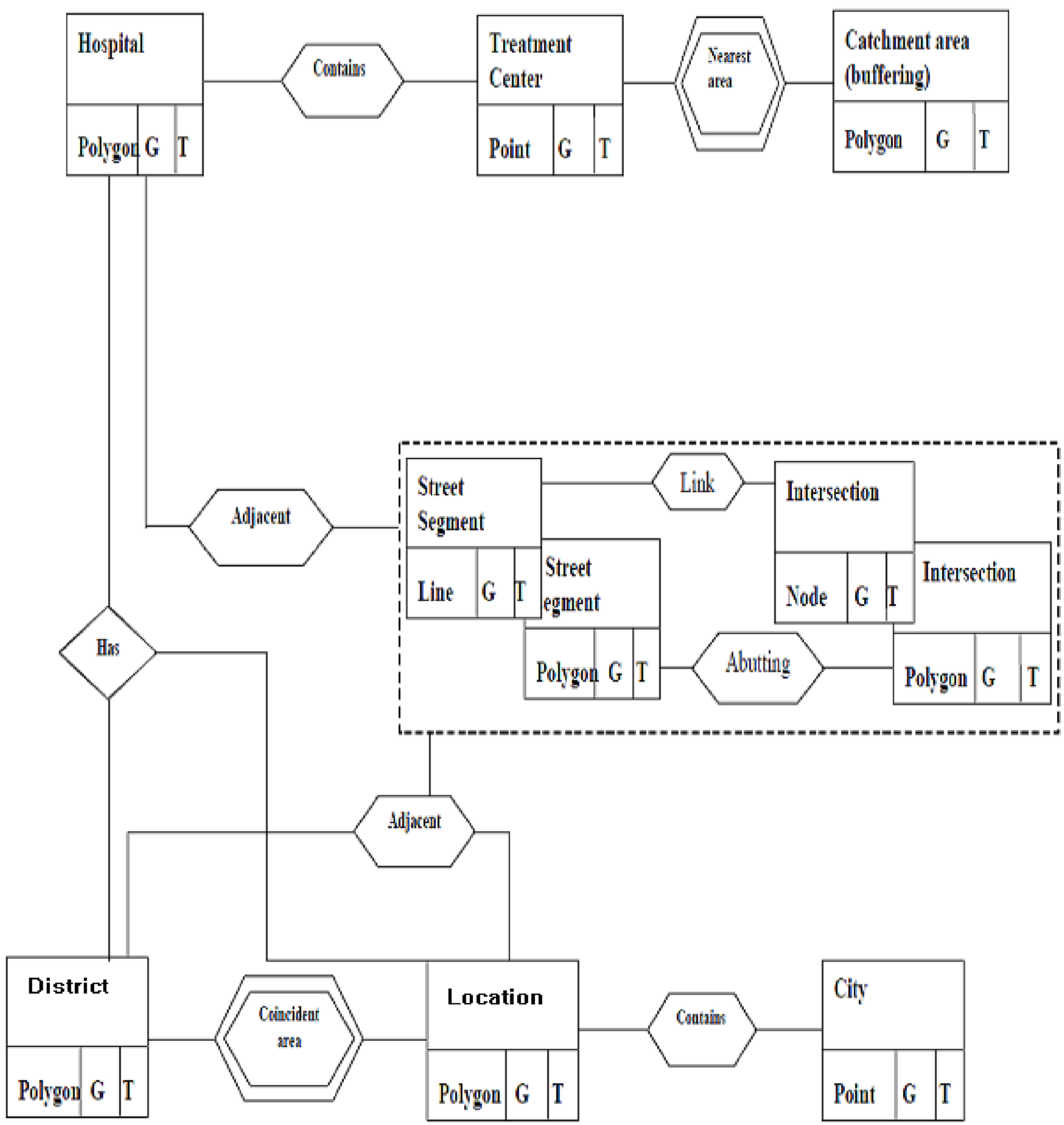
All the information was represented in layers on the maps. The maps have a layer that demonstrates demographic data, health facilities locations and when a user points over them, they get highlighted giving information about each center and further, on clicking one can get extra information. Another layer that views the number of cases of each treatment center displays all calculated indicators. The result of queries and analysis were expressed as layer on the map.

4.3.2.2 Conceptual model

"Conceptual data modeling is used to identify data content and describe data at abstract or conceptual level". (Elmasri and Navathe,1994). The importance of conceptual modeling comes from the fact that it defines a structure that promotes communication between data procedures, tools and data consumers. (Steve Grisé, ESRI). There are numerous techniques/tools that have been developed in order to be used for conceptual database modeling activities. However not all of them support the GIS database. Extended Entity Relationship technique has been developed in order to cover spatial objects and relationship representation, the topological attributes associated with spatial objects and spatial operations art of the relationships among data objects in GIS are implemented via software function. For example, the spatial coincidence relationship derived through the use of topological overlay spatial operation whose software function is unlike business applications where all relationships are explicitly represented in the database and the associated software only oriented to query and report generation functions. In summary, there are three types of relationships that were represented in a geographic database with an "object view" orientation:

- Relational relation represented by means of keys (primary and secondary);
- Spatial relationships represented in the GIS portion of the database by topology;
- Spatial relationships that exist only after a calculation is made on the (x, y) coordinates.

Figure 4.9: Conceptual Data Model



Source: Author, 2013

4.4 Physical Design

4.4.1 Database Design-Non Spatial Database

1. District Table

Table Name	Districts Table		
Primary Key	District Code		
Foreign Key			
Purpose			
Data Obtained From			
Field Name	Data Type	Length	Description
District Name	Varchar	70	District Name

2. Location Table

Table_Name	Location Table		
Primary Key	Location_ID		
Foreign Key	District_ID User_ID		
Purpose	It shows the number of health services and treatment centers, in addition to its demographic data for specific locality.		
Data Obtained From			
Field Name	Data Type	Length	Description
Location Name	Varchar	70	Location Name

3. Health Facilities Table

Table Name	Hospital_Table		
Primary Key	HF_ID		
Foreign Key	***Location_ ID User_ID***		
Purpose	This table describes the required information about each health facility in each district. Whether it is a hospital, health center or a dispensary.		
Data Obtained From			
Field Name	**Data Type**	**Length**	**Description**
HSP Name	*Varchar*	*70*	*Hospital Name.*
HSP Class	*Varchar*	*30*	*Were; classify as either government private or NGO.*
No. Doctors	*Integer*	*20*	*Number of doctors in the facility.*
No. Beds	*Integer*	*20*	*Number of beds in the facility.*
Type	*Enum*	*5*	*Hospital, Health center or dispensary.*
Specialization	*Varchar*	*70*	*If the entry type is hospital this field will get a value to identify whether it is a dentist, Obs or general hospital. Otherwise null.*

4. New Case Table

Table Name	New_Case_Table			
Primary Key	Center_ID			
Foreign Key	***HF_ ID User_ID***			
Purpose	This table shows each treatment center's information. Regarding the number of new TB cases aggregated every quarter month.			
Data Obtained From	TBMU / DOT Center's quarterly report.			
Field Name		**Data Type**	**Length**	**Description**
Year		*Enum*	*2*	Shows whether the center is working as *TBMU or DOT Center.*
Quater		*Integer*	*50*	Month name.
New cases		*Integer*	*20*	Number of new TB cases for the specified month.
Smear –ve		*Integer*	*20*	Number of cases classified as smear negative from the sputum test.
Smear +ve		*Integer*	*20*	Number of cases classified as smear positive from the sputum test.
Extra pulmonary		*Integer*	*20*	Number of cases classified as extra pulmonary.
Pulmonary		*Integer*	*20*	Number of cases classified as pulmonary.
Relapse		*Integer*	*20*	
Age Group	*<15*	*Integer*	*20*	Number of cases in each age group; age groups are divided into 6 categories starting from less than 15 years old up to above 50 years old.
	15-19	*Integer*	*20*	
	20 -29	*Integer*	*20*	
	30 -39	*Integer*	*20*	
	40-49	*Integer*	*20*	
	>50	*Integer*	*20*	
Male		*Integer*	*20*	How many cases are male?
Female		*Integer*	*20*	How many cases are female?

5. Treatment Outcome Table

Table Name	**Treatment Outcome Table**		
Primary Key	ID_NO		
Foreign Key	***TCenter ID/ User_ID***		
Purpose	To keep information about the treatment outcomes and performance indictors of each center. Indicators put in this table are those used to plan strategies and set policies.		
Data Obtained From	Calculation is done on data that has been kept in both TB/HIV new cases tables to get these measurements.		
Field Name	**Data Type**	**Length**	**Description**
Quarter	*Varchar*	*2*	Illustrate which year quarter for specified center; reports are generated by year quarter.
New cases	*Integer*	*50*	Total number of cases on specified quarter.
Defaulters	*Integer*	*20*	The rate of those patients who relapse from treatment. Started treatment and stopped before completion.
Died	*Integer r*	*20*	Rate of death among TB patients within a specified quarter.
Cured	*Integer*	*20*	Rate of cases that completed the treatment and tested as positive. Type.
Complete	*Integer*	*20*	Rate of cases that just complete the treatment but are not tested.
Transferred	*Integer*		Rate of those who changed their treatment center to another.

4.4.2 The Program Structure

This prototype is Geographic Information System to be used for managing TB data in the District coordinating centers (TBMU) and supporting the decision making process by visualizing treatment reports on a map, thus making extracting knowledge from these reports easier and clearer for decision makers. In order to achieve these objectives, several tools have been used to design and develop the system with specified requirements. They have been chosen considering the resource constraints in the environment, and the prohibitive cost of licenses. Most of the tools are open source making the implementation process of this system uncomplicated and trouble-free. These tools are shown and described below:

4.4.2.1 ArcGIS 10.1

This a package consisting of a group of geographic information system software products produced by ESRI (Environmental Systems Research Institute) that allow one to view, create, edit and query maps including capabilities for data manipulation and analysis. ArcGIS (version 10.1) is used in this system to prepare maps in Kitui County to create the geodatabase and map layers. Identify health services' locations, catchment areas and treatment outcomes.

4.4.2.2 File Maker Pro

FileMaker Pro is a cross-platform relational database application from FileMaker Inc., formerly Claris, and a subsidiary of Apple Inc. It integrates a database engine with a GUI-based interface, allowing users to modify the database by dragging new elements into layouts, screens, or forms. It contains features to build your own functions and copy, paste, and import them into any FileMaker Pro database, build or modify databases faster by importing multiple tables at once, build applications where all menus are hidden, run comprehensive reports on all elements of the database schema, pinpoint problem areas in scripts and Script Triggers, NEW–Turn off Script Triggers when debugging to fine tune the troubleshooting process, monitor fields, variables and calculations while troubleshooting.

4.5 System Requirements

Table 4.5 Hardware Requirements

Hardware	Minimal Requirement
Processor	Intel Pentium Processor 1.0
Memory	1 GB RAM
Hard disk	60 GB Hard disk
Monitor	1,024x768 resolution
Network	1 Gbps speed

Table 4.6 Software Requirements

Software	Minimal Requirement
ArcGIS 10.1	GIS Software to create and edit geodatabase maps or higher.
File maker	Database management system to create and update database tables and export data to GIS compatible file format.
Documentation and designing tools	Microsoft office 2007–Word and Visio. Microsoft project.

4.6 Test Plan

The test plan identifies the specific functional tests that are to be done in order to ensure that all functional requirement and none functional requirement of the system have been fairly met.

4.6.1 Test Methodology

The test was done as follows:-

1. **TB management application**
   All the three type of users were tested to ensure that their access rights and privileges allowed them perform their duties
2. **Maps**
   This is to ensure that maps are created properly and each element (polygons, lines and points) is drawn correctly in order to avoid getting wrong spatial analysis or have ambiguous results.
3. **Reports**
   The reports were cross checked to ensure that they met the output objective of the system.

4.7 Testing Phases

An error based testing has been conducted to test and validate the developed system, the strategy of this test has been done according to the following stages: First, testing of the developed application was conducted using unit test. Second, overall system tests were performed to validate all basic functionalities that were identified according to the user requirements. This was achieved using black box testing. Then an integration test was start to ensure that the system is well integrated and all components are working properly also to detect and fix errors. And finally was users' acceptance of the tests.

4.7.1 Unit Test

Unit test validates the basic functionality. It concludes classes (smallest unit), package data (attributes) and operators (method, services) that manipulate this data. The entry criteria for these tests are stated below:

- Functional requirements have been defined, written, and approved;
- Class diagram is ready to use;
- Configuration management process for handling version control, error handling, hardware and software configuration is defined and adhered to;
- The exit criteria for the unit test are: All modules defined in the unit test criteria have been UT tested successfully;
- Unit Test results have been documented and incorporated into overall testing results summary and approved.

4.7.2 System Test

This test is performed to validate that the system meets all organization requirements. Black box testing verifies the functions processing of valid and invalid inputs. The entry criteria for this test are:

- Test environment (hardware, software) has been installed, tested, and ready for use in this phase of testing;
- All data to be tested has been created, populated and/or re-initiated as needed;
- All test conditions and test cases are finalized.

The exit criteria are:

- The code satisfies the business requirements;
- All critical and major defects are resolved.

4.7.3 Integration Test

Integration test is done to ensure that the all integrated units in the system have been performed smoothly, and the assembly of the systems components collaborates as intended, as well as, that the system has the expected behavior. The entry criteria are:

- System testing exit criteria has been met;
- Use cases are ready for use;
- Test environments (hardware, software, test tools) have been installed, tested, and ready for this phase (SIT) of testing;
- Test data has been defined and populated into the system as necessary.

Exit criteria are:

- All critical and major defects are resolved.

4.7.4 Acceptance Test

The purpose of testing customer acceptance of the developed system is to determine the usefulness of the system and also to measure the functionality and ease of use.

Entry criteria:

- Customer acceptance test plan have been defined;
- Test environments (hardware, software, test tools) have been installed, tested and ready for this phase of testing;
- Test data has been defined and populated into the system as necessary.

Exit criteria:

- All requirements are determined to be fully functional;
- Test results have been documented, incorporated into the Test Summary Results document, and approved;
- The application has received approval for implementation;

CHAPTER FIVE

DESIGN IMPLEMENTATION AND TESTING

5.0 Prototype Implementation

As per Chapter Three of this project, this system was implemented using vertical prototyping combined with waterfall model. The key functions were identified, implemented and fully tested. How the functionalities are concentrated in the main purpose of the project, which is supporting the decision making process by producing a valuable report in visual format in addition to giving spatial dimension. Kitui County was taken as a case study and the results were illustrated in table 5.1and figures 5.1-5.11 below.

Results

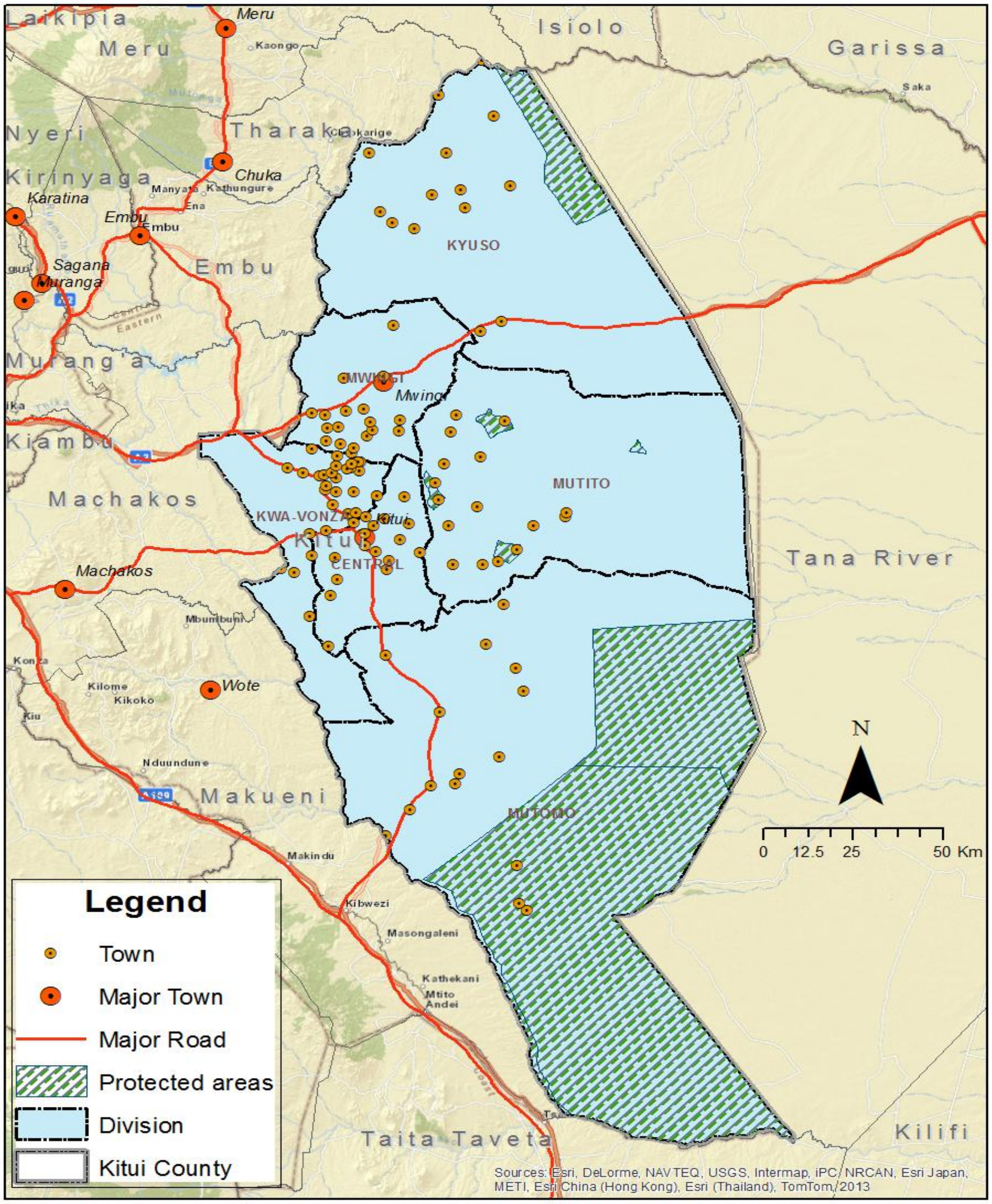
1. Most of the major towns and roads report the highest prevalence of TB as a figure.
2. The highest number of deaths was reported in the central part of the county and also in areas with less facility coverage and accessibility.
3. There was a clear relationship between the total number of reported cases, the total death and the total number of patients living with HIV and AIDS.
4. The treatment outcome of the data set was summarized in table 5.1 below.

Table 5.1 TB Treatment Outcome Table

Outcome		**Percentage**
Cured		34%
Death rate		6.4%
Death by age	0-15	10%
	16-20	8%
	21-54	65%
	55-89	17%
Out of control		2.4%
Treatment completed		33.3%
HIV		29%
Dead and HIV positive		13.5 %

Figure 5.1 is the general map of Kitui County. The map demonstrates administration blocks, main towns, major roads, agricultural regions and restricted areas. This gives an indication of probable distribution of TB incidents and there by give general areas of interest to the county TB stakeholders and decision makers.

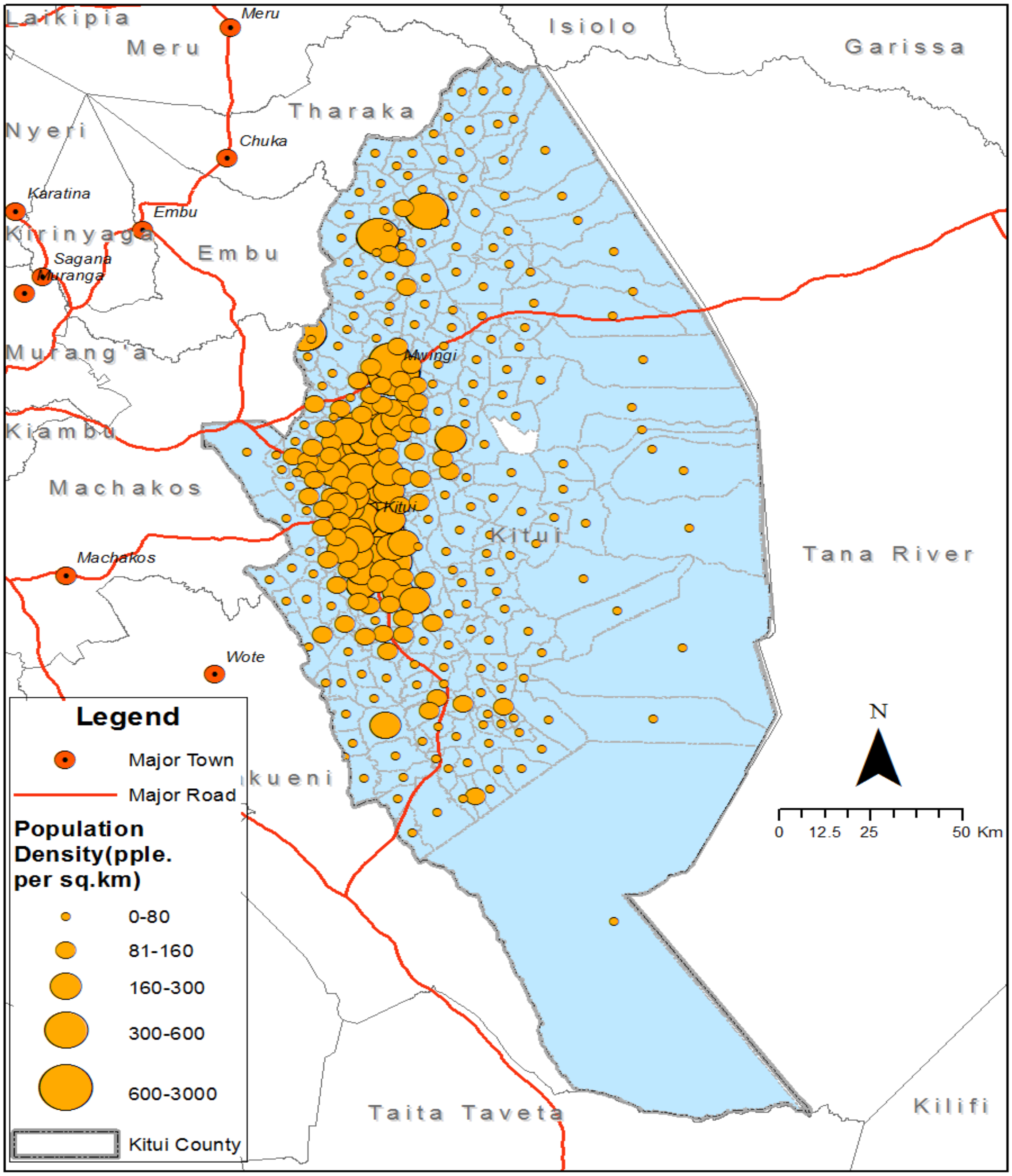
Figure 5.1: Kitui County General Information



Source: Author, 2013

The demographic layer represents distribution of population per square kilometer according to sub-location-administrative areas. See figure 5.2 below. The higher the population is, the higher the TB prevalence. This can be used to predict the sub location with higher TB infection rate.

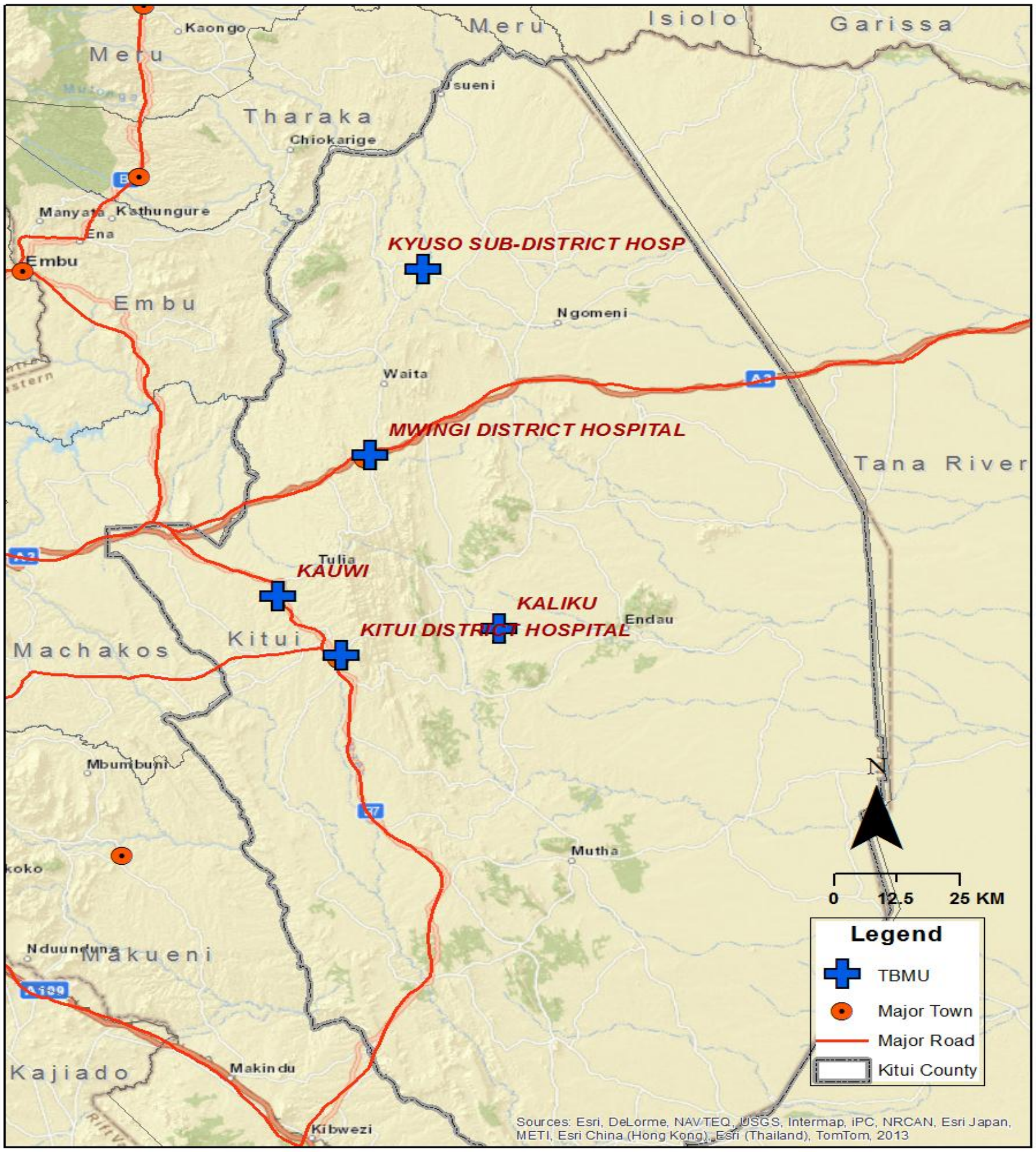
Figure 5.2: Kitui County Population Density per Sub location



Source: Author, 2013

Figure 5.3 shows the distribution of TB Management Units. The distribution is based on the location of the district hospitals and not any substantial information on population density, poverty index, rate of TB infection or any other TB contributing factors. The outcomes that follow below, can assist stakeholders to determine location of new TBMUs.

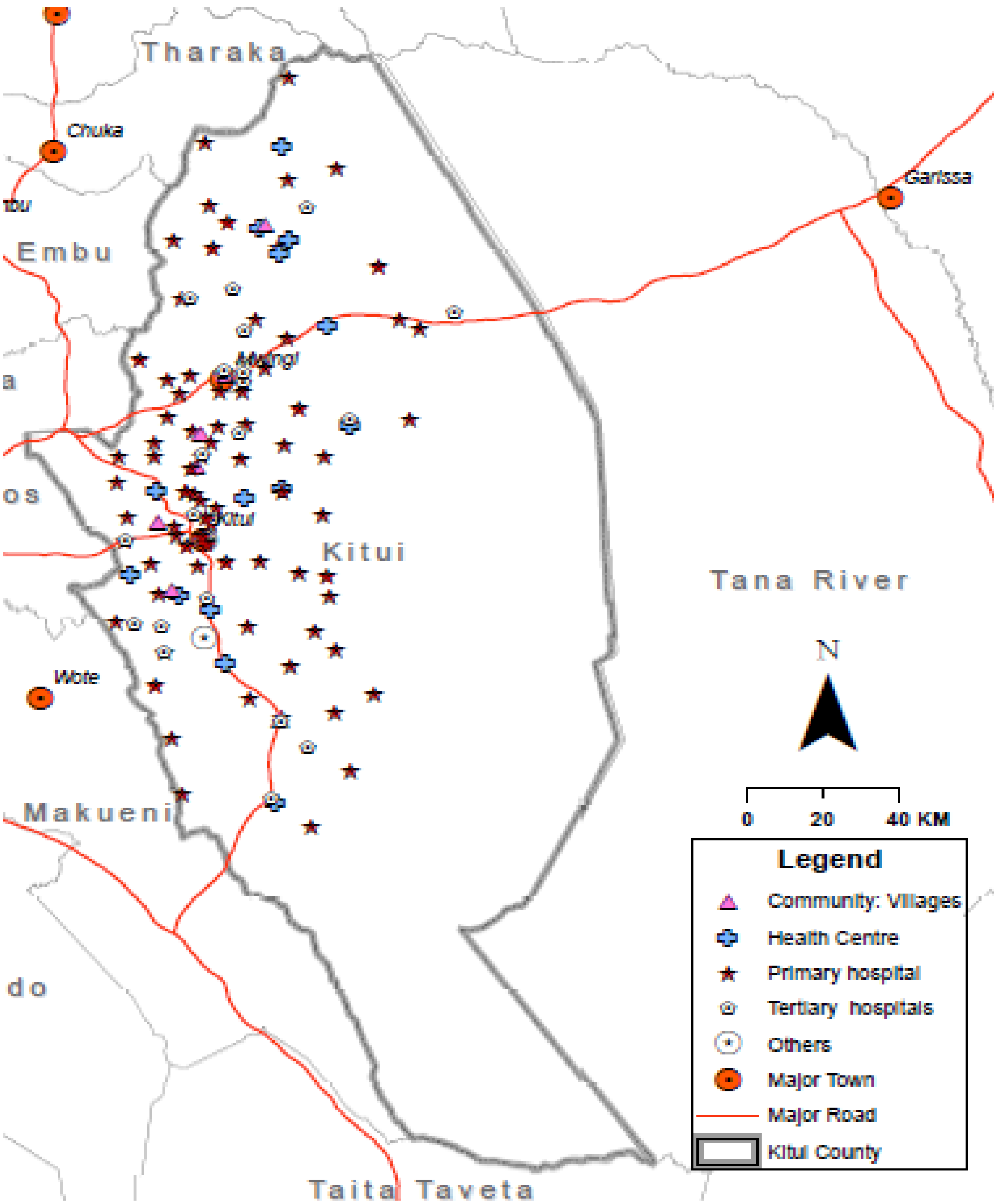
Figure 5.3: Kitui County TB Management Units



Source: Author, 2013

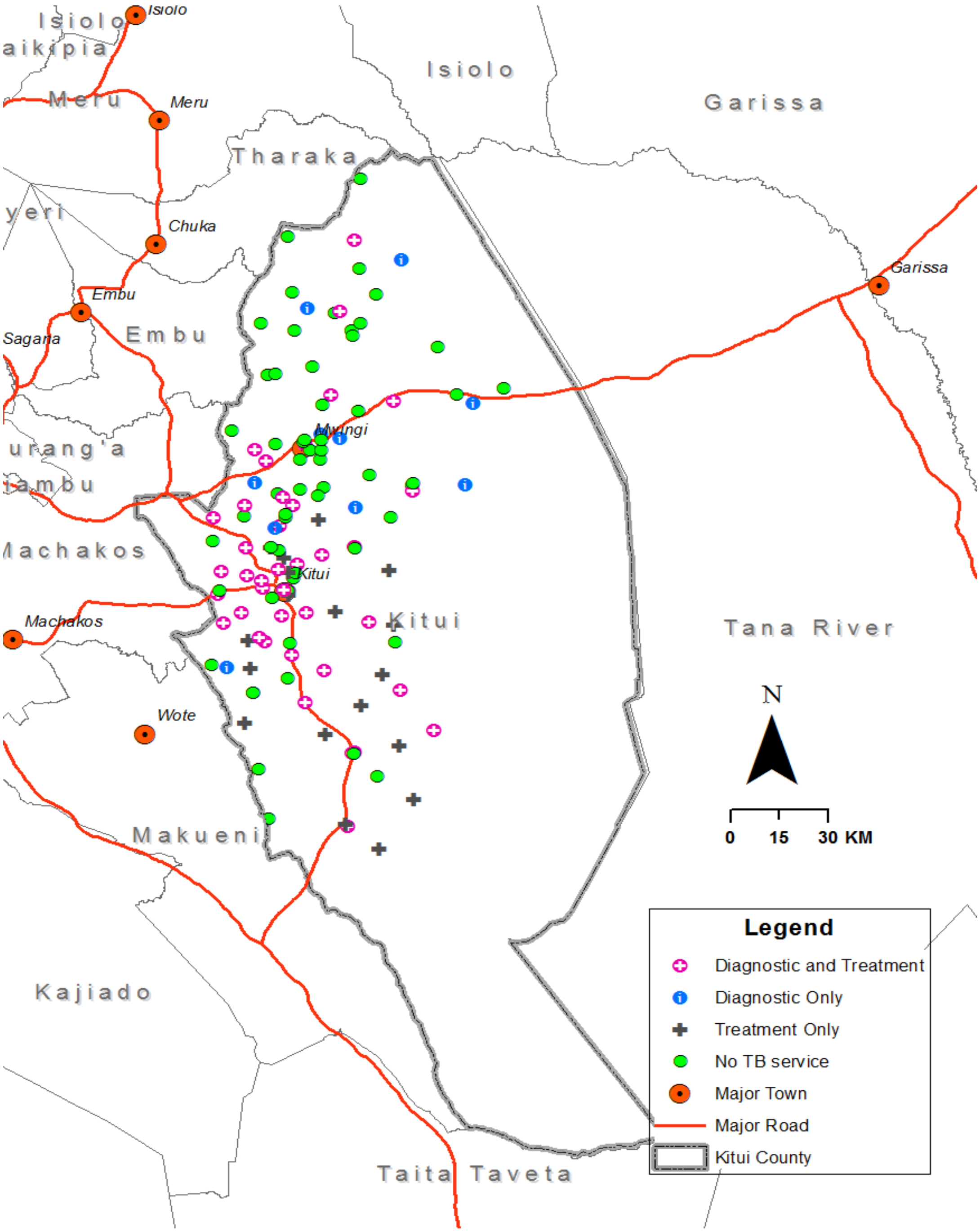
The heath care system in Kenya comprises different players. These include government, church based organization, private hospitals and NGOs. The facilities are further classified as health centers, primary hospitals and tertiary hospitals. Any of these can act as TBMU, diagnostic or treatment centers. The allocation of a facility as either of these is not informed by any information. With spatial distribution of the health facilities as in figure 5.4 and the current treatment and diagnostics centers as in figures 5.5, these can be used to determine the facilities that can hence be promoted in future as TBMUs, diagnostics or treatment centers.

Figure 5.4: Health Facilities Layer



Source: Author, 2013

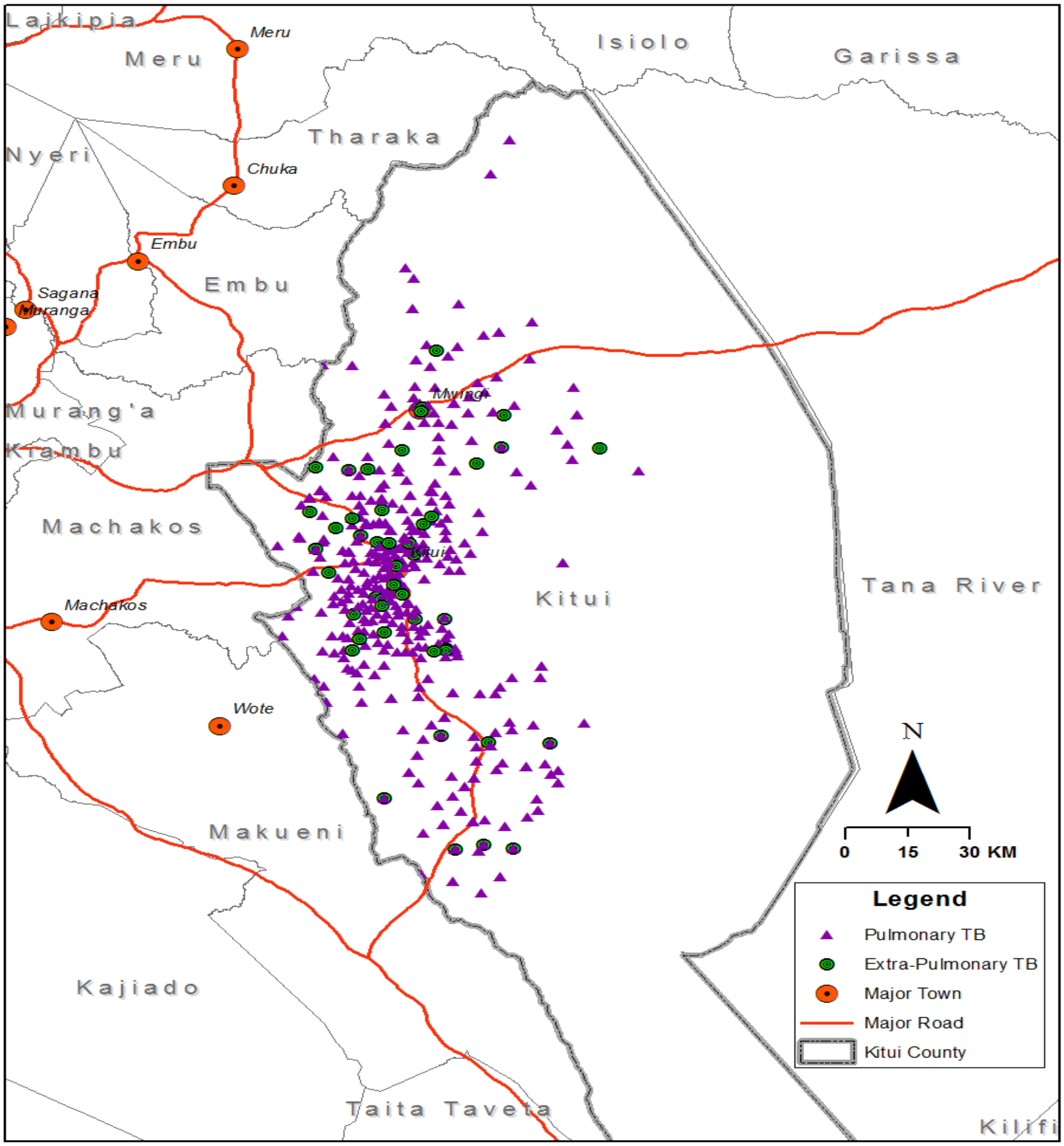
Figure 5.5: County Diagnostic and Treatment Centers



Source: Author, 2013

TB is classified as pulmonary (P) and extra pulmonary (EP). The pulmonary type affects only the respiratory system while extra pulmonary affect other body organs and therefore are considered to be more fatal and contribute to more TB deaths. The layer below figure 5.6 shows the distribution of the two types of TB in Kitui County. This can help determine the allocation of resources and personnel in the county for the diagnostic and treatment guided by the type of TB.

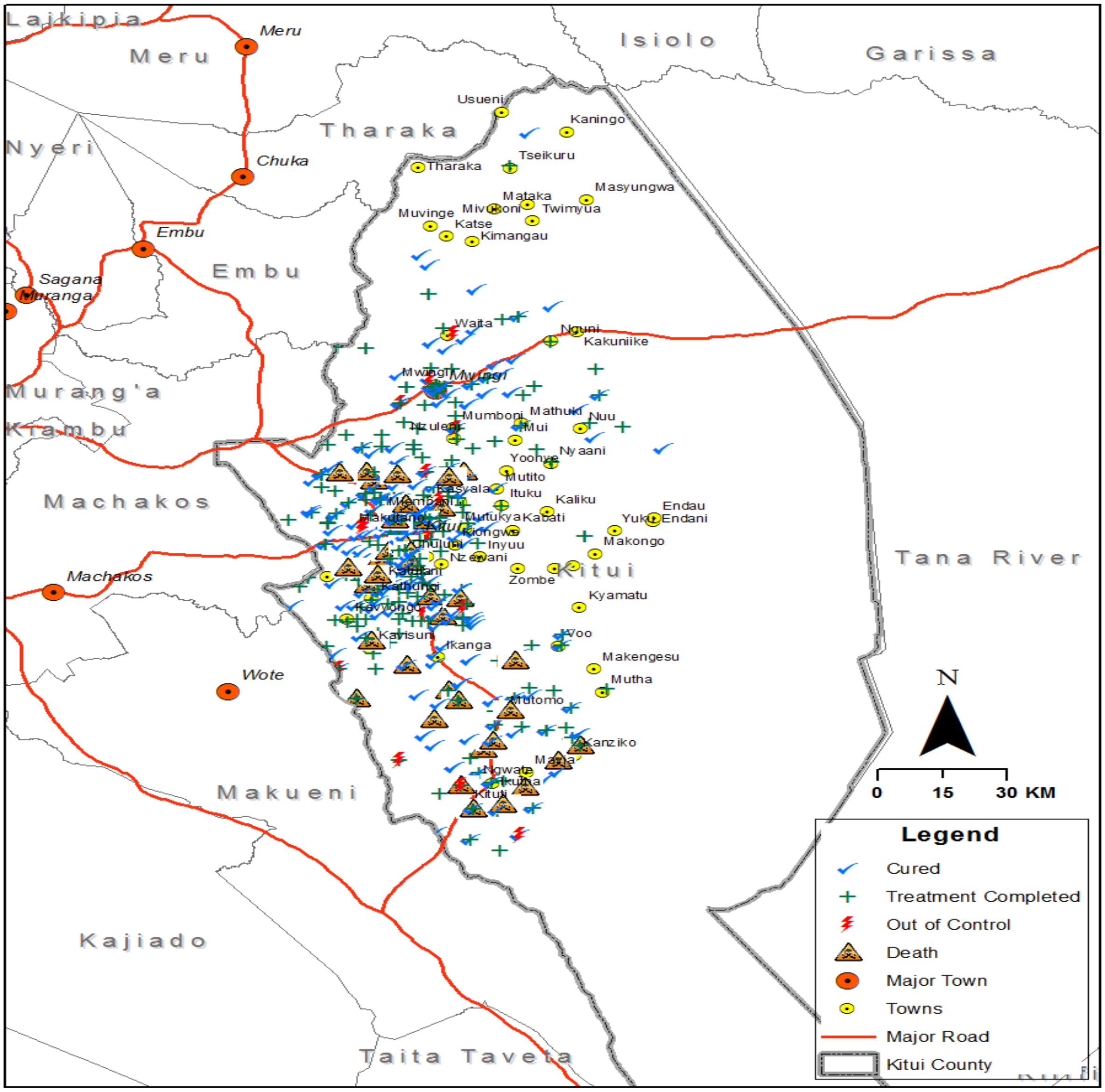
Figure 5.6: TB Distribution by Type



Source: Author, 2013

TB treatment outcomes are classified as either cured (C), treatment completed (TC), died (D) or out of control (OOC). The next layer depicted in figure 5.7, shows distribution of TB incidences classified by outcome. It visualized the areas as high cured rate, death rate and out of control. This information can assists the stakeholders determine the areas with high density of TB incidences, the type of TB, best outputs of cured that are completed and the worst cases of out of control and deaths in order to ask the question why and be able to come up with strategies plans to improve on these services.

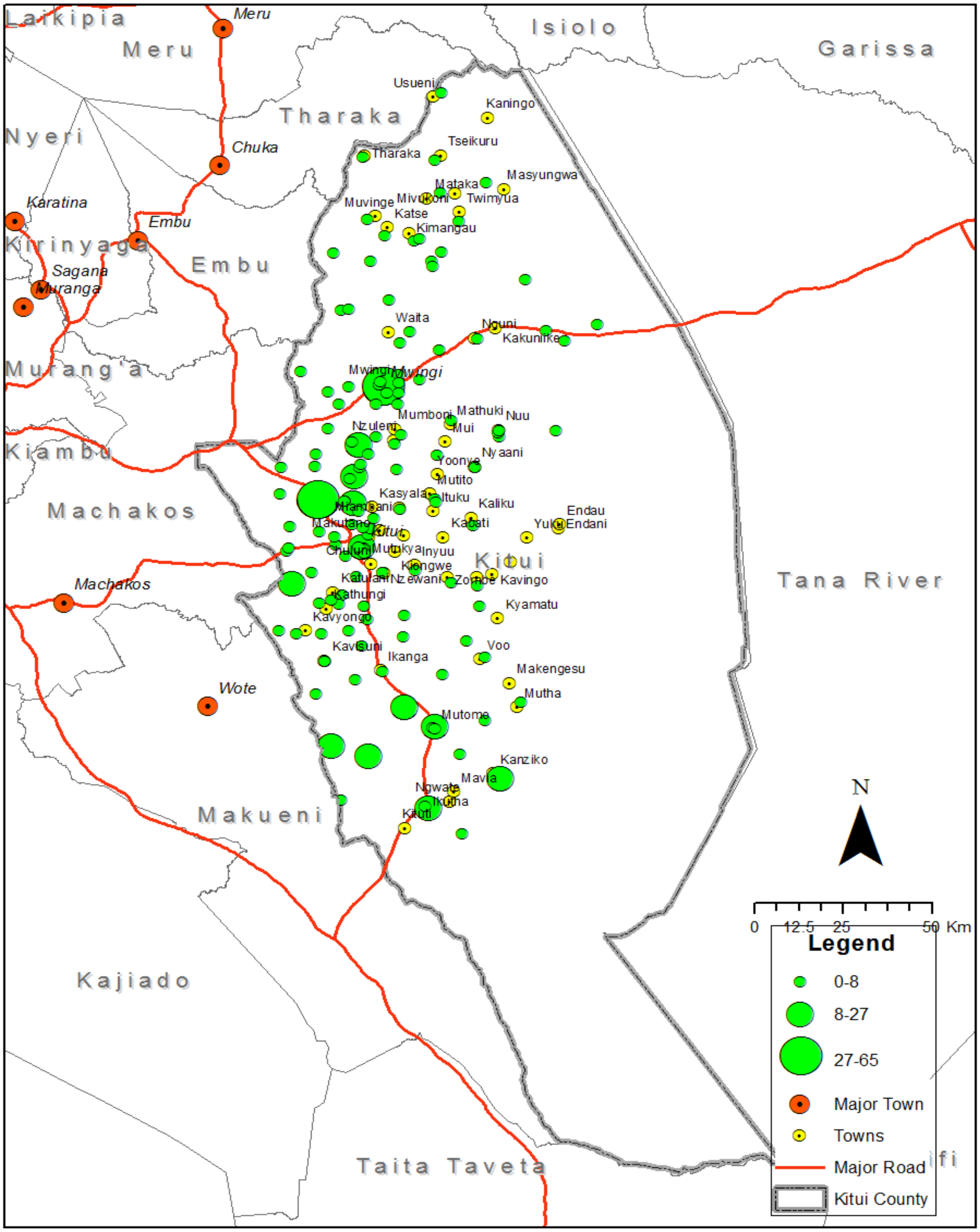
Figure 5.7: TB Distributions by Outcome



Source: Author, 2013

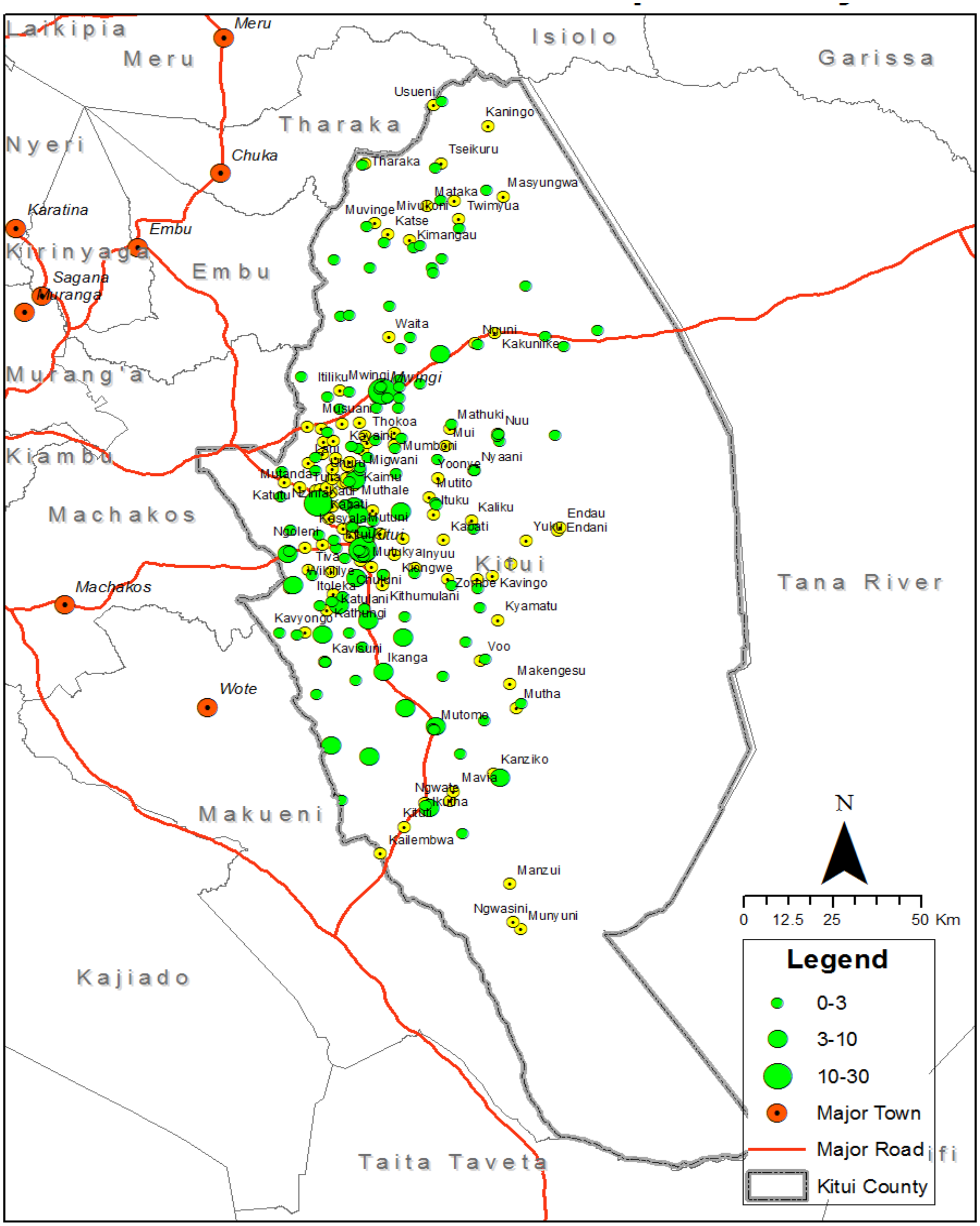
Indicators are calculated at central unit to determine the performance of the facilities. The layer below, figure 5.8 shows the aggregation of new case per facility combined with other information such as population density figure 5.2. Outcomes such as Total cured per facility and Total deaths per facility in figures 5.9, and 5.10 can be used to determine which facilities and resource are strained, find out areas with poor performance and high death rate and therefore use this to ask question why this is so, investigate the causes and come up with strategies which will be used to make decision on allocation of resources, personnel, setting up of new facilities and upgrading of the existing ones in order to improve on service delivery.

Figure 5.8: New TB patients Distributions per Facility



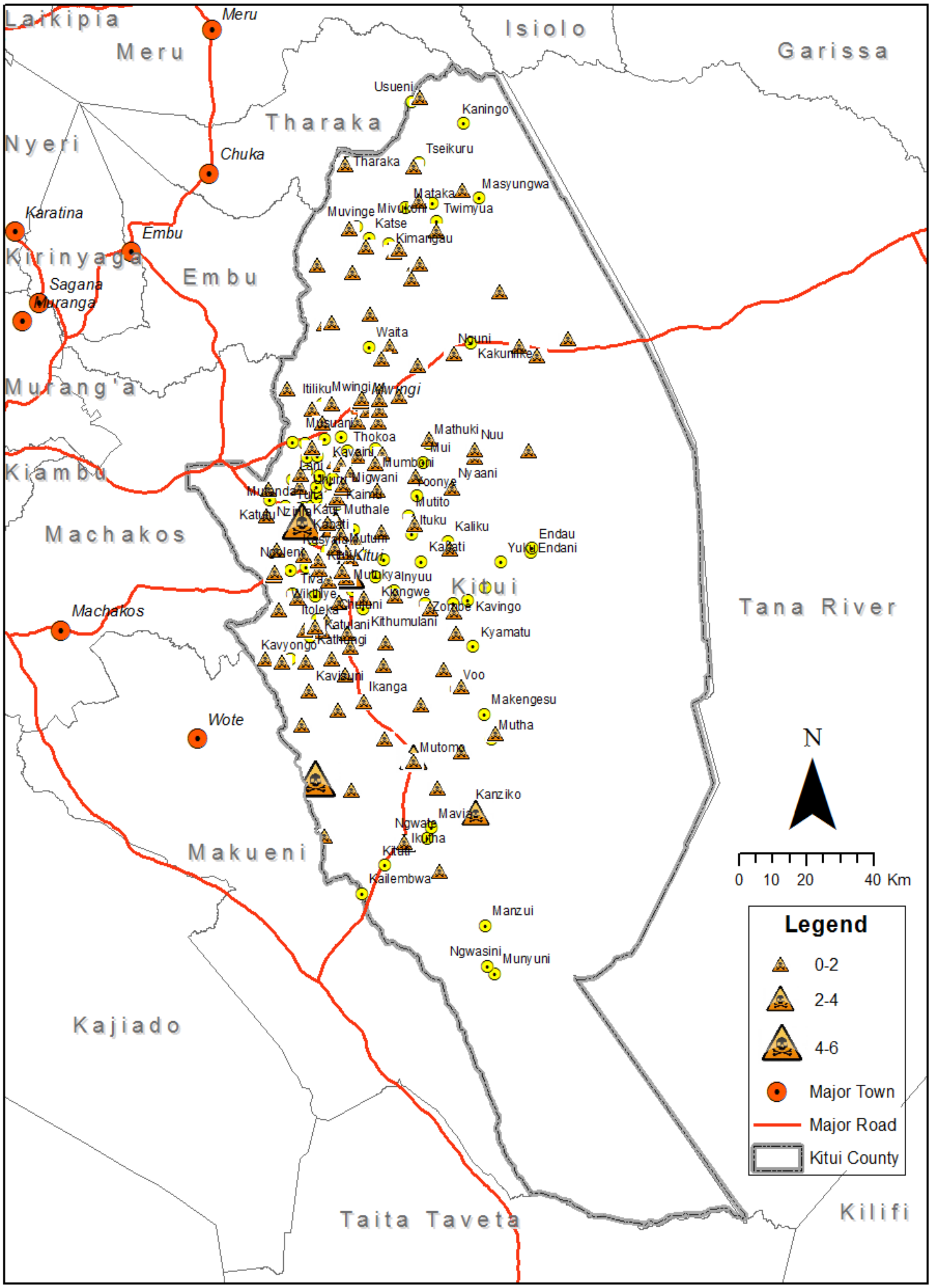
Source: Author, 2013

Figure 5.9: Total Cured Per Facility



Source: Author, 2013

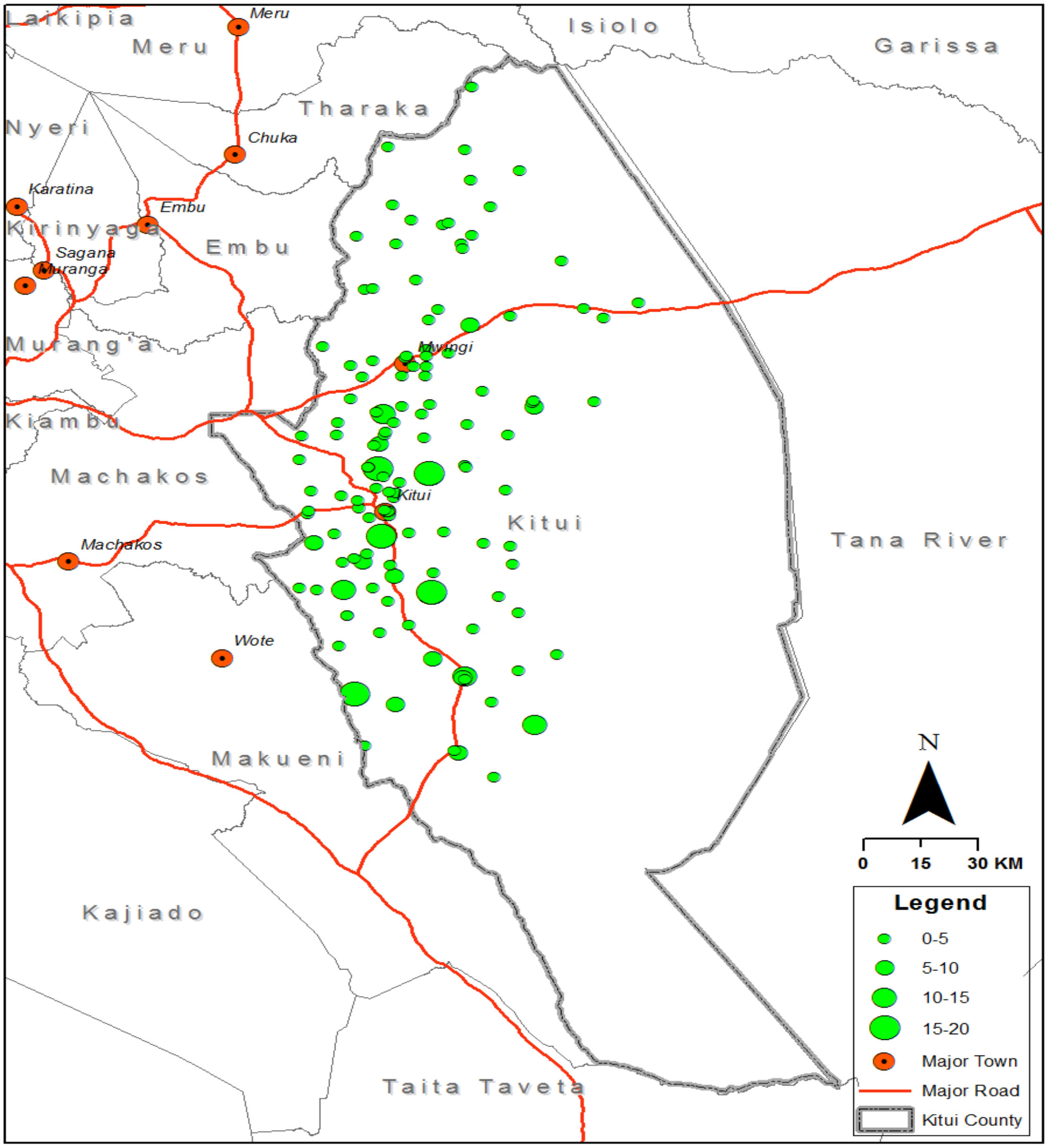
Figure 5.10: Total Deaths per Facility



Source: Author, 2013

Most people with TB may test negative since the immune system may try to fight out the infection, but weakened by HIV makes them susceptible to TB infection. There is evidence of correlation between TB infection rate and HIV rate. This is depicted in figure 5.11 below. Combined with deaths outcome in figures 5.10, this can shed more light on the hot spots which need to be given more attention in terms of TB and ART resources to prevent further deaths.

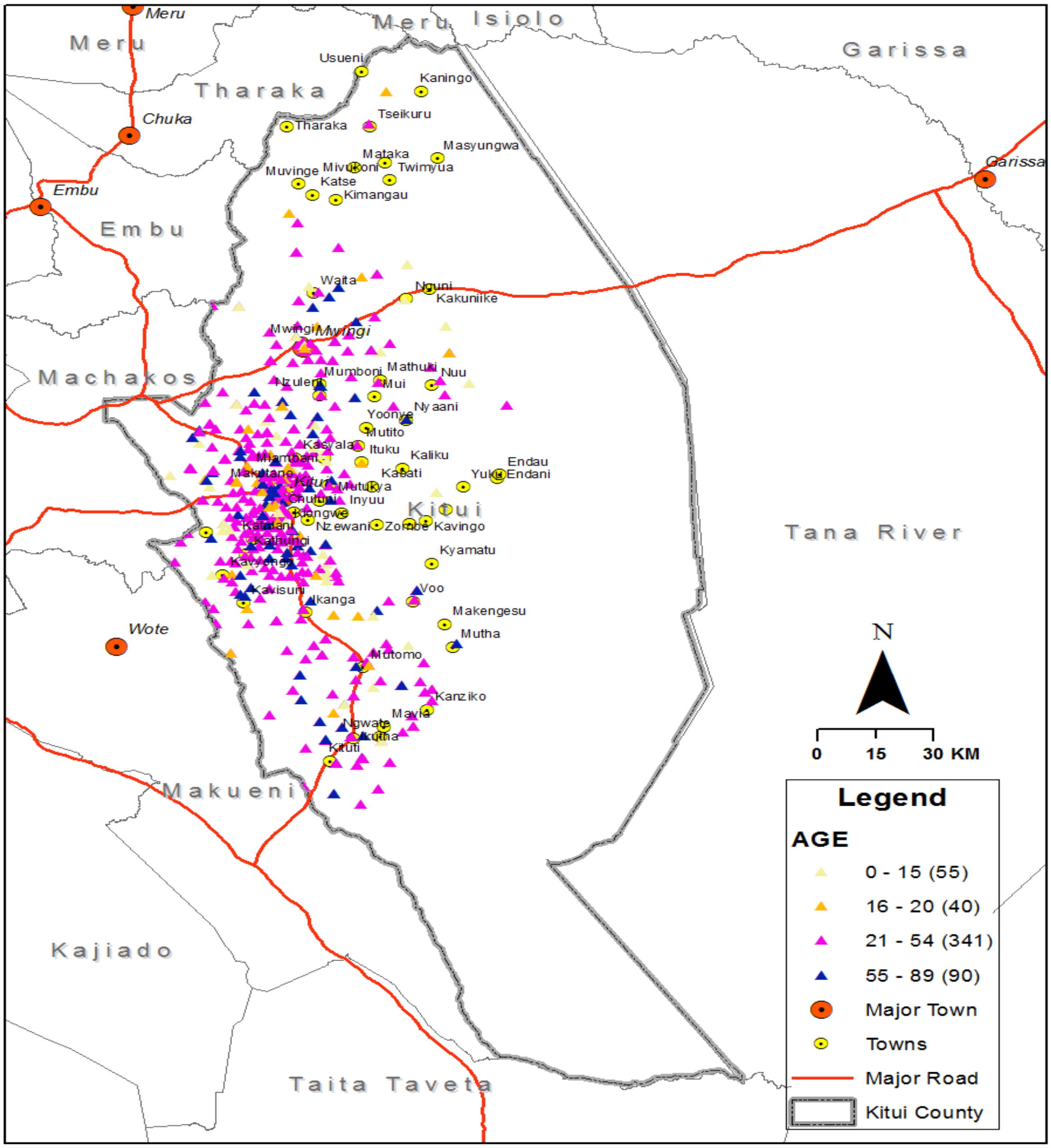
Figure 5.11: Total Aggregate of HIV and TB Patients per Facility



Source: Author, 2013

The data was further queries and it showed that the age group mostly affected by the TB scourge is the age between 21 and 55 at 55 % as depicted in table 5.1. Figure 5.11 shows the most productive age in terms of human resource and reproduction not only for the county but the whole nation. With this statistics is an indicator that the human capital is under threat and that measures should be put in place to educate these groups on the early indicators of TB infection and made aware of the diagnostic centers nearest to them for early diagnostics and treatment.

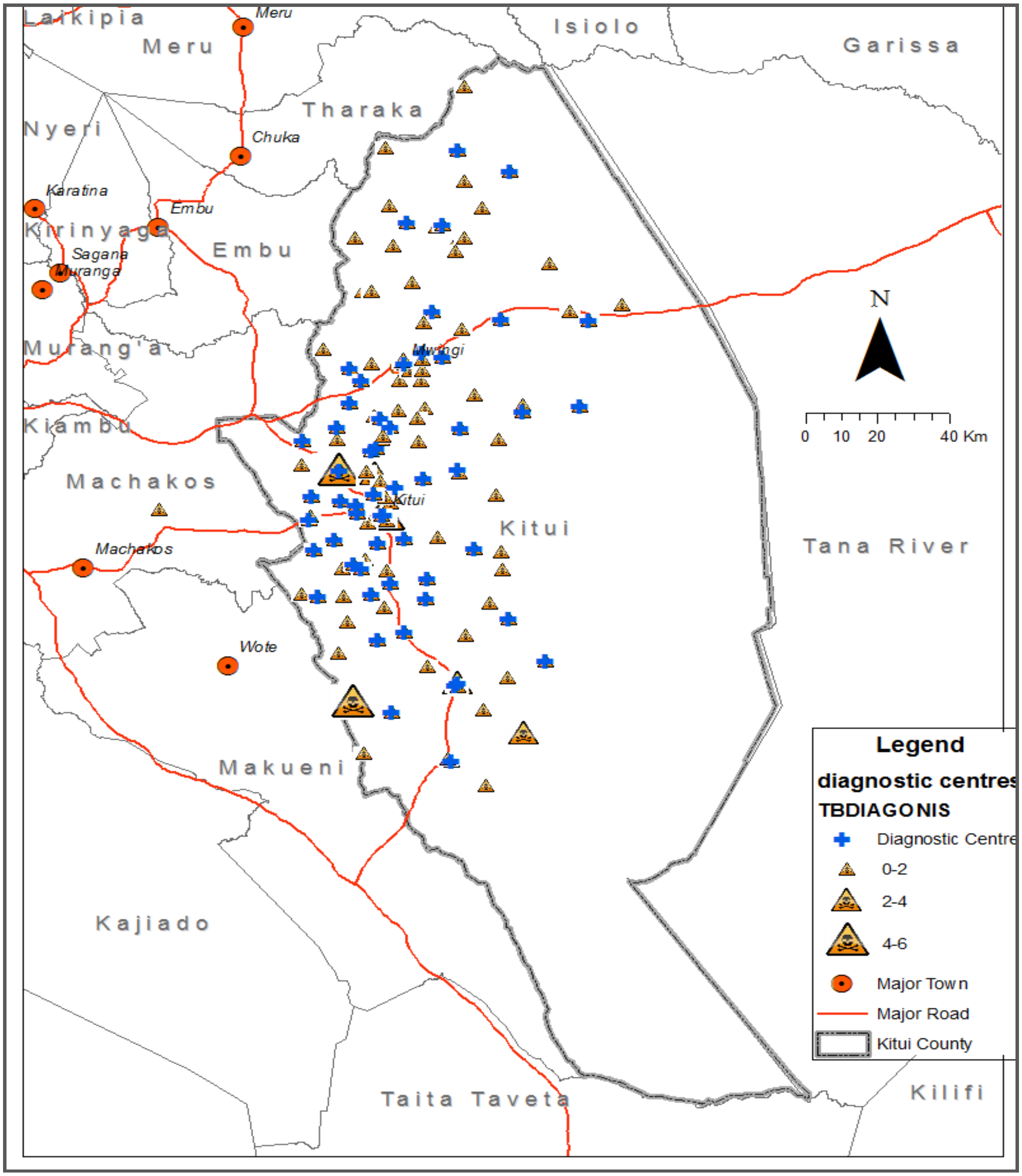
Figure 5.12: TB Patients by Age



Source: Author, 2013

Most of the deaths are not only evident along major towns, along and roads as depicted in figure 5.10 but also in areas with poor health facility coverage as in figure 5.12, with an overall average death rate of 6.4% as in table 5.1.

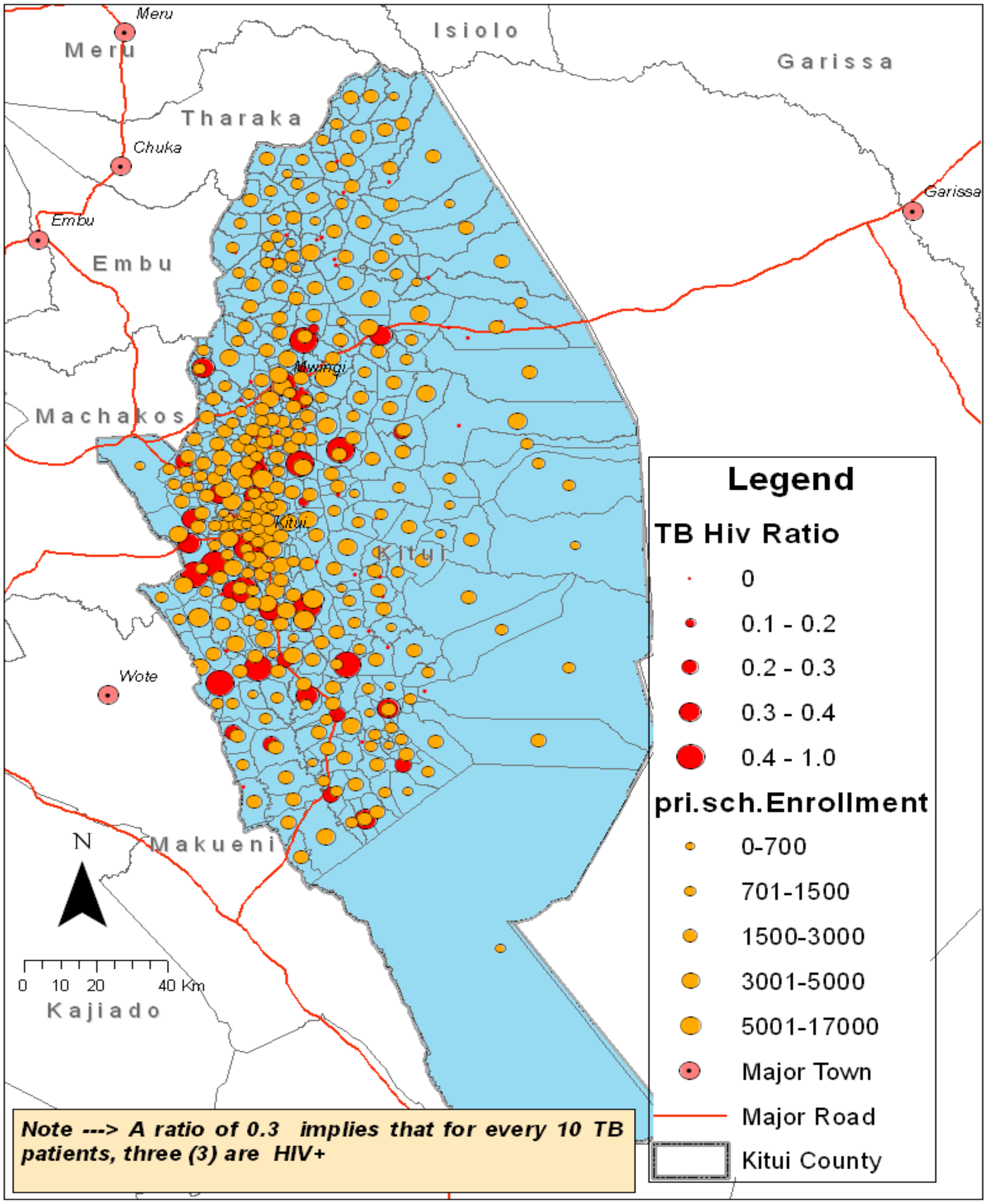
Figure 5.13: Aggregated Deaths and Diagnostic Facility Mapping



Source: Author, 2013

Some areas reported unusual patterns in terms of TB infection, deaths and HIV infection, hence high deaths. Further, correlation with enrollment in primary school data showed those areas with poor enrolment reported high infection rate, that is, for every 10 patients with TB 4patients were HIV positive as in figure 5.13 below.

Figure 5.14: Correlated Mapping between TB/HIV Infection and the Education Enrollment



Source: Author, 2013

Test Results

According to the test plan described during the design phase that is maps and the reports, the only remaining part is testing for the whole application. This is because only a vertical prototype has been developed. This can be done on the next phase at TBMU centers.

CHAPTER SIX

DEPLOYMENT STRATEGY

6.0 Roll out Strategy

As earlier indicated in the analysis, there is no Geographic Information System in the TBMU in the county. The system was to be rolled out in pilot method where by it would be implemented in some TBMUs, then extend to other locations after determining how it works after any issues arising are addressed. Direct conversion of files will be employed for the whole system.

6.1 Change Management

The aim of chain management is to coordinate the coming changes across the entire organization and do so when they occur as well as address how this new system would ensure that the new changes are positive.

This includes analyzing the current situation, that is, what is desired and then showing what will happen if the current situation doesn't change. The new system will ease the way reports are produced and also in keeping the centers' data. Therefore, adoption to the new system would not require huge effort and the benefits of transition are much more than the cost of it.

6.2 Data Migration

Data migration plan shows what activities should be performed, when and by whom before commencing operations. It includes technical and organizational aspects. These comprise the need for revising management policies, assessing cost and benefits, motivating adoption and conducting training from an organizational point of view. On the other hand, installing hardware and software and converting existing data from the technical side.

6.2.1 Technical Aspects

- ***Install Hardware:*** All hardware that is required for the new system has already been identified in the analysis and design phase.
- ***Install Software:*** All required software to be installed for the new system has already been identified

in the analysis and design phase.

- ***Data Conversion:*** Since there is no current system established, all existing data is in hard copy format, therefore, effort was needed for data entry to be done in order to convert all center's information in to an electronic format. Hiring new staff as part time staffs for data entry was required at the first stages after installation of the new system.

6.2.2 Organizational Aspects

- ***Revise Management Policies:*** In this case there was slight need to change or revise the management policies. The quarterly reports that each center must send to the central unit were not required because using the new system would eliminate this need since the staff in the central unit were now able to generate reports by accessing it online. Moreover, management policies should support the new adoption using measurement, rewards and resource allocation.
- ***Assess Costs and Benefits:*** An economic feasibility study needs to be performed in order to assess and measure the cost of using the new system and compare it with expected benefits to clearly identify whether it would be useful to use it or not.
- ***Motivate Adoption:*** Work on increasing the trust of its success among people in the organization so as to motivate them to use the new developed system. Publicize the migration plan and support the change agents. Another important point is to start differentiating between ready adopters, reluctant adopters, and resistant adopters to know the right action to be taken and with whom.
- ***Conduct Training:*** A training session was conducted to demonstrate to the staff how to use and interact with the new system and what their roles and responsibilities would be. Computer based training is the most suitable training method for this system although it costs a lot more but in the long run it is more effective and efficient.

6.3 Business Continuity Plan

Post the implementation of the new system, a support and help desk will be needed to some extend. This support can be achieved using different method such as providing an online support or on demand training in addition to considering all possible problems that might occur and identifying the suitable response to them either by implementing FAQ, getting direct help or requesting for change in case of system incorrectness.

6.4 Expected Organizational Benefits

The newly developed system was expected to improve the process of making decisions in health planning and management more effective. The improvements would be achieved by capabilities of the new system to produce very useful reports in visual formats.

CHAPTER SEVEN

DISCUSSION AND CONCLUSION

7.0 Achievements

After achieving the main purpose of this study which has been done through studying relevant literature, analyzing different similar work and models done by different people in the world, collecting data, modeling the data, analyzing and visualizing it, a number of findings were identified as follows:

- There exists significant evidence of TB cases concentration along major towns and roads in Kitui County;
- The study revealed high prevalence of TB in highly populated areas than areas with a low population density;
- High death was not only revealed in highly populated areas but also in areas with poor facility coverage and poor road networks;
- The study revealed a correlation between the rate of TB infection and HIV and further, there is high evidence of correlation between high death rate (8%) and rate of HIV, 29% of all TB patients who are HIV positive;
- Areas with a low education enrolment rate reported high TB and HIV prevalence at the high ratio. For every 10 TB case reported 4 of them are HIV positive;
- Productive ages of between 20-55 reported the highest percentage of TB infection at 65% of all cases reported;
- The research reveals evidence of lack of criteria in allocation of TB testing and treatment centers;
- There is a need to employ GIS and emphasize on its usage to enhance decision making in the county.

7.1 Constrains and Challenges

Using GIS in resource constrained environment poses a number of challenges that must be overcome to realize the full potential of this technology. In particular, in the implementation of these prototypes, the following challenges were faced:

- Problem in obtaining accurate geographical location addresses;
- Difficulty in accessing medical records due to confidentiality issues and government bureaucracy;
- Unreliability and incompleteness of government records;
- Challenges posed by the vastness of the county;

- Difficulty in obtaining licensed copied of ArcGIS resulting to a problem in integrating it with external databases;
- Lack of infrastructures; such as, telecommunication and electric power that will limit the system implementation.

7.2 Recommendation

Looking at the achievements this far and the constrain faced during the implementation of the prototype in order to achieve full potential in employing GIS to contain the menace of TB in the county, the researcher recommends the following:

- All The TBMUs adopt the use of GIS in the study of the spread of TB within the areas of jurisdiction to b e able to make informed decision of the establishment of new TB testing and treatment centers and allocation of resources and services be based on indicators such as high new cases, high relapse rate, death rate and population density and not on the existence of government health facilities;
- More accurate mapping of physical addresses be based not only on primary schools, but also on other establishment such as churches, towns, secondary schools, mountains and waterholes (dams and boreholes). This will assist in getting more accurate geo-references and therefore assist in coming up with more accurate visualization;
- Recording of the patients information be done more accurately and be more readable;
- Genuine software and tools be used during the implementation in achieve more interoperability hence, a stable and reliable system;
- Infrastructure like telecommunication and power system be improved at the TBMUs. This can be achieved by embracing modern telecommunication system such as cellular telephony and green power and the use of solar panels supplemented with gasoline generators.

7.3 Research Contribution

Despite the challenges and constrains that may be faced the, the project has contributed in spatially modeling TB infection in Kitui County and showing that if GIS is fully harnessed, it can assist TB epidemiologists, decision makers, health practitioners to visualize and understand health problems and be able to respond in a more timely and appropriate manner.

7.4 Further Research

The research is still very young and there is room for further research work which can be undertaken visualizing TB and any other infectious diseases in the whole country. Further, the study can be visualized using the new online GIS visualization tools which are more powerful and give more insights.

REFERENCES

Alexander G. (2000), Marginalization and Health Geometrics, *Journal of Biomedical Informatics*, 36 (4), pp. 1-2.

Anthony C. G. and Markku, L. (2003), *GIS and Health*, European Science Foundation, London: Taylor and Francis.

Booch G. (1986), Software Engineering with Ada, 2nd edition. San Francisco: Benjamin/Cummings.

Chainey S. and Ratcliffe J. (2005), GIS *and Crime Mapping*, New Jersey: John Wiley & Sons.

Connolly, T. Begg, C. and Conn, A. (1998*), Database Systems: A Practical Approach to Design, Implementation, and Management*, Boston: Addison Wesley.

Croner, M. *et al.*, (1996), *Visualization of Disease Surveillance Data with Geostatistics*, Atlanta: Centers for Disease Control and Prevention.

DeMers M. N. (2009), *Geographical Information Systems*, New Jersey: John Wiley & Sons.

Don J. and Pandu W. (1995), *GIS for Health and the Environment: Proceedings (IDRC):* Srilanka.

Elliot P. *et al.*, (1992), *Geographical and Environmental Epidemiology*, New York: Oxford University Press.

Fortin M. J, and Dale M. R. (2005), *Spatial Analysis: A Guide for Ecologists*, UK: Cambridge University Press.

Fotheringham A. and Rogerson P. (2009), *The SAGE Handbook of Spatial Analysis*, London: SAGE Publications Inc.

Getis A. (2008), *Essays on Geography and GIS*, ArcNews Magazine.

Goodchild M. (1992), Integrating GIS and Spatial Data Analysis, *Geographic Information Systems*, 6(5), pp. 407-423.

Hightower, A. *et al.*, (1998), Geographic Information System Applied to a Malaria Field Study in Western Kenya. *The American Society of Tropical Medicine and Hygiene, 58*(3), pp. 266-272.

Howe G. (1989), *Historical Evolution of Disease Mapping in General and Specifically of Cancer Mapping*. Recent Results of Cancer, Berlin.

Kandwa R. (2009), Health *GIS and HIV/AIDS studies: Perspective and Retrospective*, 42 (4), pp. 748-755.

Lang L. (2000), *GIS for Health Organization*, ESRI Press.

Longley P. *et al.*, (2001), *Geographic Information System and Science*, New Jersey: Wiley & Sons.

Maantay J. and McLafferty S. (2011), *Geospatial Analysis of Environmental Health*, London: Springer Science.

Maher D. and Raviglione M. (2005), Global Epidemiology of Tuberculosis, *Clinics in Chest Medicine*, Vol. 26, pp.167–182.

McDonnell R.A. (1998), *Principals of Geographical Information Systems*. London: Oxford University Press.

Mesgari M. and Masoomi Z. (2008), GIS Application in Public Health as a Decision Making Support System and its Limitation in Iran. *World Applied Sciences*, 3(1), pp.73-77.

Melnick AL (2002), *Introduction to Geographic Information Systems in Public Health*, Maryland: Aspen Publishers.

Montana L. and Mishra V. (2000), *Spatial Modeling of HIV Prevalence in Kenya*, DHS Working Papers. pp. 1(7).

Murayama Y. and Thapa R. B. (2011), *Spatial Analysis and Modeling in Geographical Transformation Process*, London: Springer Science.

Nachmias C. and Nachmias D. (2004), *Research Methods in Social Sciences*, 5th Edition, London: Arnold.

Novick P. F. and May G. P. (2008), *Public Health Administration: Principles for Population-Based Management*, Massachusetts: Jones & Bartlett publishers.

Reissman S. *et al.*, (2001), Use *of Geographic Information System Technology to Aid Health Department Decision Making about Childhood Lead Poisoning Prevention Activities*

Richards T. *et al.*, (1999), *Public Health Reports*, *Centers for Disease Control and Prevention*, Atlanta: Buford Hwy.

Touray K. *et al.*, (2010), Spatial Analysis of Tuberculosis in an Urban West African Setting. *Tropical Medicine and International Health*. 15(6), pp. 664-672.

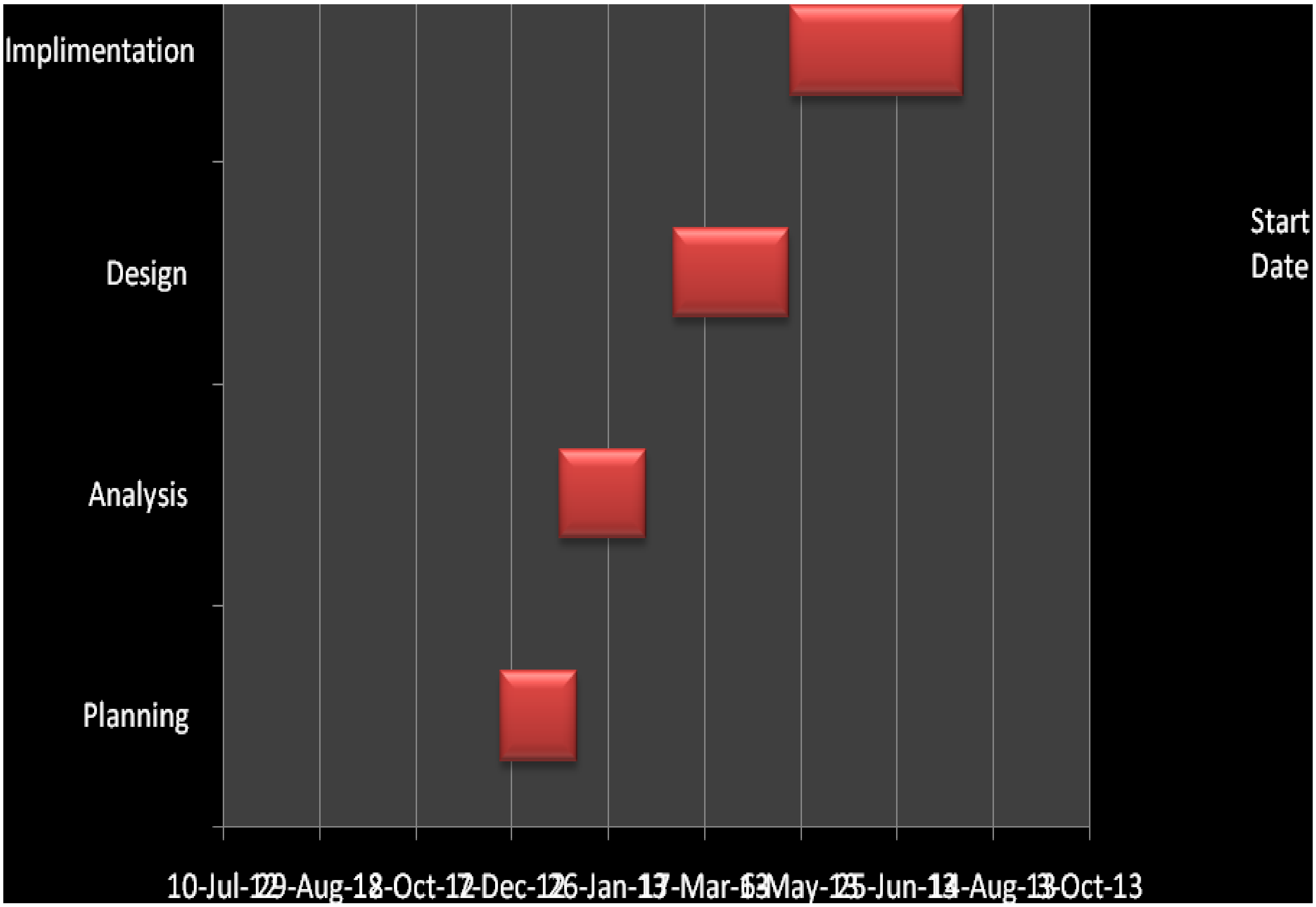
Verjee F. (2008), An *Assessment of the Utility of GIS-based Analysis to Support the Coordination of Humanitarian Assistance, UK*: Proquest Information and learning.

WHO (1997), *Treatment of Tuberculosis. Guidelines for National Programs*, Geneva.

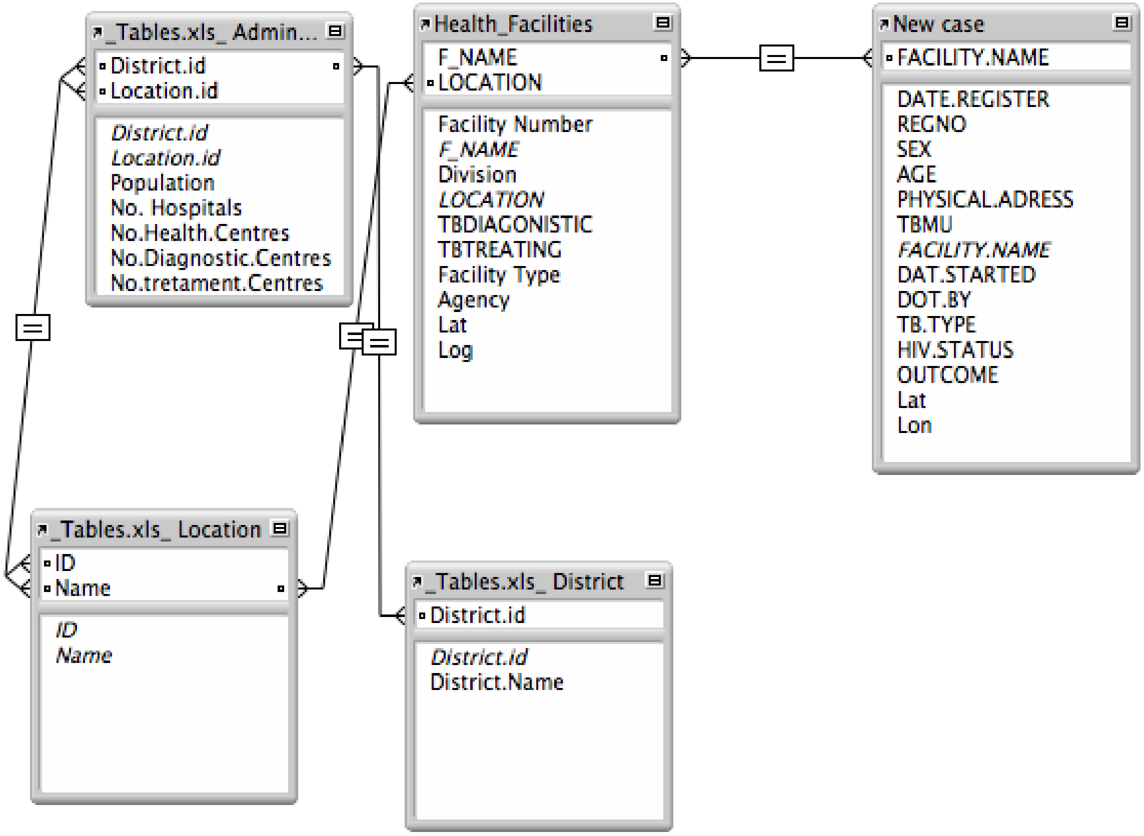
WHO (2005), *TB Emergency Declaration Issued by WHO Regional Office for Africa*
Available from:http://www.who.int/tb/features_archive/tb_emergency_declaration/en/index.html [23 October 2011].

WHO (2009), *Global Tuberculosis Control: A short Update to the 2009 Report*.
Available from:http://www.who.int/tb/publications/global_report/2009/update/tbu_9.pdf [23 October 2011].

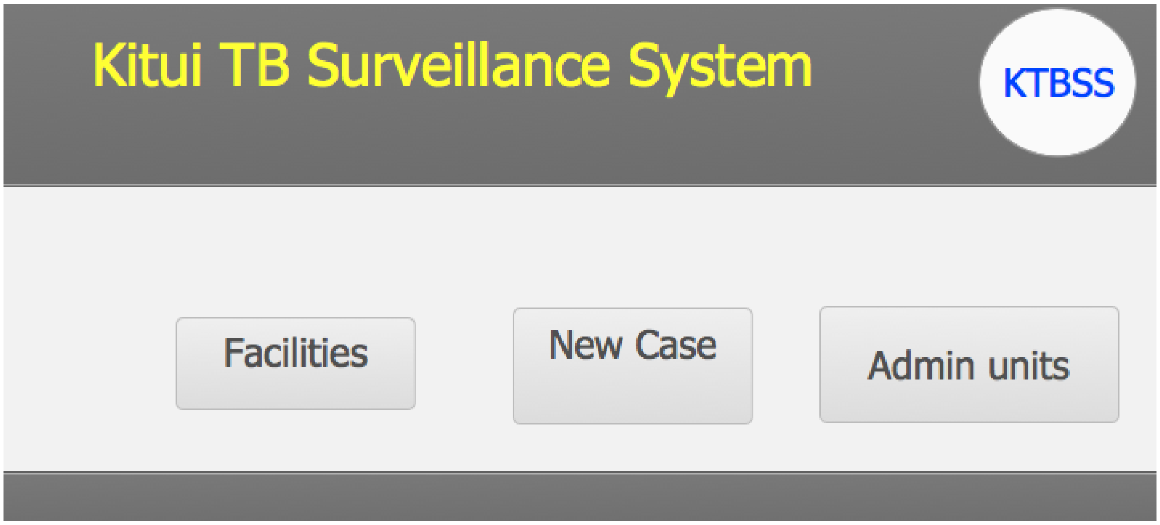
Appendix (A) Gantt Chart



Appendix (B) ERD



Appedix (C) Data Entry Screens



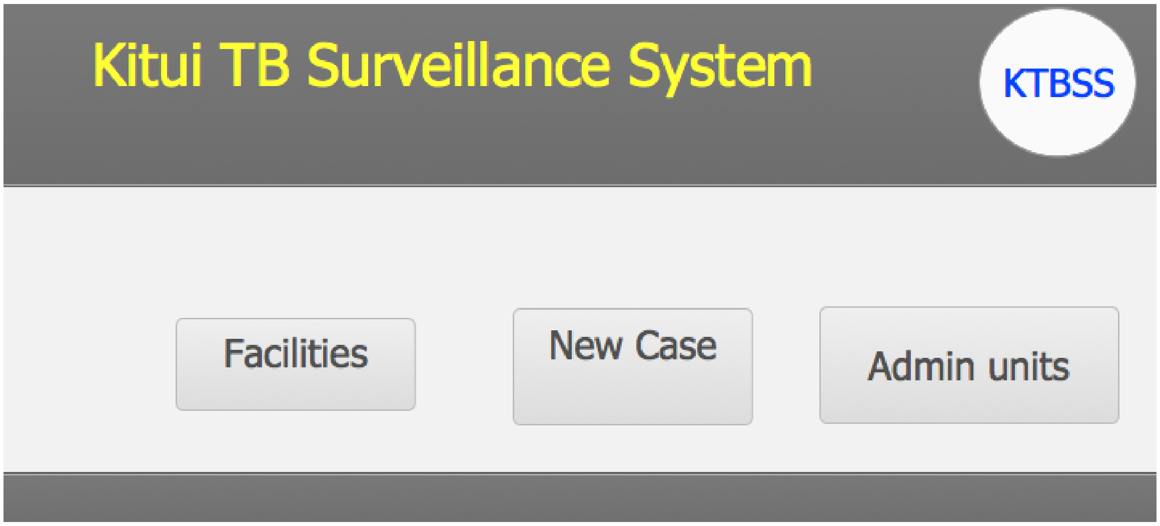
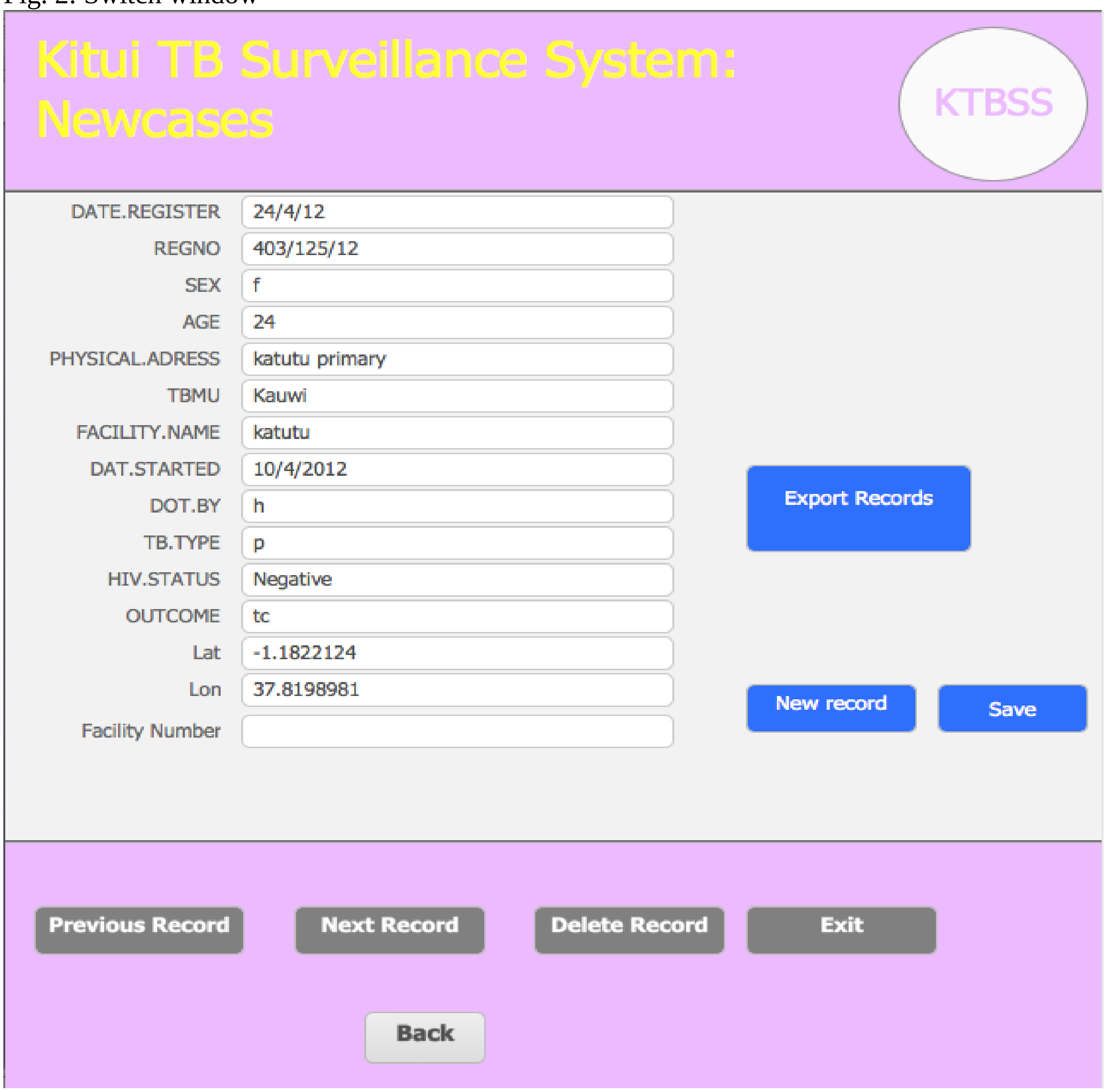


Fig. 2: Switch window




Kitui TB Surveillance System:
Health Facilities
KHTSS
Facility Number
50
F_NAME
YATTA
Division
YATTA
LOCATION
YATTA
TBDIAGONISTIC
Y
TBTREATING
Y
Facility Type
3
Agency
MOH
Lat
-1.46519
Log
37.8399
Export Records
New record
Save
Previous Record
Next Record
Delete Record
Exit
Back