UNIVERSITY OF NAIROBI

FACULTY OF SCIENCE AND TECHNOLOGY

Logistic regression modelling of factors associated with the prevalence of tuberculosis in Embakasi, Nairobi.

BY

OUMA CALVINCE ODHIAMBO

I56/39133/2021

A Thesis Submitted to the Department of Mathematics for Examination in Partial Fulfillment of the Requirements for the Award of Degree of Master of Science in Biometry of the University of Nairobi.

September,2022

Logistic regression modelling of factors associated with the prevalence of tuberculosis in Embakasi, Nairobi.

Research Report in Biometry, 2022

Ouma Calvince Odhiambo

School of Mathematics
College of Biological and Physical sciences
Chiromo, off Riverside Drive
30197-00100 Nairobi, Kenya

Master Thesis
Submitted to the School of Mathematics in partial fulfilment for a degree in Master of Science in Biometry

University of Nairobi School of Physical Sciences
Submitted to: The Graduate School, University of Nairobi, Kenya

Abstract

The research project is aimed towards establishing risk factors associated with the prevalence of tuberculosis(TB) in Nairobi Embakasi. This study was conducted on patients visiting chest clinic at Mama Lucy Kibaki Teaching and Referral Hospital which is the main health facility serving the Embakasi area of Nairobi County. Tuberculosis is a extremely infectious disease which is brought about by a bacterium called *Mycobacterium tuberculosis* which usually attacks the lungs. The bacterium also has the ability to attack other parts in the human body such as the kidney,liver,brain and the spine.The highly infectious disease is a major threat to world health, therefore early mapping of the possible cases,quarantining the individuals and offering the necessary treatments are the best and most effective ways of limiting the spread and incidences.In comparison between the developed and developing worlds, the developed world has successfully limited the prevalence in the last few decades, this has been achieved through early diagnosis and treatments as TB is highly preventable with efficient measures in place(European Respiratory Review).In most third world countries, however, the situation is quiet the opposite as most of the countries report a reemergence trend of active TB cases.This is mostly associated with the Sub Saharan region of the African continent where South Africa is also ranked as one of the leading countries with highest number of infections globally.Tuberculosis in Kenya is a major burden both to the government, societies and development partners who pump massive amount of resources so as to curtain the spread and minimize the mortality rate of the infectious disease.According to W.H.O Kenya is ranked 13^{th} amongst the the 22 countries which has about 80% of the total global cases of TB. Tuberculosis affects individuals of all age categories and anyone is at risk of getting infected. In Kenya most of the cases are reported between the age group 15-44 years which forms the economically productive part of the population. The study was conducted in Mama Lucy Kibaki Teaching and Referral Hospital which is located in Embakasi West between Umoja II and Komarock estates in Nairobi Eastland area and serve majority of the population which live in the low income estates which are characterized by poor sanitation and overcrowding. The study's main goal was to identify risk variables for forecasting the incidence of tuberculosis infection in the locality.The data records obtained at the Hospital Chest clinic formed the source of data used for this study. In the study Alcoholism and congestion were the most significant factors for the prevalence of tuberculosis in Embakasi.The prevalence rate of tuberculosis amongst the five sub-counties of Embakasi were found to be similar from the study lastly the likelihood of tuberculosis in Embakasi sub-counties was 4.661 times more likely than that of the national prevalence.

Master Thesis in Mathematics at the University of Nairobi, Kenya.
ISSN 2410-1397: Research Report in Mathematics

DISTRIBUTOR: School of Mathematics, University of Nairobi, Kenya

Declaration and Approval

This report is my original work and has not been presented previously in any institution of higher learning for the award of any academic degree or certification.

	03/07/2022
Signature	Date

Ouma Calvince Odhiambo
Reg No. I56/39133/2021

In my capacity as a supervisor of the candidate's dissertation, I certify that this dissertation has my approval for submission.

	03/07/2022
Signature	Date

Dr John M. Ndiritu
School of Mathematics,
University of Nairobi,
Box 30197, 00100 Nairobi, Kenya.
E-mail: jndiritu@uonbi.ac.ke

Dedication

I would like to dedicate this project to my whole family, Hani Mohamed and her family for their support during my post graduate studies. Also i dedicate this project to Philip Arot and his family. My close friends Naphtali Omondi,Migwi,Jesse,Cindy.To Rehema Damah and Fallis Issak your contributions were huge and highly appreciated.

Contents

List of Figures

List of Tables

Acknowledgments

First, I thank Almighty God for the wisdom granted which has enabled me reach this far with my studies, i also express my gratitude to my college based supervisor *Dr. Nelson Ndiritu* and other lecturers of *University of Nairobi, School of Mathematics* for their support and words of encouragements. Also,i would like to acknowledge my family for their support throughout the study period. Finally, i would like to appreciate the staff *Mama Lucy Kibaki Hospital* chest clinic for their corporation. ....

Ouma Calvince Odhiambo

Nairobi, 2022.

1 Chapter 1: Introduction

1.1 Background information

Mycobacterium tuberculosis causes tuberculosis. On March 24, 1882, famous scientist Robert Koch first identified the extremely contagious illness. TB attacks practically all of the body's organs, although it primarily attacks the lungs. It transmits when someone in near vicinity to a tuberculosis patient breathes in contaminated air.Tuberculosis can lay dormant for years before spreading to others or developing signs [Nambiar et al.(2016)Nambiar, Daza, an

TB manifests itself within two phases. First someone vulnerable who is subjected to a TB patient gets infected and, depending on several variables, may develop the disease later. A situation that disrupts the balance involving immune responses and *Mycobacterium tuberculosis* in infected persons thereby increasing the likelihood of developing tuberculosis. The factors that determine this balance have been studied for quite some duration. These causes were discovered that they could be "endogenous" or "exogenous" to recipients [Cress et al.(2017)Cress, Leitz, Kim, Amore, Suzuki, Linhardt, and Koffas]. The majority of research have looked at both host-related and ecological TB risk variables independently. Different designs were employed in the research. As a result determining the relative influence of these variables difficult [Falzon et al.(2011)Falzon, Jaramillo, Schünemann, Arentz, E
Furthermore, there have been a number of research conducted in countries with limited resources [Broekmans et al.(2002)Broekmans, Migliori, Rieder, Lees, Ruutu, Loddenkemper, and Ravigli
The resurgence of tuberculosis infection, which primarily affects poor nations, necessitated research into underlying risk factors for forecasting active tb incidence so as to enhance tuberculosis control measures.

In poor countries, tuberculosis, and other illnesses such as malaria, cholera has produced health and financial issues. Identifying and managing the active cases is the most important step in limiting the disease's incidence. In the last decade, tuberculosis incidence has decreased dramatically in majority of the developed countries. This has resulted in a realistic objective of eliminating tuberculosis (TB) by successful therapy of present tuberculosis cases and care of dormant tuberculosis infection to prevent the disease from developing. This is not the situation in underdeveloped nations, where tuberculosis is a major cause of death as a result of large and growing number of tuberculosis cases, particularly in sub-Saharan Africa [Mukadi et al.(2001)Mukadi, Maher, and Harries]. Kenya is ranked 13^{th} from the list of 22 countries that combined account for roughly 80% of global tuberculosis cases. According to a research, the reported incidence of tuberculosis new infections increased from 61 to 327 per 100,000 people between 1987 and 2020.

[Cassidy et al.(2009)Cassidy, Hedberg, Saulson, McNelly, and Winthrop]. The illness too was discovered to be causing a public health concern among residents of Nairobi Embakasi, Nairobi county. The research seek to discover some of the causes of tuberculosis cases in Embakasi as well as the rate of spread of tuberculosis in the area by use of data obtained from patient's record at Mama Lucy Kibaki Hospital. Study outcome of this research will help the stakeholders involved to raise tuberculosis awareness to the general public in the area of study and beyond as well as provide appropriate health care support and programs. Also the national government through the ministry of health can allocate resources needed to counter the situation. As a result, the period from exposure and detection and treatment will be significantly reduced, limiting the spread of tuberculosis.

1.2 Problem Statement

StopTBpartnership,Kenya 2020 report showed that tuberculosis as the number 4 cause of death among communicable,maternal,neonatal and nutritional diseases. The prevalence of TB in Embakasi is not well established. The high population in Nairobi has largely been the significant factor associated with the prevalence of TB in the area. Logistic regression model will be used in this case to establish the relationship between congestion,level of education,nutrition,smoking,alcoholism,ventilation in the houses among other factors and the prevalence of tuberculosis.The research study therefore seek to establish a logistic regression modelling of factors associated with a TB prevalence in Embakasi.

1.3 Research objectives

This is the set of the outcomes we expect to gain from the study.The objectives for our research are outlined in the subsequent subsections.

1.3.1 General objectives

The general objective of this study is to investigate the predisposing factors for the prevalence of active tuberculosis cases in Nairobi Embakasi.

1.3.2 Specific objectives

i. To find out main risk factor associated with pulmonary tuberculosis occurrence in Embakasi, Nairobi County.

ii. To determine the frequency and prevalence of pulmonary tuberculosis in Embakasi, Nairobi County.

iii. Compare the rate of spread of active tuberculosis cases amongst the Embakasi subcounties.

iv. To compare the average rate of tuberculosis cases in Embakasi, to the rate of tuberculosis spread nationally.

1.4 Research questions

To obtain the research goals, the research will consider the following question.

i. In Nairobi Embakasi, which factor could be most responsible for the spread of tuberculosis?

ii. What percentage does Nairobi Embakasi contribute to the county and national TB prevalence?

iii. Which of the Nairobi Embakasi sub-counties has the highest prevalence of tuberculosis?

iv. What is the most effective way to minimize the spread and infection rate of the TB infection in Nairobi Embakasi?

v. Is there a significant difference in the rate of tuberculosis infections between the males and females in Embakasi.

1.5 Hypothesis

A statistical testing will be used to compare between the different factors which contribute to the prevalence of active tuberculosis cases in Nairobi Embakasi.

i. For the first research question, the null hypothesis;
H_0 : In Nairobi Embakasi,there is no major risk-variable/factor for predicting cases of active TB.
The alternative hypothesis;
H_1 : In Nairobi Embakasi, there is a major risk variable/factor for predicting cases of active TB.
$H_0 : \mu_i = \mu_j$.
$H_1 : \mu_i \neq \mu_j$.

ii. For the second research question, the null hypothesis;
H_0 : The rate of tuberculosis prevalence in Nairobi Embakasi is not statistically different from the national prevalence.
The alternative hypothesis;
H_1 : The rate of active tuberculosis prevalence in Nairobi Embakasi is statistically

different from the national prevalence rate.
$H_0 : \mu_i = \mu_j$.
$H_1 : \mu_i \neq \mu_j$.

iii. For the third research question, the null hypothesis;
H_0 : The rate of active tuberculosis prevalence in Nairobi Embakasi sub-counties is similar .
The alternative hypothesis;
H_1 : The rate of active tuberculosis prevalence in Nairobi Embakasi sub-counties is not the same.
$H_0 : \mu_i = \mu_j$.
$H_1 : \mu_i \neq \mu_j$.

iv. For the fourth research question, the null hypothesis;
H_0 : The rate of active tuberculosis prevalence between the males and females is similar .
The alternative hypothesis;
H_1 : The rate of active tuberculosis prevalence between the males and the females is not the same.
$H_0 : \mu_i = \mu_j$.
$H_1 : \mu_i \neq \mu_j$

1.6 The study significance

Through the study individuals or residents of Nairobi Embakasi will be informed about the key predictive variables for T.B infection in the region hence increasing the community TB knowledge. The county administration will compare the prevalence of tuberculosis in Nairobi Embakasi to the national prevalence. The outcomes of the study will also help the government, doctors, and community health workers raise tuberculosis knowledge in the area and provide required health care services and appropriate resources. As a result, time between infection, detection and treatment will be significantly be reduced, limiting the spread of tuberculosis.

1.7 Study justification

The outcomes of the study will give a picture of tuberculosis prevalence during the study period, as well as its impact on country's tuberculosis situation. A statistical modelling linking the risk factors and the TB prevalence in Embakasi was also necessary as this may enhance the TB case detection.The findings provide an opportunity to critically re-engineer TB control strategies that provide a robust response towards the detection and treatment of all TB cases placing Kenya on the road towards achieving the vision of ending TB by 2030.

1.8 Limitations of the study

The study uses secondary data and therefore the the data may not be a well representation of the whole population. Also due to the small sample size, the results obtained may not give accurate situation o the tuberculosis prevalence in the region.

1.9 Chapter Overview

The need to conduct the research was clearly outlined in this chapter as tuberculosis endemic in the Embakasi region has not fully being explored. The research questions and objectives will help in giving the research a defined blue print which aim towards establishing the underlying factors which are associated with prevalence of tuberculosis in Embakasi.

2 Chapter 2: Literature Review

2.1 Introduction

Tuberculosis which is caused by a bacterium *Mycobacterium tuberculosis* is a disease that is extremely contagious. It is widespread and, in many cases, fatal. Mycobacterium tuberculosis is the bacteria that causes tuberculosis [Kumar et al.(2007)Kumar, Toledo, and Patel]. TB can affect any region of the body, however it is most commonly found in the lungs. It's an airborne disease spread by people who have active tuberculosis and disseminate respiratory fluids through coughing or sneezing. Latent TB is a type of tuberculosis that has no symptoms. If left untreated, latent infections can progress into the active tuberculosis infection, this may result into mortality of affected people [Konstantinos(2010)].

Tuberculosis that primarily impacts the respiratory system is referred to as pulmonary tuberculosis. In the case that the tuberculosis affects any other part of a human body and not the lungs, it is referred to as the extra - pulmonary tuberculosis. Appetite loss, loss of weight, fever, excessive sweating, tiredness, and chills are all symptoms of the flu [Mandell(2010)]. Pulmonary tuberculosis accounts for 90% of active tuberculosis infections. PTB symptoms include a persistent cough that produces sputum for a period exceeding fourteen days, chest pain, and, modest amounts of blood, in some circumstances Infection with pulmonary tuberculosis can progress to other parts of the circulatory system of the body such as the pulmonary artery, resulting in serious bleeding [Halezeroğlu and Okur(2014)]. Extra-pulmonary tuberculosis develops in 15-20% of active cases. Extra PTB primarily affects young children and immuno-compromised people. HIV-positive people account for half of the cases. The genitourinary system,, pleura, joints,bones, lymphatic and the central nervous system are the most commonly affected regions of the body with extra-pulmonary tuberculosis [Jindal(2011)]. The active PTB, therefore formed the primary basis for this research,that is, tuberculosis disease and the associated health dangers,epidemiology within the linked population.Notably,there is need for additional studies relating to extra-pulmonary tuberculosis.

2.2 Theoretical Literature Review

2.2.1 Tuberculosis infection risk factors

People are more susceptible to tuberculosis infections due to a variety of circumstances. According to [Organization et al.(2011)], HIV is the leading cause of tuberculosis infection worldwide, with 13 percent of tuberculosis infections also testing positive for HIV. As stated by [Gibson et al.(2008)Gibson, Abramson, Wood-Baker, Volmink, Hensley, and Costabel], HIV infection rates in the African region south of the Sahara are quite high. Persons with HIV get active tuberculosis in 30% of cases, while people without HIV acquire active tuberculosis in 5-10% of cases. Additional determinants for tb infection include poor nutrition and congestion, according to [Lawn and Zumla(2011)]. TB is regarded as the primary cause of poverty. At the community level, this study looked into the level by which hunger as well as congestion issues predispose people to infection of TB. The influence of HIV on tuberculosis was not investigated in this study.

People who may be more vulnerable to tuberculosis infection are children having close proximity or contact with tuberculosis infected persons,health care workers in tuberculosis wards or clinics,marginalized communities, those in low-income residential areas,prisoners and staff in those facilities,people living in refugee camps and the employees in these areas, passengers in crowded public transport system and markets are also considered to be high at risks. TB infection is further exacerbated by associated illnesses such as *Cor pulmonale* and silicosis. Other factors, such as alcoholism and diabetes, can increase the risk of developing tuberculosis [Society et al.(2000)]. Cigarette smoking is also a health hazard associated with tuberculosis incidences, smokers have a double percent change risk of getting TB infection compared to non-smokers. According to [Lawn and Zumla(2011)] TB infection has a hereditary component whose overall value has not been determined.

This study looked into the prevalence of tuberculosis infection in children aged 0 to 14, as[van Zyl Smit et al.(2010)van Zyl Smit, Pai, Yew, Leung, Zumla, Bateman, and Dheda] well as adults aged 15 and older, as well as the role played by smoking cigarette and drinking in pulmonary tuberculosis cases in a low-income communities. population or socio-demographic characteristics are also risk factors for tuberculosis cases.Example, males are more prone to tuberculosis infection than their female counterparts, particularly among the unmarried category With reference to a West African research report,this is the case. The male and female genders were considered in this study to assess the risk group for tuberculosis infection in a resource-poor society. Housing circumstances can also play a role in TB infection. A study in West Africa found that TB infection was higher in mud-walled dwellings with filthy floors than in concrete houses [Lienhardt(2001)] TB infection can be exacerbated by poor ventilation as stated by the report of the West African research which was conducted in health care centres across the region . This research looked into some of the impacts poor air circulation in both semi-permanent and congested story apartments plays in the spread of tuberculosis in the study region.

2.2.2 Transmission of tuberculosis

TB is spread by people who have active pulmonary TB. This occurs when people sneeze, spit, sing, speak, or cough, releasing infectious aerosol droplets. Because the dose for tuberculosis infection is so small, a single cough or sneeze can produce up to 40,000 droplets, each of which can spread the disease. Tuberculosis transmission is influenced by several factors, including the number of infectious droplets expelled by TB active individuals (as opposed to latent TB individuals, who are not contagious), the duration of exposure, the house's ventilation, the virulency of the *Mycobacterium tuberculosis* infection, as well as the infected individual's level of response to the strain. The disease takes less than a month for a fresh infection to become contagious so as to spread from the host to close contacts. The level of contact and duration spent interacting with an active case or persons infected are the most important factors for TB spread. Tuberculosis is spread through prolonged contact with an infectious individual in enclosed surroundings. People who have had extended, regular, or close contact with tuberculosis are especially vulnerable to infection [Gounder et al.(2015)Gounder, Zulz, Desai, Stenz, Rudolph, and Raymond]. A positive tuberculosis patient can spread the disease to up to 15 individuals everyday [Ragonnet et al.(2019)Ragonnet, Trauer, Geard, Scott, and McBryde]. Those individuals residing in the same house as an infected individual and also sleep together are at the greatest risk of contracting tuberculosis [Singh et al.(2005)Singh, Mynak, and Kumar]. People who live in the same residence as someone who has TB are at risk of contracting the disease. Overnight cough has been linked to increased tuberculosis spread between individuals living in the same house. The presence of an infected person in the home was not investigated as a TB risk factor in this study.

2.2.3 Tuberculosis diagnosis

TB is diagnosed by identifying *Mycobacterium tuberculosis* in a clinical sample of sputum. Sputum samples are collected for evaluation at least three times in the morning. The bacterium can also be identified through blood or sputum culture, which requires about 2 months because the bacteria in the organs grow at a slower rate. Treatment-care is therefore frequently started even without the presence of the pathogens cultures have been identified [Pai et al.(2008)Pai, Zwerling, and Menzies]. Results of samples obtained from the sputum diagnosis were evaluated in the investigation.

2.2.4 Tuberculosis prevention

The most effective way of tuberculosis prevention and control is newborn vaccination. The Bacillus Calmette-Guerin (BCG) vaccine is employed, and it reduces the chance of getting infected by approximately 20% also the further growth of tuberculosis disease by 60%. After ten years, the immune response that has been created generally begins to wane. Early detection and adequate treatment are the most effective ways to prevent TB from spreading in active cases (Lawn et al., 2011). The effectiveness of the Bacille Calmette-Guérin (BCG) immunization in the vaccinated infants were determined.

2.2.5 Pulmonary tuberculosis treatment.

Antibacterials can be used to effectively treat tuberculosis. Many antibiotics used to treat tuberculosis are ineffective because of the unique structure and chemical makeup of Mycobacterium tuberculosis' cell wall, which inhibits medicines from passing to fight the bacteria. Antibiotics like isonicotinic acid hydrazide and rifampin are commonly used to treat tuberculosis. Treatment usually lasts a few months. Only one drug is used to treat latent tuberculosis. A mixture of antibiotics is used to treat active tuberculosis illness. This lowers the chances of the bacteria developing antibiotic resistance (CDC, 2011). Directly observed therapy is recommended by WHO (DOT). This is when a health care practitioner observes a patient taking medication.This reduces the percentage of individuals who fail to take their antibacterium medications as prescribed [Mainous III and Pomeroy(2001)]. DOT appears to be ineffective in reminding people of the significance of treatment [Muture et al.(2011)Muture, Keraka, Kimuu, Kabiru, and Ombeka.]. The world health organizing body recommends that first time TB be treated for six months. Isoniazid, rifampicin, ethambulol, and other antibiotics are used in this treatment.pyrazimamide. In the event of Isoniazid resistance, Ethambulol can be administered as an alternative in the last four to three months of treatment [Organization et al.(2009)]. If TB infection reoccurs after treatment, another test is done so as to discover the specific antibiotic the patient seem resistant to before therapy begins. If someone has multidrug-resistant tuberculosis, they should be treated for 18 to 24 months. At least four efficient antibiotics are used in this treatment [Matteelli et al.(2014)Matteelli, Roggi, and Carvalho]. Therapy-resistant tuberculosis can be a severe society health problem in many developing nations, as the management process takes a long time and costs a lot of resources. Being resistant to the two effective first-line TB medications, rifampicin and isoniazid, is characterized as being a multi-drug resistant TB i.e. MDR-TB. EDR-TB is the tuberculosis type that is extensively drug-resistant to three or more of the six second-line medication categories of medications. Incomplete treatment, incorrect usage of prescription drugs, the use of drugs that are of compromised quality all contribute to this. The widespread presence of MDR-TB and medicine resistance TB was not investigated in this study.

2.2.6 Overview of tuberculosis burden

Global tuberculosis situation

About a third of the global human population has been infected with Mycobacterium tuberculosis [Organization et al.(2014)]. Every year, about 1% of the population contracts a new infection. A survey conducted by the Centers for Disease Control and Prevention (CDC), most *Mycobacterium* tuberculosis infections do not result in tuberculosis disease infection, and 88 to 93 percent of infections are immuno-suppressed or asymptomatic. In 2012, 9.6 million pulmonary active tuberculosis incidences were reported [Ahmed(2018)]. According to a 2010 world health organization an estimate of over 8 million new cases of tuberculosis were detected, with 1.3-1.5 million fatalities. The majority of these tuberculosis is the second most frequent infectious disease that causes death after HIV/AIDS. Since 2005, the prevalence of tuberculosis has been reducing, and TB incidence has been decreasing since 2002. (WHO, 2011). Between 1990 and 2010, China's tuberculosis mortality rate fell by almost 75%, and cases of new infections fell by 20% between the period 2010 and 2015. Tuberculosis is more prevalent in developing countries than developed countries.

This is demonstrated by the fact that around 79% of the people in several Asian and African nations gives positive results for tuberculosis tests, while just 6-11% of the US population tests positive. Due to reasons such as the prolonged and expensive diagnostic stages, challenges encountered when developing a new reliable vaccine , requirements of long periods of treatment,occurrence of tuberculosis cases which were resistant to the drugs during the 1980s and also increase in tuberculosis cases as a result of HIV infection, efforts to completely control tuberculosis disease have been dramatically depressed (Lawn et al.,2011).

In 2018, Switzerland was leading globally with the highest tuberculosis incidence rate, with 1,200/ 100,000. India had about 2,000,000 new reported cases (WHO, 2020).However,tuberculosis incidences is highly reduced in developed world, and most of the cases are reported from the in metropolitan regions. In 2021, the global rate of tuberculosis spread per 100,000 persons was 321, the rate in the American region was 46, African region was at 890, East Mediterranean region was 526, South-East Asian region was 749, European region was at 113, and Western-Pacific region was at 241. (WHO, 2020).Tuberculosis rates have dropped considerably in other nations, such as the United States. TB primarily affects the immuno-compromised and the elderly in many locations. In general, the 22 countries with the highest TB burden account for 79% of active infections and over 80% of fatalities worldwide [Harries et al.(2018)Harries, Lin, Kumar, Satyanarayana, and Takarinda].

Overview of TB spread in Kenya.

Kenya presently is positioned 15^{th} among the 22 nations with a high TB burden. According to the WHO Global Tuberculosis Report from 2016, this is the case. According to the data, TB cases jumped from 11,625 in 1990 to 169,065 in 2016. The average yearly rise rate for all kinds of TB during the last ten years was 7%. In 2018, the case notification rate for all forms of tuberculosis went from 178 out of 100,000 to 280 for every 100,000 persons, and for sputum- positive pulmonary TB cases climbed from 46 for every 100,000 to 98 for every 100,000 cases. In 2020, the number of PTB cases with a positive sputum smear increased by 4% over 2008.In 2009, both males and females aged 25-34 had the highest TB notification rate and the highest HIV prevalence. In the last ten years, this has been the pattern. After the age of 24, males begin to outnumber females in terms of TB notification [Wangara et al.(2019)Wangara, Kipruto, Ngesa, Kayima, Masini, Sitienei, and Ngari].

In 2020, 117,083 instances of all types of tuberculosis were reported, this was a 4 percent drop from the 119,365 cases reported in 2019. Increased incidence of HIV in the general population, estimated at 7.1 percent, has significantly contributed to the increasing health burden as a result of tuberculosis in Kenya. From the ministry of health data for 2020, 41 percent of TB patients were HIV-positive. A total of 113 MDR(multi drug-resistant) tuberculosis incidences were reported to WHO in 2019. and at the end of the year 2020, the country had put 189 individuals on treatment, out of these 81 were receiving treatment for the first time. within the calendar year. The prevalence of drug-resistant tuberculosis is estimated to climb in the next years [MBOGA(2018)]. In a total of 37,402 patients, the 2019 cohort's tuberculosis treatment success rate was 86 percent for new sputum-positive tuberculosis infections. According to the WHO 2019 Global Report, the rate of detection of all the forms of tuberculosis was at 87%.The republic of Kenya was able to meet 70 of the 85 global tuberculosis control targets in 2019. These achievements were largely associated with the improved access to health case in the country, also the improved infrastructure in the facilities such as well equipped testing labs with 2,136 AFB microscopy centers (which translates to one microscopy center for every 21,589 persons) as well as over 3,617 tuberculosis centres of treatment, as well as highly skilled health personnel,frequent review meetings, and increased community based tuberculosis awareness education. To provide every necessary care to TB patients, the private and public health sectors have also been integrated in this project (WHO, 2019). The overall number of TB cases reported in 2012 was 89,568. In 2011, 88 percent of new smear-positive and culture-positive cases were successfully treated (USAID and DNTLD,2020).According to a regional demographic survey on tuberculosis spread in North Eastern region, 9.2 percent of 400 tuberculosis-screened participants were positive for tuberculosis, revealing a high level of tuberculosis spread in the community [Mohamud(2004)].

Rural areas in Kenya,Homa-Bay county has an intermediate TB prevalence. According to a study conducted between 2005 and 2007, roughly 17% (1122) of patients attending a TB laboratory at Homa-Bay county teaching and referral hospital tested returned a sputum-positive result for tuberculosis [Emmanuel et al.(2010)Emmanuel, Shitandi, et al.].

2.3 Empirical Literature Review

Enos M.,,et al, 2016; Kenya tuberculosis prevalence survey,2016.The survey identified 305 prevalent TB cases translating to a prevalence of 558 [95%CI 455–662] per 100,000 adult population.The highest disease burden was reported among people aged 25–34 years (716 [95% CI 526–906]), males (809 [(95% CI 656–962]) and those who live in urban areas (760 [95% CI 539–981]).Compared to the reported TB notification rate for Kenya in 2016, the prevalence to notification ratio was 2.5:1.

The WHO 2016 TB report in Kenya.The report estimated that there were 169,000 cases of TB in Kenya in 2016 and 29,000 deaths (excluding deaths linked to HIV and TB co-infection), making TB the fourth-leading cause of death in the country.

StopTBpartnership,kenya 2020.TB incidences in the country was at 139,000 with 17,000 being children.The number of people who developed DR-TB was at 2,500,470 were on treatment and 346 individuals were successfully treated. The report also stated that 35,000 TB patients were also HIV co-infected. The mortality rate had increased by 5% to 21,000 in 2020, this made TB the number 4 cause of death among communicable,maternal,neonatal and nutritional diseases.

Norah K Maoke.„et al,2016: Spatial and temporal distribution of notified TB cases in Nairobi county 2012- 2016. A total of 70,505 cases of TB were notified in Nairobi County between 2012 and 2016,the males were 42,927(61%) with a male to female ratio of 3:2 and HIV co-infection rate of 38%.The median (interquartile range (IQR)) age of the cases was 32 (25–41) years.Those aged 25 to 34 years were 23,198(33%).Tuberculosis cases in children (0-5years) accounted for 2,623(4%) of the cases. Most of the cases were reported by public health facilities 43,913(62%) with the rest being reported by private20,243(29%) and prisons 1,722(2%). Kasarani and Kamukunji constituencies had the highest number of cases accounting for 7,055 and 6,923 (10%) each respectively, while Embakasi East and Dagoretti South had the least cases.

2.4 Chapter Overview

This chapter explains about some related studies to this research topic.Through this it was established that TB is the number four cause of death among communicable,maternal ,neonatal and nutritional diseases. Also,the mortality rate due to TB had increased by 5% in 2020 compared to 2017. The leading factor to the prevalence of TB in Kenya was found to be poor sanitation and also inaccessible medical care in rural areas.

3 Chapter 3: Research Methodology

3.1 Introduction

This chapter describes the model used in statistical study of tuberculosis prevalence as well as the area of the study and data to be used. When choosing the model,the nature of the variables studied were taken into account and the multiple binary logistic regression model was deemed as the best option.

3.2 Study Design and Data

The study used an analytic cross-sectional research method to collect data on the underlying factors for the spread of chronic pulmonary tuberculosis and other associated health risks. The patients visiting Mama Lucy Kibaki Teaching and Referral Hospital were the population of interest. Individuals that came to the facility between January of 2020 and March of 2022 and were exhibiting TB complaints were chosen at random for inclusion in the study. This data was obtained from the records at the hospital's health information and records department. The findings of the participants' lab sputum tests were registered in the Tuberculosis records register. Subjects with a positive phlegm test and those with a negative phlegm test were categorized into two groups so as to establish the factors for the spread of tuberculosis in Nairobi Embakasi.

Socio-demographic and economic characteristics were among the factors investigated. The gender and age to establish the especially vulnerable age cohort and gender to tuberculosis infection, level of education, profession, tuberculosis knowledge, tuberculosis medication history, type of the housing where the individuals stay and their ventilation status, alcohol addiction, diet, ease of access to a medical facility, cigarettes abuse, and overcrowding in the residential buildings were all investigated.

3.2.1 Area of study

Nairobi Embakasi can be broadly divided into 5 sub-counties namely: Embakasi East,Embakasi West,Embakasi North,Embakasi South and Embakasi Central. The study area borders; Makadara, Ruaraka, Kasarani and Mathare Sub-counties. The area also borders with Machakos County. The area has a number of sub-county health facilities with the main referral hospital being Mama Lucy Kibaki Teaching and Referral Hospital which is located in Embakasi Central sub-county. The study area is a densely populated area with most residents staying in high-rise flats. It is considered a low and middle-income region as most of the residents in the area are those who have informal employment. The major source of livelihood for the residents are small scale business and temporary employment in industries located in industrial area of Nairobi city county and Machakos county. Most of the residents speak Swahili and English languages coupled with their local ethnic dialects.

The map below shows the Nairobi city county and its sub counties. This also include our study area.

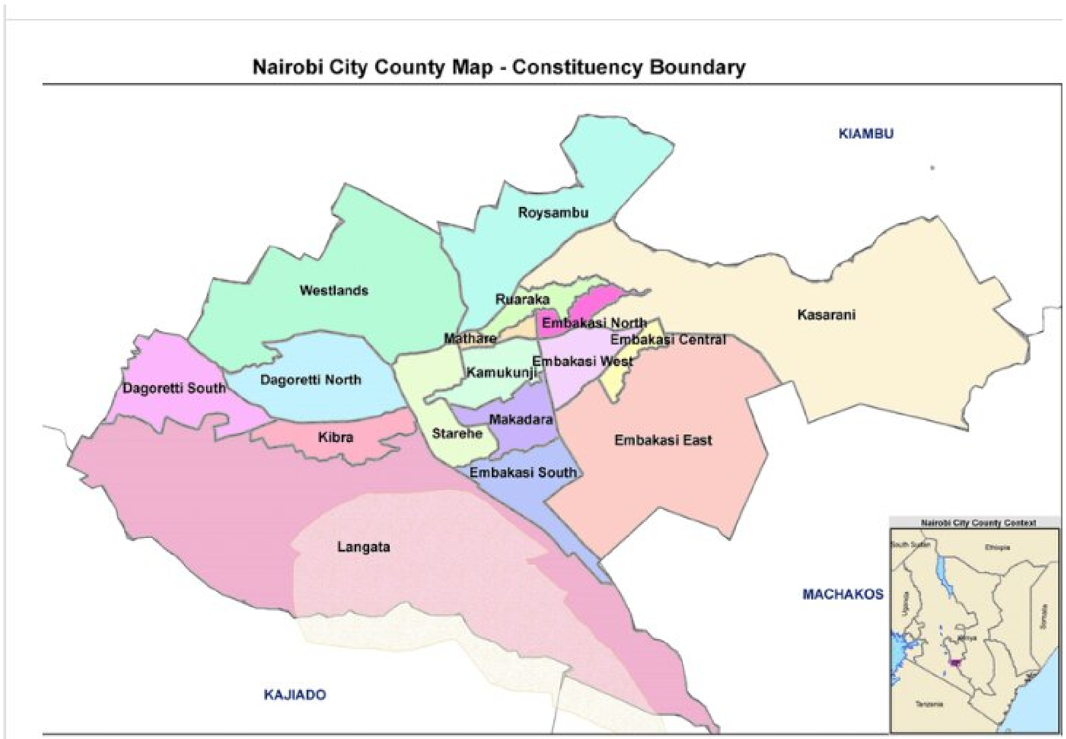

Figure 1. Nairobi map showing all the sub-counties

3.2.2 Data Collection

The data used was obtained from the hospital records of patients who visited Mama Lucy Kibaki teaching and referral hospital having symptoms of tuberculosis. According to the records, the patients were mostly tested for Covid-19, tuberculosis and common flue. For tuberculosis tests, a sputum sample was collected and analyzed in the laboratory to check for the presence of the *Mycobactrium tuberculosis* bacteria which causes the tuberculosis disease. The results of the tests were entered into the hospital's tuberculosis records. Data such as the gender, age of the patient, their physical address, duration of stay in their locality and the date of tests and the test results or diagnosis were also captured into the clinic's records. The results were recorded as either tuberculosis positive which meant that the recorded phlegm had traits of the bacteria while those which didn't have the bacteria traits were recorded as negative.

3.2.3 Sample studied

Criteria for inclusion

Those individuals who visited Mama Lucy Kibaki teaching and referral hospital tuberculosis clinic exhibiting tuberculosis symptoms were the primary target group. Individuals should have been residents of the study are for at least half a year. Also the individuals should be in the hospital's tuberculosis registers . The sample comprised of individuals in the age brackets 0-14, and those with age bracket 15-80.The research involved both female and male gender.

Sample size calculation

The study used a sample size of three hundred(300) individuals. A total of 109 were females and 191 of the respondents were males. This was the entire list of patients who had visited the chest clinic at Mama Lucy Kibaki teaching and referral hospital during the period from January,2022 to April 2022. The youngest of the subjects being 6 months old and the oldest being 80 years old.

3.2.4 Data analysis tool

Data obtained from the hospital was analyzed using R-programming statistical software. Each of the hypotheses were tested at 5% level of significance using the chi square test and the Fisher's exact test. The chi-square test was used to check if the difference between the observed data and expected data was as a result of chance or some causative relationship. The Fisher's exact test was used in analysis of a contingency table to asses if a relation does exist or does not between the various categorical variables used in our study. At the end logistic regression will be employed to asses the underlying risk factors that contributed to the spread of tuberculosis in Embakasi Nairobi and their levels of contribution to our research problem.

3.3 Emperical model

3.3.1 Logistic Regression Model

The use of logistic regression involves describing the probability of a categorical data based on some earlier observed set of data.The model makes a prediction of an outcome or expected variable by examining the association between a single or numerous existing independent variables. The binary logistic regression model is the most popular amongst the generalized linear models(GLM). In health sciences, most of the response outcomes are binary or categorical with two possible outcomes , therefore the binary logistic regression model can be useful in finding the risk or probability of developing a disease or health condition over a given duration if the factors involved are known.In this project, the outcome variables are binary in nature and therefore a logistic regression model to model the underlying risk factors for the commonness and spread of tb in Embakasi area of Nairobi.

The components of a GLM Y are:

i. A random component. the random component specifies the conditional distribution of the response Variable given the set of predictors.

ii. Systematic component.This is the quantity which incorporates the information about the predictor variables into the model;

$$\eta = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k$$

iii. Link function. This specifies a function which combines the random component and the systematic component. let E[Y/X] =μ and if g(μ) = η, then g is the link function.

$g(\mu) = (\frac{\mu}{1} - \mu)$ helps in modelling log of the odds that a given event will be observed. The function is referred to as the logit link and is most useful when $0 \leq \mu \leq 1$

3.3.2 Description of variables

Table 1. List of variables investigated

Variable	type of variable		levels of variable
Gender	numeric	2	(0= males, 1= females)
Age	cont.	7	(0-14,15-24,25-34,35-44 ,45-54,55-64s, above 65)
Education	numeric	4	(0= no formal, 1= primary,2= secondary,3= tertiary)
Employment	numeric	3	(0= unemployed,1= formal,2=informal)
Awareness	numeric	2	(0= little or no awareness,1= aware)
Ventilation	numeric	2	(0=poor,1=good)
Congestion	numeric	2	(0= uncongested,1=congested)
Alcoholism	numeric	2	(0=alcoholic,1=non-alcoholic)
Smoking	numeric	2	(0=smoker,1=non-smoker)
Nutrition	numeric	3	(0=1 meal,2=2 meals,3= 3 meals)
TB history	numeric	2	(0=no previous TB infection,1= had previous TB infection)

3.3.3 Multiple logistic regression model

The existence of additional factors may impact the effect of one variable on a dependent variables as a result of modifying the effects (Chap, 2013). It is preferable to evaluate a wide set of variables and determine those closely connected to the predictor variables (Chap, 2013). As a result, multivariate logistic regression is employed for risk assessment to give a detailed thorough analysis. A linear set of predictors are used in multiple logistic regression model. These may include; a person's age,education level, history of other illnesses etc . These are categorized as variables which are binary in nature, from here we then we pick the dummy binary variable and use it as our reference for the others. For instance, let k represent a set of predictors for a binary variable $(\mathbf{X})$ having specific values $x_1, x_2, ...x_k$. Then the multiple binary logistic regression model is given by;

$$\pi(\mathbf{X}) = \frac{exp(\beta_0 + \beta_1 x_1 + ... + \beta_k x_k)}{1 + exp(\beta_0 + \beta_1 x_1 + ... + \beta_k x_k)};$$

$$= \frac{exp(\mathbf{X}\beta)}{1 + exp(\mathbf{X}\beta)}$$

$$= \frac{1}{1 + exp(-\mathbf{X}\beta)}$$

The value π represents the probability of success i.e the binary response variable of interest is observed.

The linear form of the binary logistic regression is:

$$log[\frac{\pi(\mathbf{X})}{1-\pi(\mathbf{X})}] = \beta_0 + \sum_{i=1}^{k} \beta_i x_i :$$

where the β_0 represents the intercept and the β_i represent the partial regression coefficients.

3.3.4 The model assumptions

The assumptions made in this case include:

i No multicollinearity: when two or more of the independent variables are substantially correlated amongst each other.When multicollinearity is present, the regression coefficients and statistical significance become unstable and less trustworthy..

ii Independence:Each of the observations (data points) should be independent.i.e x_i are fixed independent variables and

iii Linearity:Logistic regression fits a logistic curve(the probability associated with each outcome across independent variable values) to binary data. Logistic regression assumes that the relationship between the natural log of these probabilities (when expressed as odds) and the predictor variable is linear.

iv No Outliers:

The multiple logistic regression model will be helpful in evaluating the type of association the variables have on the prevalence on tuberculosis in Embakasi: the gender, age of the individuals, education level of the individuals, type of employment, the awareness level of the individuals to the tuberculosis infection, level of ventilation in the houses, congestion in the residential houses, alcoholism, smoking, infection history of the individuals where individuals were categorized as those who had previous TB infection and those who have not had the infection as well as the number of meals an individual had in a day amongst the individuals who participated in the study. An interaction term amongst the significant factors will also be included in the reduced model so as to minimize the effects due to correlation of the investigated variables.

During the modelling process in logistic regression, a function may be used to convert an ordinary least squares regression equation to the logistic regression equation and vice versa. The multiple logistic regression model is given as:

$$logit(\pi) = ln(odds) = ln(\frac{\pi}{1-\pi}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_i x_i$$

In the above equation, solving for the antilog on both sides, we get the equation:

$$\pi = \frac{e^{\beta_0+\beta_1x_1+\beta_2x_2+...+\beta_ix}}{1+e^{\beta_0+\beta_1x_1+\beta_2x_2+...+\beta_ix}}$$

which is useful in determining the probability that our event of interest will be observed.

In the above equation:
π = probability of the occurrence of the event under consideration for example, the prevalence of tuberculosis is due to smoking.
$\beta_{i's}$ = regression coefficients
$x_{i's}$ = predictor variables

The equation above therefore will be used to predict the probability of high tuberculosis prevalence given the underlying predictors.The research therefore also will use a multiple regression model to predict the factors associated with the prevalence of tuberculosis in Embakasi due to the presence of many predictors The use of linear regression model would not be appropriate in this case because the outcome probabilities may be negative or more than one.

The probability of non-prevalence of tuberculosis is therefore given as: 1-$\pi(\mathbf{X})$

$$= \frac{1+e^{\beta_0+\beta_1x_1+\beta_2x_2+...+\beta_ix_i} - e^{\beta_0+\beta_1x_1+\beta_2x_2+...+\beta_ix_i}}{1+e^{\beta_0+\beta_1x_1+\beta_2x_2+...+\beta_ix_i}}$$

The odds of prevalence of tuberculosis is given as;

$$= \frac{1}{1+e^{\beta_0+\beta_1x_1+\beta_2x_2+...+\beta_ix_i}}$$

The log of the odds of tuberculosis prevalence;

$$ln(\frac{\pi(x)}{1-\pi(x)}) = \beta_0+\beta_1x_1+...+\beta_ix_i$$

3.4 Parameters estimation and Hypothesis testing

In logistic regression model the the parameter estimation i.e the $\hat{\beta}$ follows the solution to the maximum likelihood estimation.

3.4.1 Maximum likelihood estimation

The method of maximum likelihood is important when remodelling the data when given the estimates of the parameters. In the analysis data may be given the values 0 or 1 for predictors which are continuous in nature and for the categorical set of predictors the dummy variable is identified and assigned the value 0 or 1.

Testing for the hypothesis: $H_0 : \beta_{i's} = 0.$ and in the case, the null hypothesis is rejected then it shows that not all the $\beta_{i's} = 0$ and therefore the logistic regression becomes a better predictor than the mean of the response variable Y for the outcome probability.

The aim of maximum likelihood estimation in logistic regression is to identify the variables which results into the highest probability for the data observed.The equation is obtained from the distributive probability of the response variable.

Let y_i be the Bernoulli responses of a given population with sample size n, therefore the joint pdf of Y is given as:

$$L(\beta;\mathbf{y},\mathbf{X}) = \prod_{\mathbf{i=1}}^{\mathbf{n}} \pi_{\mathbf{i=1}}^{\mathbf{y_i}}(\mathbf{1}-\pi)^{\mathbf{1-y_i}}$$

$$= \prod_{i=1}^{n} \left(\frac{exp(\mathbf{X_i}\beta)}{1+exp(\mathbf{X_i}\beta)}\right)^{y_i} \left(\frac{1}{1+exp(\mathbf{X_i}\beta)}\right)^{1-y_i}$$

The equation above yields the log likelihood given as;

$$\ell(\beta) = \sum_{i=1}^{n} [y_i log(\pi) + (1-y_i) log(1-\pi)]$$

$$= \sum_{i=1}^{n} [y_i \mathbf{X_i}\beta - \mathbf{log}(\mathbf{1}+\mathbf{exp}(\mathbf{X_i}\beta))]$$

The MLE are the β values that maximizes the above function. For instance using a technique such as iterative reweighted least squares or equating the first partial derivatives to 0 gives the critical points i.e. the $\hat{\beta}$

3.4.2 Hypothesis testing and confidence I

For large samples, the Wald confidence limits for β_j (i=1,..,k) is calculated by:

$$\hat{\beta}_j \pm Z_{\frac{\alpha}{2}}\{se(\hat{\beta}_j)\}$$

Also the 100(1 - α)% confidence limit for the odds ratio e^{β_j} can be calculates as:

$$exp[\hat{\beta}_j \pm Z_{\frac{\alpha}{2}}\{se(\hat{\beta}_j)\}]$$

The null hypothesis β_j = 0 depicts that the probability of success in a logistic regression model is independent of the X_j and the Wald large sample test statistic given by ;

$$Z = \frac{\hat{\beta}_j}{se(\hat{\beta}_j)} \sim N(0,1): \quad \beta_j = 0$$

For the alternative hypothesis $H_1 : \beta_j \neq 0$, the Z^2 will have a chi-square distribution with 1 degree of freedom.

The overall null hypothesis $H_0 : \beta_1 =, ..., = \beta_k = 0$ can be tested using the likelihood ratio test which is given by; χ^2_{LRT} =$2[logL(\hat{\beta}) - logL(0)]$.

i $L(\hat{\beta})$: maxi. likelihood of the complete model. i.e that where all predictor variables are included.

ii $L(0)$: max. likelihood of the simple model where only the intercept is included.

In cases where H_0 is rejected, it shows that all the predictor variables significantly contribute to the prediction of the response variable.

3.5 Evaluation of goodness of fit

The goodness of fit test evaluates how well the model fits a given set of observed responses. the test considers the data being the fit of the complete model, which is the extensive test model fit available. Each observed variable has its own parameter in the saturated model (Agresti, 2017). To confirm the lack of fit of the saturated model,through the test we can confirm if every parameters in the complete model and not in the simple model equals to 0.The model's deviation is measured by the likelihood ratio test statistic.

3.5.1 Deviance statistic and likelihood ratio test

The deviation statistic is calculated as follows:

$$Deviance = -2(L_m - L_s)$$

where L_m represents the max. log likelihood of the simple model i.e model with only the intercept and L_s represents the Max. log likelihood of the saturated model i.e the model with the intercept and predictors.

Let

$$Deviance_{null} = -2lnL_2 \;\; where \;\; L_2 = \frac{\text{null model likelihood}}{\text{saturated model likelihood}}$$

and

$$D_{simple/fitted} = -2lnL_1$$

We can then evaluate the impact of a single predictor or a set of predictors by subtracting the deviance of the model from the null deviance and comparing the results with the ch-square test statistic having df equivalent to the difference in the estimated number of parameters.

$$D_{simple/fitted} - D_nulll = -2lnL_1 - (-2lnL_2)$$
$$= -2ln\left(\frac{L_1}{L_2}\right)$$
$$= -2ln\left(\frac{\text{fitted model likelihood}}{\text{null model likelihood}}\right)$$

The model is said to be fit if the overall model deviance is significantly smaller than the null deviance and one can make a conclusion that a predictor or the set of predictors improved the model fitted significantly. The model is said to be unfit if the test gives a large value of the test statistic and a smaller p-value.

3.5.2 The Chi-square test for homogeneity

In a rxc contingency table. Let x_1 and x_2 denote two types of categorical variables, x_1 having r levels and x_2 having c levels respectively. The sample therefore have a rxc permutations of groups. Taking into account the null hypothesis H_0 such that probabilities of each cell is equal to some predetermined values π_{ij}.When the null hypothesis H_0 holds, then the values of the frequencies expected becomes e_{ij}=n_{ij}, which indicate the values of the expectations $E(n_{ij})$ of a data sample n with n_{ij} (Agresti, 2007).

Chi-square homogeneity test determines if the samples represented by the r rows (r independent samples) are homogeneous in relation to the proportions of the variables in categories c. This is shown when the number of observations or variables in j^{th} category are equal to those in the r populations. In this case, the data can be said to be homogeneous, (Sheskin, 2000). To investigate if the data is significantly different from the null hypothesis H_0, the frequencies observed n_{ij} and the anticipated frequencies e_{ij} are compared. Testing for the homogeneity, we write the null and alternative hypotheses as:

H_0: All the proportions of the same column in the rxc table are equivalent in the population under study that the study samples reflect.

H_1: All proportions of the a similar column in a rxc table are nonequivalent for a minimum of a single column in the population under study that the samples reflect.

The chi-square test statistic for testing null hypothesis, H_0 given by:

$$\chi^2 = \Sigma \frac{(O_k - E_k)^2}{E_k}$$

where:
χ^2 = Chi-squared
O_k = Observed value
E_k = Expected value
The chi-squared test statistic for a large sample size with r number of rows and c number of columns, degrees of freedom is given by:

$$df = (r-1)(c-1)$$

The chi- squared test static have the value zero for every i^th and j^th variable in cases where O_{ij} =E_{ij}. In cases where the sample size n is fixed, $O_{ij} - E_{ij}$ gives bigger difference of about 32 which implies a strong evidence against the null hypothesis H_0. . As the expected values E_{ij} increases, the chi-squared estimation also improve. Expected value $E_{ij} \geq 5$ is generally adequate for a good estimation.

The study will use the chi-square test to compare the proportion of individuals infected having tuberculosis s a rxc contingency table. i.e. number of columns are greater than 2. the test also presumes that the row totals are pre-determined before the commencement of the collection of data. Also, some of the assumptions made include:

i The analysis employs categories which are mutually exclusive.

ii These data were derived from a random sample group of n independent observations.

iii Every cell in the contingency table has an anticipated frequency of 5 or greater.

3.5.3 Fisher's exact test

In a 2x2 contingency table, the independence and or homogeneity relates to an odds ratio $\theta = 0$. Assume that n_{ij} cell counts obtained from two binomial samples withdrawn from each of the four cells. Then, taking into account a table having equal number of columns and rows totals as that of the data observed. this will generate a minimal null sample probability distribution of the count of cells that is independent of any unknown parameter (Agresti, 2007).

Table 2. A 2x2 Contingency Table

	Column 1	Column 2	Total
Row 1	n_{11}	n_{12}	n_{1+}
Row 2	n_{21}	n_{22}	n_{2+}
Total	n_{+1}	n_{+2}	n

The counts in the cells follows a hypergeometric distribution when conditioned based on the restricted table above. For each of the marginal totals in the rows and columns, n_{11} is useful in determining those other three cell counts. Hypergeometric formulae in this case express probabilities of the cell counts in relation to n_{11} only, (Agresti, 2007). Given θ = 1, then probability of a given value n_{11} is:

$$P(n_{11}) = \frac{\binom{n_{1+}}{n_{11}}\binom{n_{2+}}{n_{+1}-n_{11}}}{\binom{n}{n_{+1}}}$$

n_{i+}, n_{+j} respectively relates to the i^{th} row total and j^{th} column total.

Testing for the null hypothesis i.e. the homogeneity, and that the p-value is obtained by summing all probabilities of the hypergeometric outcomes as favorable as the alternative hypothesis from the outcomes observed. The contingency tables that have a bigger n_{11} also have a bigger sample odds ratio. That is,

$$\hat{\theta} = \frac{n_{11}n_{22}}{n_{12}n_{21}}$$

which strongly favors the H_1: $\theta > 1$.

The study uses Fisher's test of homogeneity to compare the fraction of individuals who tested positive for the tuberculosis disease in a 2 x 2 contingency table. The assumptions made in this case are similar to those of the Chi-square test for homogeneity only that we don't take into consideration the assumption of smaller unobserved (expected) frequencies.

3.6 Fitting data into the logistic regression model

In situations where the model does not achieve convergence i.e. When a model fails to converge, it means that the coefficients are meaningless since the iterative process failed to identify suitable solutions. Convergence failure can arise for a variety of reasons: the data being multicollinear(unacceptably high levels of correlation between predictor variables thereby increasing the standard error of the coefficients), sparsity of the data(large proportion of the data having zero cell counts), full separation(the predictors gives a perfect prediction for the criterion) and also when the proportion of predictors to cases is large.

Odds ratio

For two by two contingency table, the odds ratio(OR),θ is the measure of degree of association between the variables i.e. the outcome and exposure variables.
The estimate is given by:

$$\hat{\theta} = \frac{n_{11}n_{22}}{n_{12}n_{21}}$$

The *OR* indicates the odds that an event will happen given a certain exposure is present, in comparison to the probability that odds that the same event will occur given that, the exposure is absent.An *OR* greater than 1 shows that the occurrence of the event has increased, an *OR* less than 1 shows that the occurrence of the event has decreased and an *OR* = 1 shows that the occurrence of the event remains unchanged.

The 100(1-α)% confidence interval for $\hat{\theta}$ is:

$$exp[log\hat{\theta} \pm Z_{\frac{\alpha}{2}}\{se(\hat{\theta})\}]. \quad se(\hat{\theta}) = \sqrt{\frac{1}{n_{11}} + \frac{1}{n_{12}} + \frac{1}{n_{21}} + \frac{1}{n_{22}}}$$

and the $se(\hat{\theta})$ = $se(log\hat{\theta})$

Difference of proportions

Let π_1 and π_2 represent the success probabilities for samples in row 1 and those in row 2 respectively and let π_1 - π_2 denotes the proportions difference which compares the chances of success between the rows. Then p_1 and p_2 can be used to indicate the proportions of success of the samples and can be given as:

$$p_1 = \frac{n_{11}}{n_{1+}} \quad and \quad p_2 = \frac{n_{21}}{n_{2+}}$$

and p_1 - p_2 can be used as a unbiased estimate for π_1 - π_2. The estimate of the standard error of the difference in proportion $p_1 - p_2$ can be obtained using the formulae below in cases where the samples contained in the two rows are independent binomial.

$$se = \sqrt{\frac{p_1(1-p_1)}{n_{1+}} + \frac{p_2(1-p_2)}{n_{2+}}}$$

For large samples the 100(1-α)% CI for the difference of proportionπ_1 - π_2 can be found as:

$$(p_1 - p_2) \pm Z_{\frac{\alpha}{2}}(se)$$

The study will use the difference in proportion method to determine the difference in the TB prevalence amongst the sub-counties of Embakasi and also compare the general prevalence in Embakasi and the national tuberculosis prevalence.

3.7 Chapter Overview

The role of demographic and socioeconomic factors in tuberculosis predisposition was examined. Gender, for example, is one of these determinants. In the study area, the gender that is especially vulnerable to tuberculosis infection was explored. The age bracket that is more susceptible to tuberculosis infection was also identified. The participants in this study ranged in age from 15 to 80 years old for one group and the other group made up of children aged 0 to 14 years old. The study sample's educational level was explored to see if it was a risk factor for tb infection in the study region. Individuals were divided into two groups, that is those with only primary education and below, and the second group was made of those individuals with post primary school education. The role of occupation as a factor was explored to see if tuberculosis was linked to formal or informal work in the research area.

Individual's tuberculosis awareness level was also investigated to see if the spread of the infection had any correlation to the subject's understanding of the infection in terms of the symptoms, some of its causes and the transmission mechanisms. In addition the tuberculosis history of the individuals were also of interest as the study seek to find out if the tuberculosis re-occurrence was also a also a contributing factor. The nature of the housing, was also a factor of interested. the study was seeking to determine if the ventilation status of the houses could be a contributing factor for tuberculosis spread in the households. The number of people living in the same house were also investigated to find out the household's congestion level. Social health habits such as smoking and alcoholism were also taken into consideration to determine if they had some influence on the risks of tuberculosis infection. Also the nutrition levels were investigated to see if individuals which had proper nutrition and those which did not have proper nutrition had a different susceptibility level for tuberculosis.The ease of access to the health facility was also investigated. This potentially have an impact on how long it takes to seek medical help and thereby a shorter diagnosis and treatment time which may limit the community spread.

The fitted multiple logistic regression model was used to explore the effects of the predictor(s) on the odds of prevalence of tuberculosis. The model assumptions were listed, the model was also evaluated to see how well it fits the data.

4 Chapter 4: Data Analysis and Results

4.1 Introduction

The chapter contains results obtained from the analyzed data as well as the interpretations. The findings are divided into four categories. In the first section we discuss the various socio-demographic variables such as the gender and age.In the second and subsequent sections the findings from the chi-square test,the Fisher's exact test and the fitted logistic regression models are discussed. Here, the prevalence of tuberculosis among different categories and compared using a contingency table.

4.2 Tuberculosis prevalence factors

In Embakasi some demographic characteristics have been discovered to play a role in the prevalence of pulmonary tuberculosis. This section looks into the association between pulmonary tuberculosis prevalence and the predisposing factors.

4.2.1 Prevalence by gender

From the lab results, 7 out of 10 males tested for tuberculosis returned a positive test result while for the female gender 6 out of every 10 sample tested returned a positive tuberculosis test result. This translated to a percentage prevalence of 69.6% amongst the males and 65.14% amongst the females who attended chest clinic at Mama Lucy Kibaki Hospital.

Table 3. Tuberculosis prevalence according to gender

		Tuberculosis Cases			
		Positive	**Negative**	**Total**	**Positive%**
Gender	Males	132	59	190	69.63
	Females	72	37	110	65.14
Total		204	96	300	

Testing for the difference in the prevalence of tuberculosis between the male and the female. that is the hypothesis that:
H_0 : The rate of active tuberculosis prevalence between the males and females is similar .
The alternative hypothesis;

H_1 : The rate of active tuberculosis prevalence between the males and the females is not the same.
$H_0 : \mu_i = \mu_j$.
$H_1 : \mu_i \neq \mu_j$

Table 4. Fisher's test for tuberculosis prevalence according to gender

	Odds ratio	Proportion difference	P-value
	1.227	0.04496	0.249
95% CI	(0.7441, 2.024)	(-0.06482, 0.1547)	

At α level 0.05, the p-value of the Fisher's test is 0.249. We therefore fail to reject the null hypothesis that the prevalence of tuberculosis between the males and females in Embakasi is the same. Given the odds ratio value =1.227, this implied that the pulmonary tuberculosis amongst males was 1.23 times more likely than amongst the females in the same area. This finding was also in agreement with the 2016 Kenya tuberculosis survey which discovered that in Kenya the cases of tuberculosis are 2.5 times higher in males than in females.

4.2.2 Prevalence by smoking habits

The occurrence of tuberculosis in smokers was found to be at 78% and amongst the non-smokers was found to be at 61%.

Table 5. Tuberculosis prevalence among individuals who smoke and those who do not smoke

		PTB cases			
		Positive	**Negative**	**Total**	**Positive%**
Smoking	Yes	81	22	103	78.64
	No	121	76	197	61.73
Total		202	98	300	

The Fisher's exact test also was performed to test an hypothesis about the prevalence rates amongst the smokers and non-smokers. we have the null hypothesis, H_0: the spread of tuberculosis amongst smokers and non-smokers is the same and the alternative hypothesis, H_1: The rate of prevalence of tuberculosis amongst smokers and non- smokers is significantly different. i.e. H_0: $\mu_{smokers} = \mu_{non-smokers}$ vs H_1:$\mu_{smokers} \neq \mu_{non-smokers}$

The results of the Fisher's test are indicated in the table below.

Table 6. Tuberculosis prevalence amongst smokers and non-smokers, Fisher's test result.

	Odds ratio	Proportion difference	P-value
	2.133	0.1553	0.004
95% CI	(1.236, 3.683)	(0.04461, 0.2660)	

At α = 0.05, the p-value= 0.004 and therefore we reject the null hypothesis that the rate of tuberculosis prevalence is the same amongst individuals who smoke and non-smokers. Also the odds ratio was found to be 2.133. This meant that an individual who smokes was 2.133 times more likely to develop tuberculosis than their counterparts who do not smoke. This finding was also found to be consistent with the findings of [Fox and Menzies(2013)] on their research published on the American Journal of Epidemiology which showed that smoking doubles the risk of tuberculosis as it leads to weakening of the body's immune response. The study highlighted that Russia had the highest proportion of tuberculosis cases associated with smoking at 31.6% of the TB cases with a 95% confidence interval of (15.9 , 37.6).

4.2.3 Alcoholism and tuberculosis prevalence

From the study,individuals who were considered as alcoholic had a tuberculosis prevalence of 75.41% while those who did not consume alcohol/ non-alcoholic had a tuberculosis prevalence of 58.97%.

Table 7. Summary table of Alcoholics and Non-alcoholics and tuberculosis Prevalence

		Tuberculosis Cases			
		Positive	**Negative**	**Total**	**Positive%**
Alcohol	Yes	138	45	183	75.41
	No	69	48	117	58.97
Total		204	96	300	

To ascertain if there is a significant different on tuberculosis prevalence between individuals who are alcoholic and those who are non alcoholic, we consider the following hypothesis; H_0: the spread of tuberculosis amongst alcoholic individuals and non-alcoholic individuals is the same and the alternative hypothesis, H_1: The rate of prevalence of tuberculosis amongst alcoholic individuals and non- alcoholic individuals is significantly different. i.e. H_0: $\mu_{alcoholic} = \mu_{non-alcoholic}$ vs H_1:$\mu_{alcoholic} \neq \mu_{non-alcoholic}$
. A Fisher's exact test was performed to make this comparison and the results were as shown below.

Table 8. Tuberculosis prevalence between alcoholic and non-alcoholic individuals, Fisher's test result.

	Odds ratio	Proportion difference	P-value
	2.174	0.171	0.002
95% CI	(1.325, 3.567)	(0.06290, 0.2791)	

At α = 0.05 the p-value = 0.002, the test therefore rejects the null hypothesis that the the rate of tuberculosis prevalence amongst individuals who are alcoholic and those who are non-alcoholic is the same. This is also supported by the odds ratio value = 2.174 which shows that an individual who suffers from tuberculosis is 2.174 times more likely to be an alcoholic when compared to their non-alcoholic counterparts. This fining was also consistent with the study finding of (Sameer.,et al,2014) published in the European respiratory journal which stated that alcohol-linked tuberculosis cases were on the rise between the period 2000-2014, as the alcohol usage caused 22.02% of incidence cases with a 95% confidence level of (19.7-40.77). Excessive usage of alcohol negatively affects the body's immune system and therefore causes higher vulnerability to diseases such as tuberculosis.

4.2.4 Proper nutrition and tuberculosis prevalence

The study aimed to find out if the number of meals individuals occasionally had in a day could be a contributing factor to tuberculosis prevalence. The findings were such that, those who mainly took one meal a day has a tuberculosis infection rate of 81%, those who occasionally had two meals a day had a positivity rate of 69% and those who regularly had three or more meals a day had a positivity rate of 63%.

Table 9. Nutrition level and tuberculosis prevalence

		Tuberculosis Cases			
		Positive	**Negative**	**Total**	**Positive%**
meals per day	1	22	5	27	81.48
	2	115	52	167	68.86
	3	67	39	106	63.21
Total		204	96	300	68

A chi-square test was performed to ascertain if there is a major difference in the number of meals taken each day and the rate of tuberculosis prevalence. The hypothesis considered were:H_0: the spread of tuberculosis amongst individuals who took a single mean each day, those who took two meals each day and those who had three or more meals each day were all the same and the alternative hypothesis, H_1: The rate of prevalence of tuberculosis was different amongst individuals who had a single meal each day, those who had two meals each day and those individuals who had more than two meals each day. i.e. H_0: $\mu_{onemeal} = \mu_{twomeals} = \mu_{threemeals}$ vs H_1:$\mu_{onemeal} \neq \mu_{twomeals} \neq \mu_{threemeals}$

The chi-square test statistic was such that: $\chi^2_{.05,2}$ = 3.431 and the p-value = .1799
At 5% confidence confidence,we fail to reject the null hypothesis that the rate of tuberculosis prevalence in Embakasi is similar despite of the number of meals an individual had in a day. According to the Ministry of Health, Kenya in their tuberculosis and nutrition report of 2015,(TIBU,2015), the report revealed that under nutrition increased the risk of infections such as tuberculosis,leprosy and other lung diseases. The study also showed that their was a significant positive correlation between high tuberculosis prevalence and high food insecurity with more men at 70% than women experiencing this burden.

4.2.5 Education level and pulmonary tuberculosis infection

The individuals were categorized into four groups. i.e. those without any formal education, those who have only primary school level education, those with secondary school education level and lastly those with post secondary school education.
The table below shows the summarized results;

Table 10. Education level and tuberculosis prevalence

		None	Primary	Secondary	Tertiary
TB cases	Yes	38	85	40	26
	No	10	42	23	36
Total		48	127	63	62
Positive %		79.17	66.93	63.49	41.93

The result showed that 79% of individuals with no formal education were positive to tuberculosis,67% of individuals with only primary education were positive, amongst individuals with secondary school education,63% were positive and lastly only 42% of individuals with post secondary school education tested positive for tuberculosis.

A chi-square test result performed to test the null hypothesis that the tuberculosis prevalence is the same despite of the individual's level of education versus the alternative hypothesis which stated that the rate of the prevalence of tuberculosis was significantly different amongst individuals with different levels of education.
The results of the chi-square test were as follows; $\chi^2_{.05,3}$ = 5.52 with a p-value = .1356. Therefore at $\alpha = 0.05$ we fail to reject the null hypothesis and conclude that in the study,

an individual's level of education was not a significant factor in determining the prevalence of tuberculosis in Embakasi. The finding contradicted a study which was conducted in Bangladesh which showed that an individual's education level contributed to an individual's tuberculosis knowledge [Li et al.(2014)Li, Ehiri, Oren, Hu, Luo, Liu, Li, and Wang]

4.2.6 Prevalence of tuberculosis as a result of the individual's tuberculosis infection history.

From the study,individual's who did not have a history of tuberculosis infection had a prevalence rate of 72.85% while those individuals who had a previous history of tuberculosis infection had a prevalence rate of 61.07%

Table 11. Previous infection history and tuberculosis prevalence

		Tuberculosis Cases			
		Positive	**Negative**	**Total**	**Positive%**
TB history	Yes	91	58	149	61.07
	No	110	41	151	72.85
Total		201	99	300	

The rate of pulmonary tuberculosis prevalence was found to be 12.52% less in individuals with a history of previous infection compared to those individuals who had no history of tuberculosis infection.

The following hypothesis was considered to check if there was a statistical association between history of tuberculosis infection and the prevalence of tuberculosis in Embakasi.
H_0 : The rate of active tuberculosis prevalence between the individuals with previous history of infection and those without a previous history of infection is similar .
The alternative hypothesis;
H_1 : The rate of active tuberculosis prevalence between individuals with previous history of infection and those without a previous history of infection is not the same.
$H_0 : \mu_i = \mu_j$. vs $H_1 : \mu_i \neq \mu_j$. The results from the Fisher's exact test were;

Table 12. History of tuberculosis infection Fisher's test results

	Odds ratio	Proportion difference	P-value
	0.5867	0.1152	0.022
95% CI	(0.3592, 0.9584)	(0.009613, 0.2208)	

From the test, the p-value = 0.02 therefore the null hypothesis was rejected at 5% level of significance as the test showed that the rate of tuberculosis prevalence was not the same amongst individuals who had a history of the infection and those individuals who did not have a history of the infection. Also from the odds ratio = 0.5867 showed that the rate of prevalence in individuals with infection history was 41.33% less likely as compared to those individuals without a previous tuberculosis history. This finding was consistent with the findings of a survey conducted in Turkey by (Aylin, et al. 2012) which showed that the frequency of newly diagnosed patients was significantly higher than those of the former tuberculosis patients.

4.2.7 Occupation and tuberculosis prevalence

The study showed that patients who had formal employment had a tuberculosis prevalence rate of 67.50%, while those who had informal employment had a positivity rate of 74.63% and the positivity rate amongst the unemployed category was at 64.63%.

Table 13. Tuberculosis prevalence and the employment type

		Unemployed	Formal	Informal
TB Disease	Yes	53	54	103
	No	29	26	35
Total		82	80	138
Prevalence %		64.46	67.50	74.63

A chi- square test of homogeneity was used to check if the the rate of prevalence was influenced by an individual's type of employment. the following hypothesis were considered; H_0: The rate of prevalence of tuberculosis was similar amongst individuals who were unemployment, those who were formally employed and those individuals who had informal employment H_1: The rate of prevalence of tuberculosis was different amongst individuals who were unemployment, those who were formally employed and those individuals who had informal employment. i.e. H_0: $\mu_{unemployed} = \mu_{formal} = \mu_{informal}$ vs H_1:$\mu_{unemployed} \neq \mu_{formal} \neq \mu_{informal}$
The results were $\chi^2_{.05,2}$ = 6.061 and the p-value = 0.0483. Therefore at α = 0.05 the null hypothesis is rejected and conclude that the rate of tuberculosis prevalence is significantly different amongst individuals in different categories of employment.

4.2.8 Ventilation and tuberculosis prevalence

Individuals who reside in houses with proper ventilation had a prevalence rate of 40.00% while individuals who resided in poorly ventilated houses had a prevalence rate of 71.48%.

Table 14. Prevalence of tuberculosis and the level of ventilation

		PTB Cases			
		Positive	**Negative**	**Total**	**Positive%**
Ventilation	Good	12	18	30	40.00
	Poor	193	77	270	71.48
Total		205	95	300	

To check for the prevalence of tuberculosis amongst residents living in properly ventilated houses and those in poorly, ventilated houses the hypothesis considered were as follows: H_0 : The rate of active tuberculosis prevalence between the individuals living in poorly ventilated houses is similar to those living in properly ventilated houses .
The alternative hypothesis;
H_1 : The rate of active tuberculosis prevalence between individuals living in poorly ventilated houses is different with those living in properly ventilated houses.
$H_0 : \mu_i = \mu_j$. vs $H_1 : \mu_i \neq \mu_j$. The results from the Fisher's exact test were;

Table 15. Ventilation and tuberculosis prevalence

	Odds ratio	Proportion difference	P-value
	0.3403	0.2565	0.008
95% CI	(0.1564, 0.7403)	(0.0779, 0.4352)	

The p-value = .008 at α = 0.05 the test showed sufficient evidence to enable the rejection of the null hypothesis in favor of the alternative and conclude that the ventilation level in the residential houses significantly contribute to the prevalence of tuberculosis in Embakasi. The odds ratio = 0.3403 also implied that incidences of pulmonary tuberculosis was 65.97% less likely in properly ventilated houses compared to poorly ventilated houses.Individual's socio-economic standings can be reflected in the type of housing they reside in. These circumstances can hugely influence their respiratory health and also lead to the growth and transmission of tuberculosis. Such conditions can be lack of proper ventilation, congestion in the houses e.t.c (Brassard.,2006).

4.2.9 Tuberculosis awareness and the rate of prevalence

The study showed that individuals who had knowledge about tuberculosis had a prevalence rate of 59.72% while the prevalence rate amongst individuals who had little to know tuberculosis awareness was at 72.80%.

Table 16. Tuberculosis prevalence and the awareness level

Tuberculosis cases					
		Yes	No	Total	Positive%
Awareness level	Yes	43	29	72	59.72
	No	166	62	228	72.80
Total		209	91	300	

Fisher's test was used to test for the hypothesis; H_0 :The rate of tuberculosis spread between the individuals with good tuberculosis awareness and those without a good knowledge of the infection is similar .
The alternative hypothesis;
H_1 : The rate of tuberculosis prevalence between individuals with good awareness level and those without a good knowledge of the infection is not the same.
$H_0 : \mu_i = \mu_j$. vs $H_1 : \mu_i \neq \mu_j$. The results from the Fisher's exact test were;

Table 17. Fisher's test result for the prevalence of tuberculosis and the awareness level

	O dds ratio	Proportion difference	P-value
	0.525	0.1461	0.014
95% CI	(0.3067, 0.8986)	(0.0252, 0.2669)	

The p-value = 0.014 at α = .05, therefore we reject the null hypothesis in favor of the alternative hypothesis which stated that the prevalence rate of tuberculosis among individuals with and those without tuberculosis awareness is different. The odds ratio = .1461 showed that the prevalence of the infection was 85.39% less likely amongst individuals with good tuberculosis awareness compared to those who had no awareness. These findings were in agreement with a study conducted in Tanzania which aimed to determine the knowledge as well as the attitude on the prevention and factors associated with tuberculosis spread in Northern Tanzania, the study revealed that 9 out of every 10 participants who had good knowledge about tuberculosis had good practises whenever they were in close contact with anyone who showed signs of tuberculosis and only 11% of individuals who did not have prior knowledge about the disease maintained good practises when in close proximity to an infected or an individual who exhibited tuberculosis symptoms and signs [Kazaura and Kamazima(2021)].

4.2.10 Congestion and tuberculosis prevalence

The rate of prevalence of tuberculosis among individuals living in congested houses was at 74.13% and the rate of prevalence amongst individuals who stayed in non-congested households was at 57.58%.

Table 18. Congestion and the prevalence of tuberculosis

Tuberculosis Cases		Yes	No	Total	Positive%
Congestion	**Yes**	149	52	201	74.13
	No	57	42	99	57.58
Total		206	94	300	

Consider the hypothesis; H_0 :The rate of tuberculosis spread between the individuals living in congested houses and individuals living in uncongested houses is similar .
The alternative hypothesis;
H_1 : The rate of tuberculosis prevalence between individuals living in congested and those in uncongested houses is different.
$H_0 : \mu_i = \mu_j$. vs $H_1 : \mu_i \neq \mu_j$. The results from the Fisher's exact test were;

Table 19. Fisher's test result for level of congestion and tuberculosis prevalence

	Odds ratio	Proportion difference	P-value
	2.042	0.1594	0.004
95% CI	(1.233, 3.382)	(0.0477, 0.2711)	

From the results, the null hypothesis is rejected at α = .05 since p-value = .004 is less than the level of significance and conclude that the level of prevalence of tuberculosis is different in houses that are overcrowded and those which are not overcrowded. Odds ratio for the test is 2.042, implies that, it is 2.04 times more likely amongst patients who live crowded houses as compared to those who live in spacious houses. The number of bedroom a house has was used to measure the level of crowding in the houses i.e. a house is said to be congested if the house does not have enough bedrooms for the number and composition of the individuals living in the house all year round.This finding was supported by previous finding of study conducted amongst the First Nations in Canada which are characterized by congested houses. In the Canadian study, it was shown that increasing the number of persons per room by 0.1 lead to an increase in the risk of new tuberculosis infection by two or more cases in the community by at least 40% [Meier-Stephenson et al.(2007)Meier-Stephenson, Langley, Drebot, and Artsob].

4.3 Logistic regression modelling of tuberculosis prevalence factors

The factors associated with the prevalence of tuberculosis in Embakasi were fitted in a logistic regression model. To study the association between the prevalence of tuberculosis and the predictor variables, the following hypothesis was tested:The null hypothesis is such that all the estimators are equal to zero, i.e. none of the parameters significantly influence the prevalence of tuberculosis and the alternative hypothesis is such that at least one of the estimates is not equal to zero.i.e at least one of the parameters significantly contribute to the prevalence of tuberculosis in Embakasi.

H_0: $\beta_i = 0$ i= 1,2...,10.

H_1: $\beta_i \neq 0$ for at least one i, i = 1,2...,10.

The results from the logistic regression model are presented in the table below:

Table 20. Tuberculosis factors and the logistic regression model

	$\hat{\beta}_i$	**Se($\hat{\beta}_i$)**	**e($\hat{\beta}_i$)**	**95% CI for e($\hat{\beta}_i$)**	**Z**	**p-value**
Intercept	0.1821	0.5242	1.2019	0.4373,3.2911	0.356	0.7217
Sex	-0.1382	0.306	0.8709	0.4781, 1.5865	-0.452	0.6515
Education	0.1167	0.2966	1.1237	0.6284, 2.0098	0.393	0.694
Occupation	-0.1396	0.375	0.8697	0.4170, 1.8137	-0.372	0.7098
Awareness	-0.3248	0.3324	0.7227	0.3767, 1.3864	-0.977	0.3285
History	-0.2007	0.3115	0.8182	0.4443, 1.5066	-0.644	0.5194
Alcoholism	0.7484	0.3656	2.1136	1.0323, 4.3274	2.047	0.0406
Nutrition	0.3208	0.3313	1.3782	0.7200, 2.6383	0.968	0.3329
Smoking	0.5259	0.3186	1.692	0.9062, 3.1593	1.65	0.0989
Ventilation	-0.2115	0.3439	0.8094	0.4125, 1.5881	-0.615	0.5384
Congestion	0.649	0.3282	1.9136	1.0058, 3.6410	1.978	0.048
Null deviance: 372.47 on 297 df; Residual deviance: 343.32 on 283 df						

The test result indicates that one or more of the partial regression coefficients is greater than zero therefore, the null hypothesis was rejected and the conclusion made is that the overall model is significant given the difference of the deviance= 29.1 on 10 degrees of freedom and the p-value = .0033

Alcoholism and congestion were the most significant factors for the prevalence of tuberculosis in Embakasi with the p-values of 0.0406,0.048 and associated odds ratio of 2.1136,1.9136 respectively. Smoking, level of ventilation in the houses were a moderate significant factors for the prevalence of tuberculosis. From the results, an individual's gender, occupation, number of meals an individual had in a day, history of tuberculosis infection, level of education and the level of awareness of tuberculosis were found to be independent of the prevalence of the tuberculosis infection.

Incorporation of interaction terms after performing backward elimination of less significant terms i.e. terms which have p-value greater than 0.05 gives a new model fitted below.

Table 21. Reduced logistic regression model of the prevalence of tuberculosis

	$\hat{\beta}_i$	Se($\hat{\beta}_i$)	e($\hat{\beta}_i$)	95% CI for e($\hat{\beta}_i$)	Z	p-value
Intercept	-0.5123	0.3634			-1.3822	0.16754
Alcoholism	1.5121	0.4583	4.5418	1.8520,11.137	3.319	0.00083
Smoking	0.552	0.3042	1.7552	0.9655,3.1915	1.851	0.06548
Congestion	1.4028	0.4247	4.0635	1.7648,9.3563	3.283	0.00089
Alco:Conge	-1.3779	0.5585	0.2529	0.0851,0.7530	2.454	0.01362
Null deviance:372.47 on 297 df; Residual deviance:340.12 on 290 df						

The test indicated that the model overall was still significant as the difference of the deviance was 32.35 on 4 degrees of freedom and p-value < .0001. From the results on the table above, the incidences of tuberculosis was less likely amongst individuals who were alcoholic given their is congestion in their residential places. Even though incidences of tuberculosis were more likely amongst individuals who were smokers, smoking was of marginal significance in predicting for the prevalence of TB.

The equation for the logistic regression model can be given as:

$$log[\frac{\hat{\pi}(x)}{1-\hat{\pi}(x)}] = -.501 + 1.513(alcoh) + .562(smk) + 1.401(cong) - 1.379(alch.cong)$$

4.4 Prevalence of tuberculosis in Embakasi

This section contains the findings of the prevalence of tuberculosis in Embakasi and compares the same to the national rate of prevalence.

4.4.1 Prevalence according to age categories

Table 22. Age categories and prevalence of tuberculosis

		Age (years)							
		0-14	15−24	25-34	35−44	45−54	55-64	65−80	**Total**
TB Cases	Yes	14	32	44	46	41	11	16	204
	No	12	17	26	23	10	6	2	96
Total		26	49	70	69	51	17	18	300
Prevalence %		53.85	65.31	62.86	66.67	80.39	64.71	88.89	68

We test the hypothesis that, the prevalence of tuberculosis amongst the different age categories are the same vs the alternative that prevalence of tuberculosis amongst the different age categories are not all similar.

At α = .05, the chi-square test result is $\chi^2_{.05,6}$ = 10.76 and p-value = .0962, we therefore fail to reject the null hypothesis and conclude that their is no significant difference in the prevalence of tuberculosis as a result of difference in ages of individuals in Embakasi. According to the United States tuberculosis report of 2020,persons aged between 0-14 years had the lowest tuberculosis incidence rate at 0.4/100,000 and persons aged above 65 years had the highest tuberculosis incidence rate at 3.4/100,000(cdc.gov,2020 report). these findings were in agreement with the study findings in this study.

4.4.2 Embakasi and the national tuberculosis prevalence

In 2020, an estimated number of people who developed tuberculosis were at 139,000 in a population of 53,771,300 out of these only 71,646 cases were reported and put on treatment(stoptb.org,2021).Mama Lucy Kibaki teaching and referral hospital which serves a total population of more than 2.1 million spreading across the entire Eastlands region of Nairobi reported a total of 4,157 out of these about 2987 were reported to have been from the larger Embakasi sub-counties having a total population of about 1,437,216. These resulted into a national prevalence of 14 per 10,000 and the Embakasi prevalence at 21 per 10,000 individuals.

Table 23. Embakasi and National TB prevalence

		Tuberculosis Cases			
		Yes	No	Total	**Cases(per 10000)**
Region	**Embakasi**	2987	1434229	1437216	21
	National	71646	53669654	53771300	14

The hypothesis tested were: the null hypothesis; H_0 : The rate of tuberculosis prevalence in Nairobi Embakasi is not statistically different from the national prevalence.The alternative hypothesis; H_1 : The rate of active tuberculosis prevalence in Nairobi Embakasi is statistically different from the national prevalence rate.i.e. $H_0 : \mu_i = \mu_j$. vs $H_1 : \mu_i \neq \mu_j$. The results from the chi-square test are shown below:

Table 24. Embakasi and National TB prevalence

	Odds ratio	Proportion difference	$\chi^2_{.05,1}$	p-value
	4.6614	0.0068	576.3803	< 2.2e-16
95% CI	(4.6163, 4.7069)	(.0056, .0080)		

The null hypothesis was rejected as the test indicated that the rate of tuberculosis prevalence between Embakasi and the National prevalence were significantly different at $\alpha =$ .05 given the p-value < 0.000001. The odds ratio = 4.661 showed that the likelihood of tuberculosis in Embakasi sub-county was 4.661 times more likely than that of the national situation.

4.4.3 Tuberculosis prevalence in different sub-counties of Embakasi

The prevalence of tuberculosis in the 5 Embakasi sub-counties were analysed and the results were as shown below.

Table 25. Embakasi sub-counties and TB prevalence

		Sub-counties in Embakasi				
		West	Central	East	North	South
TB Cases	Yes	26	98	26	32	22
	No	16	41	15	12	12
Total		42	139	41	44	34
Prevalence	%	61.9	70.5	63.41	72.73	64.71

We consider the following hypothesis for this test, the null hypothesis; H_0 : The rate of active tuberculosis prevalence in Nairobi Embakasi sub-counties is similar.The alternative hypothesis; H_1 : The rate of active tuberculosis prevalence in Nairobi Embakasi sub-counties is not the same. $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu$ vs $H_1 : \mu_1 \neq \mu_2 \neq \mu_3 \neq \mu_4 \neq \mu_5$. From the chi-square test result gives $\chi^2_{.05,4}$ = 2.135 and the p-value = .07109,therefore we fail to reject the null hypothesis at 5% level of significance and conclude that the prevalence rate of tuberculosis amongst the five sub-counties of Embakasi are similar.

4.5 Chapter overview

The study revealed that the prevalence of tuberculosis in Embakasi was higher in males compared to the females,also the age group 25-34 had the most active tuberculosis cases, amongst the individual sub-counties in Embakasi is similar at α = 0.05 as the chi-square statistic = 2.135. The overall model was found to be significant when predicting for the prevalence of TB given the factors investigated. The main risk factors for TB prevalence in Embakasi according to the reduced logistic regression model are alcoholism and congestion.

5 Chapter 5: Conclusions and Recommendations

5.1 Summary

The study aimed at finding the factors associated with the prevalence of tuberculosis in Embakasi. These factors were investigated and analyzed in this thesis report, these findings wee then compared at the sub-counties level and at the national level and appropriate inferences made.

5.2 Conclusions

The study found that the prevalence of tuberculosis in the informal and low-middle income settlements within the city is higher compared to the national rate. Embakasi area is highly dominated by low income earners and therefore may not be able to afford proper housing and a decent standard of living. Embakasi being the most populated area of Nairobi city has also contributed greatly to the tuberculosis prevalence in Nairobi and the country at large.

According to the findings, factors that significantly contribute to the prevalence and spread of tuberculosis in Embakasi were majorly related to the host. They include; alcoholism, lack of proper ventilation in residential places, overcrowding or congestion in the houses, smoking,individual's tuberculosis awareness and the individual's occupation. When all the risk factors were fitted in the logistic regression model so as to find out how they are associated with the tuberculosis infection, only the risks of alcoholism, smoking and congestion were found to be significant in predicting for the tuberculosis incidences in Embakasi sub-counties. The findings also revealed that more males were prone to tuberculosis infection as compared to their female counterparts.This may be mainly due to the fact that most of the males engage as casual laborers and also were not able to afford proper nutrition levels as they spend better part of the days away from their houses doing menial jobs, also most of the alcoholics and smokers were also males who ranged between the ages of 25 and 45.

Lastly, the study also determined that there was no significant difference in the prevalence of tuberculosis among the five sub-counties which make up the larger Embakasi region of Nairobi. As a result, they both have an almost equal level of contribution to the regional and national prevalence. Putting in place similar measures in place to curb this menace is therefore one step towards the MoH vision of ending tuberculosis in Kenya by the year 2030.

5.3 Recommendations

The prevalence of tuberculosis in Embakasi is higher compared to the prevalence at the county and national level. This means that Embakasi at a higher scale, contribute to the national tuberculosis load. To minimize this burden, the study recommends the following actions;

There should be an increased community sensitization on the problem of smoking and excessive alcoholism amongst the youths in Embakasi who are highly susceptible to the infection. This can be done through mass public education on the risk dangers posed by these habits, also engaging these individuals in meaningful empowerment projects will address this problem as they will be more responsible, improved access to health care services will also improve the diagnostic rate and hence minimize the community spread of TB. lastly, access to proper nutrition and sanitation services will also play a big role in limiting the spread of tuberculosis in Embakasi thereby tackling the prevalence.

5.4 Preliminaries

Future Research

The tuberculosis situation in Kenya has been extensively analyzed by both the governmental and non-governmental bodies. Embakasi region is probably the most densely populated area in the country and therefore may give the best case scenario on how to effectively root out tuberculosis in the country by the year 2030. Therefore the following future studies are recommended specifically for Embakasi region;

i The extend of prevalence of multi drug resistant and extensive drug resistant tuberculosis in Embakasi.

ii The extend to which HIV infection contributes to TB in the region.

iii Also from this study,there was insufficient evidence to reject the null hypothesis that the level of nutrition had an influence of the level of tuberculosis prevalence. Therefore, a further research was recommended in the study area to ascertain if the results were obtained were indeed true since most of the global and national studies were in disagreements with this study findings.

iv Also, the study finding that the prevalence of tuberculosis in Embakasi was not influenced by the education level was found to be inconsistent with most global study findings which shows that an individual's level of education could influence their tuberculosis awareness and therefore it leads to an improvement in tuberculosis control mechanism leading to lower rates of infections since most of these individuals could take the appropriate measures to limit the spread of infectious diseases.

Bibliography

[Ahmed(2018)] Amira Fadlallah Mohammed Alhassan Ahmed. *Assessment of the Effectiveness of Immunization programme Among Children Under Five Years and their mothers in Shendi locality from 2016–2017*. PhD thesis, Abdallah Ahmed Adam Belal, 2018.

[Broekmans et al.(2002)Broekmans, Migliori, Rieder, Lees, Ruutu, Loddenkemper, and Raviglione] JF Broekmans, GB Migliori, HL Rieder, J Lees, P Ruutu, R Loddenkemper, and et MC Raviglione. European framework for tuberculosis control and elimination in countries with a low incidence: recommendations of the world health organization (who), international union against tuberculosis and lung disease (iuatld) and royal netherlands tuberculosis association (kncv) working group. *European Respiratory Journal*, 19(4):765–775, 2002.

[Cassidy et al.(2009)Cassidy, Hedberg, Saulson, McNelly, and Winthrop] P Maureen Cassidy, Katrina Hedberg, Ashlen Saulson, Erin McNelly, and Kevin L Winthrop. Nontuberculous mycobacterial disease prevalence and risk factors: a changing epidemiology. *Clinical Infectious Diseases*, 49(12):e124–e129, 2009.

[Cress et al.(2017)Cress, Leitz, Kim, Amore, Suzuki, Linhardt, and Koffas] Brady F Cress, Quentin D Leitz, Daniel C Kim, Teresita D Amore, Jon Y Suzuki, Robert J Linhardt, and Mattheos AG Koffas. Crispri-mediated metabolic engineering of e. coli for o-methylated anthocyanin production. *Microbial Cell Factories*, 16(1):1–14, 2017.

[Emmanuel et al.(2010)Emmanuel, Shitandi, et al.] Nyakeriga Emmanuel, A Shitandi, et al. Spatial prevalence rate of tuberculosis as presenting to a rural hospital. *Journal of Applied Sciences Research*, (August):1008–1011, 2010.

[Falzon et al.(2011)Falzon, Jaramillo, Schünemann, Arentz, Bauer, Bayona, Blanc, Caminero, Daley, Dur
D Falzon, E Jaramillo, HJ Schünemann, M Arentz, M Bauer, J Bayona, L Blanc, JA Caminero, CL Daley, C Duncombe, et al. Who guidelines for the programmatic management of drug-resistant tuberculosis: 2011 update, 2011.

[Fox and Menzies(2013)] GJ Fox and D Menzies. Epidemiology of tuberculosis immunology. *The New Paradigm of Immunity to Tuberculosis*, pages 1–32, 2013.

[Gibson et al.(2008)Gibson, Abramson, Wood-Baker, Volmink, Hensley, and Costabel] Peter G Gibson, Michael Abramson, Richard Wood-Baker, Jimmy Volmink, Michael Hensley, and Ulrich Costabel. *Evidence-based respiratory medicine*. John Wiley & Sons, 2008.

[Gounder et al.(2015)Gounder, Zulz, Desai, Stenz, Rudolph, and Raymond] Prabhu P Gounder, Tammy Zulz, Shalini Desai, Flemming Stenz, Tsang Rudolph, Karen, and Raymond. Epidemiology of bacterial meningitis in the north american arctic, 2000–2010. *Journal of Infection*, 71(2):179–187, 2015.

[Halezeroğlu and Okur(2014)] Semih Halezeroğlu and Erdal Okur. Thoracic surgery for haemoptysis in the context of tuberculosis: what is the best management approach? *Journal of thoracic disease*, 6(3):182, 2014.

[Harries et al.(2018)Harries, Lin, Kumar, Satyanarayana, and Takarinda] Anthony D Harries, Yan Lin, Ajay MV Kumar, Srinath Satyanarayana, and Takarinda. What can national tb control programmes in low-and middle-income countries do to end tuberculosis by 2030? *F1000Research*, 7, 2018.

[Jindal(2011)] SK Jindal. Textbook of pulmonary and critical care medicine vols 1 and 2. 2011.

[Kazaura and Kamazima(2021)] Method Kazaura and Switbert Rwechungura Kamazima. Knowledge, attitudes and practices on tuberculosis infection prevention and associated factors among rural and urban adults in northeast tanzania: A cross-sectional study. *PLOS Global Public Health*, 1(12):e0000104, 2021.

[Konstantinos(2010)] Anastasios Konstantinos. Testing for tuberculosis. 2010.

[Kumar et al.(2007)Kumar, Toledo, and Patel] Ashwani Kumar, Jose C Toledo, and Rakesh P Patel. Mycobacterium tuberculosis doss is a redox sensor and dost is a hypoxia sensor. *Proceedings of the National Academy of Sciences*, 104(28):11568–11573, 2007.

[Lawn and Zumla(2011)] SD Lawn and AI Zumla. Seminar tuberculosis. *Lancet*, 378 (9785):57–72, 2011.

[Li et al.(2014)Li, Ehiri, Oren, Hu, Luo, Liu, Li, and Wang] Ying Li, John Ehiri, Eyal Oren, Daiyu Hu, Xingneng Luo, Ying Liu, Daikun Li, and Qingya Wang. Are we doing enough to stem the tide of acquired mdr-tb in countries with high tb burden? results of a mixed method study in chongqing, china. *PLoS One*, 9(2):e88330, 2014.

[Lienhardt(2001)] Christian Lienhardt. From exposure to disease: the role of environmental factors in susceptibility to and development of tuberculosis. *Epidemiologic reviews*, 23(2):288–301, 2001.

[Mainous III and Pomeroy(2001)] Arch G Mainous III and Claire Pomeroy. *Management of antimicrobials in infectious diseases*, volume 349. Springer, 2001.

[Mandell(2010)] Gerald L Mandell. Bennett's principles and practice of infectious diseases. *Mycobacterium tuberculosis*, 2:2853–86, 2010.

[Matteelli et al.(2014)Matteelli, Roggi, and Carvalho] Alberto Matteelli, Alberto Roggi, and Anna CC Carvalho. Extensively drug-resistant tuberculosis: epidemiology and management. *Clinical epidemiology*, 6:111, 2014.

[MBOGA(2018)] LYDIA NYABATE MBOGA. A statistical investigation to determine the predisposing risk factors for predicting prevalence of active tuberculosis in north pokot sub county, west pokot county, kenya. 2018.

[Meier-Stephenson et al.(2007)Meier-Stephenson, Langley, Drebot, and Artsob] V Meier-Stephenson, JM Langley, M Drebot, and H Artsob. Archived-canada communicable disease report. *Canada Communicable Disease Report*, 33(11), 2007.

[Mohamud(2004)] Abdikadir Mohamed Mohamud. *Prevalence of Tuberculosis in Central Division of Wajir District, North Eastern Province, Kenya*. PhD thesis, Kenyatta University, 2004.

[Mukadi et al.(2001)Mukadi, Maher, and Harries] Ya Diul Mukadi, Dermot Maher, and Anthony Harries. Tuberculosis case fatality rates in high hiv prevalence populations in sub-saharan africa. *Aids*, 15(2):143–152, 2001.

[Muture et al.(2011)Muture, Keraka, Kimuu, Kabiru, and Ombeka.] Bernard N Muture, Margaret N Keraka, Peter K Kimuu, Ephantus W Kabiru, and Ombeka. Factors associated with default from treatment among tuberculosis patients in nairobi province, kenya: a case control study. *BMC public health*, 11(1):1–10, 2011.

[Nambiar et al.(2016)Nambiar, Daza, and Livornese] Puja H Nambiar, Alejandro Delgado Daza, and Lawrence L Livornese. Clinical impact of vaccine development. *Vaccine Design*, pages 3–39, 2016.

[Organization et al.(2009)] World Health Organization et al. The stop tb strategy, case reports, treatment outcomes and estimates of tb burden. *Global tuberculosis control: epidemiology, strategy, financing*, pages 187–300, 2009.

[Organization et al.(2011)] World Health Organization et al. The sixteenth global report on tuberculosis. *Tuberculosis: Executive summary. WHO Rep*, 2011:1–2, 2011.

[Organization et al.(2014)] World Health Organization et al. *Global tuberculosis report 2014*. World Health Organization, 2014.

[Pai et al.(2008)Pai, Zwerling, and Menzies] Madhukar Pai, Alice Zwerling, and Dick Menzies. Systematic review: T-cell–based assays for the diagnosis of latent tuberculosis infection: an update. *Annals of internal medicine*, 149(3):177–184, 2008.

[Ragonnet et al.(2019)Ragonnet, Trauer, Geard, Scott, and McBryde] Romain Ragonnet, James M Trauer, Nicholas Geard, Nick Scott, and Emma S McBryde. Profiling mycobacterium tuberculosis transmission and the resulting disease burden in the five highest tuberculosis burden countries. *BMC medicine*, 17(1):1–12, 2019.

[Singh et al.(2005)Singh, Mynak, and Kumar] M Singh, ML Mynak, and Kumar. Prevalence and risk factors for transmission of infection among children in household contact with adults having pulmonary tuberculosis. *Archives of disease in childhood*, 90(6):624–628, 2005.

[Society et al.(2000)] American Thoracic Society et al. Targeted tuberculin testing and treatment of latent tuberculosis infection. *Am J Respir Crit Care Med*, 161:S221–S247, 2000.

[van Zyl Smit et al.(2010)van Zyl Smit, Pai, Yew, Leung, Zumla, Bateman, and Dheda] Richard N van Zyl Smit, Madhukar Pai, Wing-Wai Yew, CC Leung, A Zumla, ED Bateman, and K Dheda. Global lung health: the colliding epidemics of tuberculosis, tobacco smoking, hiv and copd. *European respiratory journal*, 35(1):27–33, 2010.

[Wangara et al.(2019)Wangara, Kipruto, Ngesa, Kayima, Masini, Sitienei, and Ngari] Fatihiyya Wangara, Hillary Kipruto, Oscar Ngesa, James Kayima, Enos Masini, Joseph Sitienei, and Faith Ngari. The spatial epidemiology of leprosy in kenya: A retrospective study. *PLoS neglected tropical diseases*, 13(4):e0007329, 2019.