RISK FOR LATENT TUBERCULOSIS INFECTION AMONG HOUSEHOLD AND NON-HOUSEHOLD CONTACTS OF TB CASES IN AN URBAN AFRICAN SETTING

by

ROBERT KAKAIRE

(Under Direction of Christopher Whalen)

ABSTRACT

Statement of the Problem: Tuberculosis remains a major global public health problem. About 25 percent of the world's population is infected with *Mycobacterium tuberculosis (M. tuberculosis)*. Although individuals with latent tuberculosis are not infectious, they are at risk of developing active disease and becoming the next pool of infectious individuals. Studies show that risk for progression to disease once infected may range from 5% in children who are recently exposed to as high as 16% among individuals infected with the human immunodeficiency virus. **Purpose:** To identify individuals in the community who are latently infected with tuberculosis. **Methods:** We conducted two inter-related studies; Study 1 was to examine the effect of household and non-household exposure to an infectious tuberculosis case on the prevalence of latent tuberculosis infection in the community. We identified index TB cases and matched them with controls who in turn enumerated their social contacts who were then approached and asked to participate in the study. Diagnosis of latent tuberculosis infection was by tuberculin skin test. The second was an incident cohort study to investigate whether differences in the prevalence of latent tuberculosis infection between men and women observed in previous studies could be due to differences in the incidence of latent tuberculosis infection. A cohort of tuberculin skin test negative individuals was

enrolled between 2014 and 2016 and were followed -up and a repeat TST was placed to determine tuberculin skin test conversion. **Main results:** Infection was highest among household case contacts and lowest in household contacts of controls. Prevalent infection among non-household case contacts was similar to that of non-household control contacts. The incidence study found that men had a higher risk for infection than women. Fifty-one percent of converters were men representing an incidence rate of 16.2 per 100PYO while 48.4% of the converters were women representing an incidence rate of 10.6 per 100PYO among women.

INDEX WORDS: Prevalence, Latent tuberculosis infection, Household and Non-household, Index TB cases, Controls, Social contacts, Incidence rate, Converter, Men, Women

RISK FOR LATENT TUBERCULOSIS INFECTION AMONG HOUSEHOLD AND NON-HOUSEHOLD CONTACTS OF TB CASES IN AN URBAN AFRICAN SETTING

by

ROBERT KAKAIRE

B.A (SS), Makerere University, Kampala 1996

M.P.H, University of Alabama, Birmingham, 2005

A Dissertation Submitted to the Graduate Faculty of the University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PUBLIC HEALTH

ATHENS, GEORGIA

2018

RISK FOR LATENT TUBERCULOSIS INFECTIONAMONG HOUSEHOLD

AND NON-HOUSEHOLD CONTACTS OF TB CASES IN AN URBAN AFRICAN SETTING

by

ROBERT KAKAIRE

Major Professor:	Christopher Whalen
Committee:	Fredrick Quinn
	Juliet Sekandi
	Hanwen Huang

Electronic Version Approved:

Suzanne Barbour
Dean of the Graduate School
The University of Georgia
May 2018

DEDICATION

This dissertation work is dedicated to my parents, Mr. and Mrs. Kakaire who have loved me unconditionally and whose good examples have taught me to work hard for things that I inspire to achieve particularly in education and life in general.

ACKNOWLEDGEMENTS

I acknowledge the research support of the National Institutes of Health (NIH), National Institute of Allergy and Infectious Diseases (NIAID) through research mechanism grant # RO1 AI093856 awarded to The University of Georgia. This work would not have been possible without the support of my dissertation committee members: Drs. Christopher Whalen, Fred Quinn, Juliet Sekandi and Hanwen Huang. I would especially like to thank Dr. Christopher Whalen, the chair of my committee. As my teacher and mentor, he has taught me more than I could ever give him credit for here. He has shown me, by his example, what a good scientist should be.

I am grateful to all of those with whom I have had the pleasure to work with on this dissertation. Special thanks go to the Epidemiology in action group (EIA) at UGA: Drs Jane and Simon Mutanga, Ronald Galiwango, Samuel Kirimunda, Maria Eugenia Castellanosu, Jonathan Sanchez-Garcia and the COHSONET study team especially Dr. Srah Zalwango based at Lubaga hospital in Uganda.

Finally, nobody has been more important to me in the pursuit of this doctorate than the members of my family whose love and guidance are with me in whatever I pursue. Most importantly, I wish to thank my loving and supportive wife, Agnes, and my three wonderful daughters, Patience, Tacey and Kacey, who provide unending inspiration to keep me going. They are the ultimate role models and I know they will one day read this dissertation.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

TB	Tuberculsosis
TST	Tubeculosis Skin Test
PYO	Person Years of Observation
NIH	National Institutes of Health
NIADS	National Institute of Allergy and Infectious Diseases
WHO	Worls Health Organization
UGA	University of Georgia
EIA	Epidemiology in Action
HIV	Human Immuno deficiency Virus
IGRA	Intereron Gamma Release Assays
DNA	Deoxyribonucleic Acid
PCR	Polymerase Chain Reaction
PPD	Purified Protein Derivative
BCG	Bacillus Calmette–Guérin
QFT	Quantiferon Test
KCCA	Kampala Capital City Authority
NTLP	National Tubeculosis and Leprosy Program
TU	Tuberculin Unit
SD	Standard Deviation
PR	Prevalence Ratio

CI	Confidence Interval
LTBI	Latent Tuberculosis Infection
ARTI	Annual Risk of Tuberculosis Infection
COHSONET	Community Health and Social Networks of Tuberculosis

CHAPTER 1

INTRODUCTION

This dissertation provides a critical review of the existing literature on latent tuberculosis infection particularly in resource limited Africa and highlights some existing gaps in knowledge that will be filled by two manuscripts. A brief perspective of work on latent tuberculosis in the households and in the community is provided. Plus a brief background of tuberculosis control using household contact investigations and its limitations are explained. A critical summary of the recently published literature on latent tuberculosis infection in developing countries where the burden of TB is particularly high is also provided. The first manuscript is about prevalence of latent tuberculosis within the households of tuberculosis cases and in the community. The second manuscript is about incidence of latent tuberculosis among men and women. The final chapter synthesizes the results from the two manuscripts and also discusses policy implications and opportunities for future research.

Tuberculosis has been around us for decades and remains a major public health problem[1]. The World Health Organization declared tuberculosis a global emergency in 1993. Tuberculosis kills nearly two million people annually mainly in low and middle-income countries, particularly in sub-Saharan Africa and Asia[3]. About 25 percent of the world's population is infected with *Mycobacterium tuberculosis (M. tuberculosis)*, the causative microorganism of tuberculosis[2]. The natural history of tuberculosis follows a two-stage process after infection, beginning with a latent

stage and thereafter tuberculosis disease or active disease[3,4] (Figure 1.1). Latent tuberculosis infection can progress to active disease stage, depending on various host, agent and environmental factors. Nearly all incidents of tuberculosis disease begin with a period of asymptomatic infection.

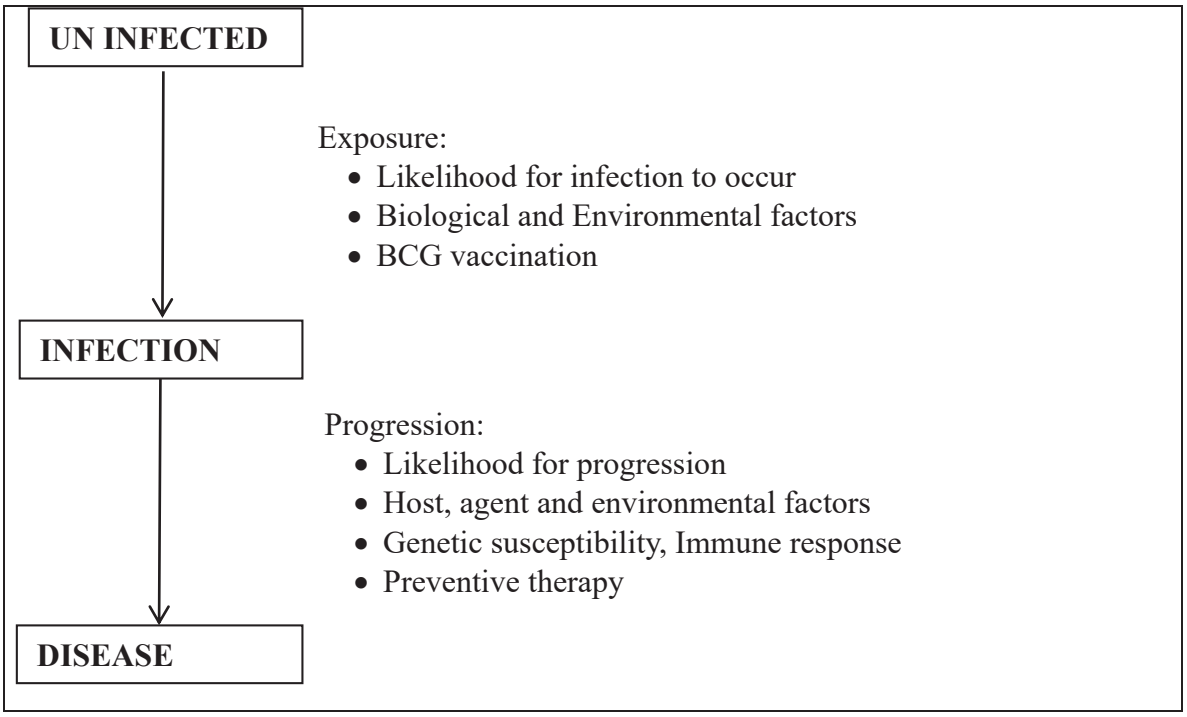

Figure 1.1 *M. Tuberculosis* Transmission model

1.1. Public health significance

People with latent tuberculosis are not infectious, but are at risk of developing active disease and becoming infectious at some poni in the future[5]. Studies show that risk for progression to disease once infected may range from 5% in children[6] to as high as 16% among individuals infected with the human immunodeficiency virus (HIV)[7]. Individuals with latent infection serve as a reservoir from which new cases of active tuberculosis arise, propagating the global epidemic[8]. Latent tuberculosis infection is not evenly distributed across the world. Resource-rich developed countries, such as the United States of America, have fewer cases of infection whereas developing countries in sub-Saharan Africa and Asia (Figure 1.1) bear the highest prevalence of infection[9].

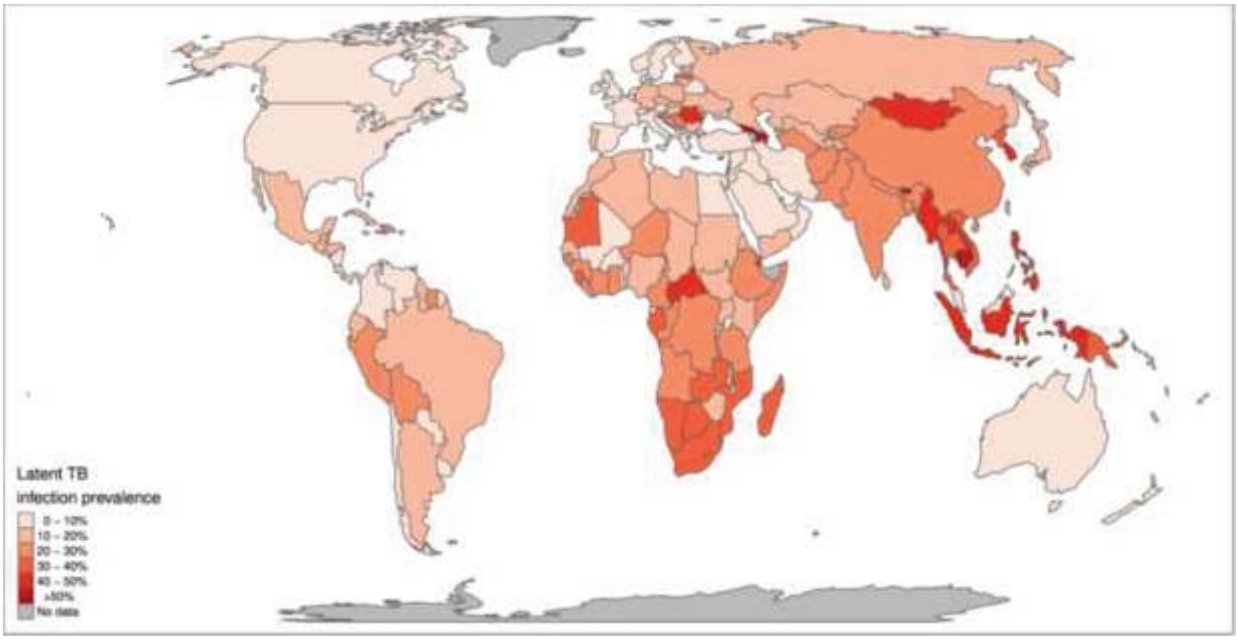

Figure 1.2 Estimated population prevalence of latent tuberculosis infection by country in 2014[9].

The global health community set itself the task of eliminating tuberculosis as a public health problem by 2050[10]. Approaches for tuberculosis control to achieve elimination include 1) early case detection and prompt initiation of effective treatment as a principle means of controlling transmission and reducing incidence of tuberculosis, 2) vaccination with Bacillus Calmette–Guérin (BCG) as a primary prevention strategy, 3) treatment of HIV infection with anti-retroviral medications to reduce the risk of disease in this susceptible population, 4) infection control through work practice and administrative measures to prevent transmission in healthcare settings, and 5) treatment of latent tuberculosis infection.

Treatment of latent infection requires undertaking screening for infection in contacts of patients with active tuberculosis and selected high-risk groups. Contacts may be categorized into close contacts such as household members and social, or casual contacts who are non-household

contacts of tuberculosis patients. The two broad categories of individuals at high risk of tuberculosis infection may be further reclassified by gender and assessed for latent tuberculosis. Mathematical modelling suggests that community treatment of latent infection is one of the effective ways to reduce incidence of active tuberculosis[11,12]. This approach would identify infected individuals in the community likely to progress to disease and provide an opportunity to treat such subclinical infections and prevent future disease. The strategy reduces the reservoir of latently infected persons from which new cases arise and provide a critical opening to interrupt progression to disease and reduce the global burden of tuberculosis disease[8,13]. Although treatment of latent infection is shown to be effective, and is often advocated by tuberculosis control programs for the prevention of tuberculosis, its practice is erratic, and seldom used in developing countries.

1.2. Study Rationale and Hypothesis

The proposed dissertation addresses two important and inter-related issues, regarding latent tuberculosis infection in an African urban setting. In the first study aim, I compare the effect of household and non-household exposure to an infectious tuberculosis case on the prevalence of latent tuberculosis infection. We hypothesized that proximity to the index TB case defined as household vs non-household status is associated with infection among contacts of infectious patients. As a result, we predicted that there will be a gradient in prevalence of infection when moving from household to non-household contacts and that this risk is greater than what is observed in social contacts of controls.

In the second aim, I evaluated the difference in incidence of latent infection between men and women. Here, we hypothesize that the probability for infection is greater in men than women because most surveys show higher prevalence of latent tuberculosis infection in men. I will

investigate potential factors for the differences in incidence between men and women. Although I propose two aims for this dissertation, it is important to know that the data for these projects comes from an integrated set of studies in Kampala, Uganda, for which I have been the project coordinator. I outline rationale further for each aim below.

a) Study Aim 1: Household and non-household contacts

To estimate the age specific prevalence of latent tuberculosis infection and tuberculosis disease among household and non-household contacts of infectious tuberculosis patients.

Tuberculosis control programs have over the years focused on detection and treatment of active tuberculosis cases - an approach very important in interrupting transmission[14]. Household contact investigation is the principal method used to detect additional tuberculosis patients and recently exposed persons with latent infection at risk for progression to active tuberculosis[14]. Household contact investigation has been a successful means of approaching tuberculosis control in industrialized countries like the US[15]. As part of the household contact investigation, a tuberculosis control program worker will diagnose and treat the index case, identify any other cases of active tuberculosis within the household, initiate treatment, and finally screen for latent infection among house hold contacts. Contacts with latent infection do not have active tuberculosis but may be treated to prevent tuberculosis disease.

In high burden settings, tuberculosis control programs typically define high-risk contacts mainly by virtue of them sharing a household with the index case even though there are other high risk settings, such as the work place of a tuberculosis case. This is so partly because it is often difficult to perform screening among contacts outside the household. Although household contact investigation is widely advocated, it its not clear it is effective in the control of tuberculosis at a population level in high burden resource limited countries. These studies are difficult to conduct,

but since there is sufficient content validity in the approach, many experts strongly advocate for household contact investigations.

Household contact investigations generally do not identify and evaluate all contacts of tuberculosis patients. Many contacts are not screened so are not be treated for latent tuberculosis infection. This approach has limited success because it does not take into consideration the many non-household contacts, who may be equally at risk of infection[16]. A South African study that integrated contact and environmental data in evaluating tuberculosis transmission in South Africa estimated up to 84% of infections occurring outside of one's own household[16]. Molecular epidemiology and community studies have pointed to high rates of community transmission ranging from as high as 64 to 72%[17-19]. Studies that use *M. tuberculosis* genotype data to identify transmission chains suggest that in high-incidence settings, up to 60% of cases of secondary disease among household contacts may be due to exposures outside the home[16,20].

In this analysis we hypothesized that being a household member of an index tuberculosis case is associated with infection. We predict that the highest prevalence of infection will be observed among household contacts of tuberculosis cases followed by non-household contacts of tuberculosis cases. The lowest prevalence is predicted to be among contacts of controls (Figure 1.2).

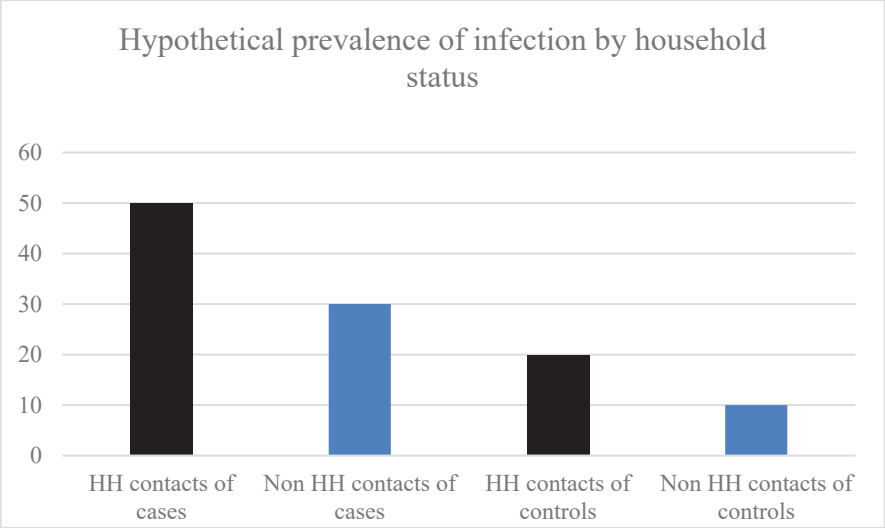

Figure 1.3. Hypothetical prevalence of infection by household status

b) Study Aim 2: Differences in latent tuberculosis infection between men and women

To estimate incidence of latent tuberculosis infection by sex and determine factors associated with infection among men and women.

Across the world, the incidence of reported tuberculosis disease is greater in men than women at all ages except in childhood, when they are higher in women[21]. Globally, sex differences in the prevalence of latent tuberculosis infection begin to appear between 10 and 16 years of age, and remain higher for men than women thereafter[22]. The reasons for the higher male prevalence and incidence are poorly understood. Moreover, this discrepancy is seen in both industrialized countries like the United States[23] and in developing countries such as Uganda[24-26]. Current prevalence data suggests that men are more likely to be diagnosed with active disease than women[27], with a male-to-female ratio of 2:1 to 3:1 globally[1]. The reasons for this discrepancy in tuberculosis disease have been widely studied. The case for latent infection is different as not much has been studied to understand whether this is the result of biologic or environmental factors that may account for the differences.

Possible Underlying Theories

Differences in the prevalence of tuberculosis between men and women may be factual or may arise from chance or bias. In theory if the risk of infection once exposed and the risk for progression to active disease once infected were the same for men and women, then the observed gender differences in infection or disease would be explained by bias. The most probable reasons for bias may be due to differential access to diagnostic tests and care or may be due to differential reporting of cases where may be male cases are reported in preference to female cases. But why do we believe that the difference between men and women is factual? In active case finding prevalence surveys[10,28,29] where unbiased random sampling methods have been used, yet prevalence of tuberculosis is still higher in men. In the same way, if the risk for infection once exposed and the risk for progression to active disease were greater for men than women, then the observed gender differences would be factual. This phenomenon would be explained by biological, behavioral or environmental factors since these factors may influence opportunities for exposure or progression[30].

Men and women may be exposed to *M. tuberculosis* in different ways because of differences in risk behaviors and social roles. For instance, males may have more social contacts than women, or may spend more time in settings that may be conducive for transmission, such as bars. Although these factors are important in tuberculosis transmission dynamics, we do not know whether this difference is biological or behavioral/environmental, nor do we know which factors are responsible for the overall difference in infection[30]. This limits our understanding of the effect of biologic sex in the transmission dynamics of tuberculosis and constrains our ability to make

recommendations aimed at targeting most-at risk groups in a bit to improve current tuberculosis control policies.

1.3. Study setting

Uganda is one of the world's tuberculosis high-burden countries with an annual risk of infection of 3%[1]. Uganda has no country specific data on the prevalence of infection however, different measures in specific sub-populations around the country have yielded consistent results with prevalence ranging from 16%[31] among adolescents in rural Eastern Uganda to as high as 49% in Lubaga division of Kampala[18,32,33]. The three complementary studies of tuberculosis transmission which provide the foundation of this thesis were conducted in Lubaga division Kampala, Uganda. Lubaga division located in the western part of the city (Figure 1.3) is one of five administrative units of metropolitan Kampala comprising 13 parishes and 135 zones similar to census tracks that are political units headed by a local council. The total population of Lubaga division is approximately 383,215 people based on a census performed in 2014[34].

Figure 1.4 Study area, Lubaga division. Source: Uganda Bureau of Statistics, 2010

1.4. Project overview

The goal of this project is to improve upon detection of latent tuberculosis infection and disease. To contribute to this goal, we will analyze data from separate but complementary epidemiologic studies of tuberculosis transmission in urban Africa, namely the community health and social networks of tuberculosis (COHSONET) study, the community tuberculosis skin test survey and the longitudinal tuberculosis transmission study. This analysis will have two separate parts, each addressing a separate research question but coming from the same underlying studies outlined above.

1.5. Dissertation deliverables

Two manuscripts will be written out of this dissertation thesis. The first manuscript will report the prevalence of latent infection among household and non-household contacts of index

cases. The research findings will provide an insight into the burden of latent infection in the general population. The findings will provide TB decision makers with information to be used in designing programs to prevent transmission of tuberculosis targeted to at risk populations.

The second manuscript will report differences in incidence of latent tuberculosis infection between men and women and provide an insight into why men appear to be carrying a higher burden of infection than women. Understanding gender-specific differences in tuberculosis epidemiology is a gateway to addressing barriers to effective care. This dissertation is in line with strategies that prevent progression of latent infection to active disease mitigated through improved latent tuberculosis detection and treatment initiation.

CHAPTER 2

LITERATURE REVIEW

This chapter provides a review of epidemiological literature of latent tuberculosis infection. The review focuses on latent tuberculosis infection incidence, prevalence, and effect of gender on the two measures, infection in specific risk groups and the general population in high income and in resource-limited settings. Where available these aspects will be considered at population level.

2.1 Cause of tuberculosis in humans

Tuberculosis is caused by a group of phylogenetically closely related bacteria, collectively known as the Mycobacterium tuberculosis complex[35]. Mycobacterium tuberculosis complex comprises of Mycobacterium tuberculosis (M. tuberculosis), M. africanum, M. canettii, M. bovis, M. microti, M. orygis, M. caprae, M. pinnipedii, M. suricattae[36]. *M. tuberculosis* is the agent responsible for tuberculosis in humans. Its survival is limited outside of the human body and there is no known animal reservoir[35]. Transmission of tuberculosis is from a person with active pulmonary disease to a susceptible person who inhales infected droplets generated from the bronchial secretions of the person with the disease.

Four broad factors determine the likelihood of transmission of *M. tuberculosis* namely the number of organisms being expelled into the air, the concentration of organisms in the air determined by the volume of the space and its ventilation, the length of time an exposed person breathes the contaminated air, and the immune status of the exposed individual[37].

2.2 Immunology

The route of entry of *M. tuberculosis* is via the respiratory tract. After inhalation, the droplet nucleus is carried down the bronchial tree and implants in a respiratory bronchiole or alveolus[38]. The ability for tubercle bacilli to establish an infection in the lung depends on both the bacterial virulence and the inherent ability of the alveolar macrophage that ingests it[39]. If the bacillus is able to survive initial defenses, it can multiply within the alveolar macrophage until sufficient to elicit a cellular immune response. Before the development of cellular immunity, tubercle bacilli spread via the lymphatics to lymph nodes and through the bloodstream[38].

In persons with uncompromised immunity, collections of activated T cells and macrophages form granulomas that limit multiplication and spread of the organism[40]. Antibodies against *M. tuberculosis* are formed but do not appear to be protective. For the majority of individuals with normal immune function, multiplication of *M. tuberculosis* is stopped once cell-mediated immunity develops. Once the organism has been contained within the granuloma, latent infection is established but if the organism is not fully contained by this initial response, then infection will progress to active disease. The ability of the host to respond to the organism may be reduced by certain diseases such as HIV infection, lymphomas and diseases associated with immunosuppression[41]. Immunosuppressive drugs such as corticosteroids may affect the ability of the host to respond to the organism. The body's response to infection provides protection against re-infection. The likelihood of reinfection is a function of the risk of re-exposure, the intensity of such exposure, and the status of the host's immune system[38]. In most individuals with latent tuberculosis infection, the combination of macrophages, dendritic cells and T cells is sufficient to

maintain a controlled, asymptomatic infection state. However, in some individuals, the infection can progress to clinical disease within weeks and for some other individuals after decades[42].

2.3 Natural history of tuberculosis

Latent tuberculosis infection is the presence of *M. tuberculosis* in an individual without clinical, imaging, or microbiologic evidence of active disease.[43] The main reservoir of *M. tuberculosis* is the patient with pulmonary tuberculosis. Diagnosis of latent tuberculosis infection is based on the tuberculin skin test (TST), which is also known as the Mantoux test or Interferon-gamma release assays (IGRAs). These diagnostic tests classify into any of these states: absence of any clinical or laboratory evidence of infection, infection without clinically active disease and active disease. The state of infection is determined by the ability of the host innate and adaptive immune systems to eradicate or control *M. tuberculosis*[44].

Tuberculosis can develop in any part of the body but disease involving the lungs, which occurs in 60–75% of cases, is critical for transmission to occur[45]. Persons living with a tuberculosis patient are at risk of inhaling particles containing infectious bacilli released into the atmosphere through speaking, cough or sneeze. Upon exposure to *M. tuberculosis*, 10–30% of individuals become infected with *M. tuberculosis*[46]. Once in hailed, the bacilli are ingested by local macrophages, multiply within the cells and within 2 weeks granuloma are formed. This is followed by the development of an immune response, signaled by a delayed-type hypersensitivity reaction within 4 weeks which control of infection[45].

Not all patients with tuberculosis develop immunological signs of infection. The initial infection is often asymptomatic but may be associated with fever, mild chest symptoms and increased inflammatory markers[47]. Three asymptomatic phases of tuberculosis infection may

occur in an individual namely the innate phase, adaptive phase and quiescent phase. Some individuals may eliminate the bacteria during the innate immune phase, without generating T cell memory[44]. In the adaptive immune phase, T cells are engaged by antigen-presenting cells, and this generates effector and memory T cells and central memory T cells which may result into infection being cleared at this stage. Most exposed individuals will enter the quiescent phase, which may persist for life. In this phase, the bacteria are contained inside granulomas[44].

In situations where the host fails to eliminate the pathogen, disease will occur[44]. In about 90% of individuals who are infected, the body's immunity either kills the bacteria, or keeps the bacilli suppressed in the granuloma causing latent tuberculosis infection. In the general population, the lifetime risk of progression from latent tuberculosis infection to active disease is about 5% to 10% however with HIV coinfection, this rises to 30%[48]. In immune-suppressed individuals such as HIV-infected persons, tuberculosis infection progresses to disease more rapidly due to the weakening of their cellular immunity. In healthy individuals, only about 10% of infected persons develop active disease and become ill.

Without treatment, 50% of patients with pulmonary tuberculosis disease die within five years, and about 25% remain sick with chronic, infectious tuberculosis which can be spread to the community. The remainder about 25% will naturally recover their healthy, due to their adaptive immune defenses, but they could become sick again due to reactivation at any time if the tuberculosis bacteria are latent[47]. *M. tuberculosis* exists in a dormant state during latent tuberculosis infection, but reactivation may occur especially when one's immune status is compromised[49]. The latency state of infection and probability of reactivation depend on the balance between host immunity and the influence of external factors[48].

2.4 Risk factors for latent tuberculosis infection.

Risk of infection once exposed to the tuberculosis bacilli is determined by both biological and environmental risk factors[50]. When an individual is exposed to an infectious tuberculosis case, the risk of becoming infected is determined mostly by the combined effect of three aspects namely the degree of infectiousness of the source case, the duration of exposure of the susceptible individual to an individual with active disease, and the degree of susceptibility of that individual to infection[51]. Some risk factors may be specific to particular populations or regions such as foreign born but are also shown to increase the susceptibility to infection[52].

For purposes of this dissertation, I will summarize these risk factors along two world health organization (WHO) broad categories i.e. high income and upper middle-income countries with an estimated tuberculosis incidence rate of less than 100 per 100,000 population and resource limited and other middle-income countries with an estimated tuberculosis incidence rate of more than 100 per 100,000 population[53].

Degree of exposure of the susceptible individual

Close contacts of infectious tuberculosis cases including household and community contacts are at a higher risk of becoming infected with *M. tuberculosis.* Household contact studies among tuberculosis patients over the years have established this effect[54]. A clear demonstration of the influence of proximity to an infectious case was shown in an airplane outbreak investigation where passengers seated within two rows of the index tuberculosis patient were more likely to have positive tuberculin skin test compared to those in the rest of the section (30.8% versus 3.6%, RR = 8.5, CI = 1.7–41.3)[55]. Clusters of patients infected with identical strains raise the importance of proximity to an infectious case as the driving factor behind community transmission of

tuberculosis. Similar results from a household contact study in Uganda found active tuberculosis common among household contacts of infectious index cases, occurring in 6 percent of contacts[32]. This partly informs the decision to target household members of tuberculosis cases for screening because probability of infection increases with the proximity as tuberculosis infection tends to aggregate in certain households[56]. The "clustering" of tuberculosis infection within families may be a reflection of infection due to proximity to a contact case within the confines of the home[57].

Crowding has been documented to increase the risk of tuberculosis infection because it increases the risk of exposure by increasing the likelihood of contact between a susceptible individual and an infectious case[58]. Individuals exposed to infectious cases in closed houses may experience an increased risk of infection compared with individuals in lesser-crowded settings. In a study that looked at crowding and tuberculosis infection by Baker and colleagues in New Zealand showed that for every 1% increase in the proportion of overcrowded households in a census block led to an 8% increase in tuberculosis incidence in that block, holding other variables constant[59]. It remains a challenge to quantify the risk associated with crowding because the extent of the risk is directly affected by housing quality.

Degree of susceptibility of the Individual to latent tuberculosis infection

Age as a risk factor may be regarded as being bi-directional in nature. Children are at higher risk of contracting tuberculosis because, the alveolar macrophage, the first line of defense in the innate immune response that amplifies the response to infection is limited in infants compared to adults[60]. This may allow mycobacteria to overpower the effects of the innate immune system prior to the initiation of an antigen-specific immune response[61]. On the other hand, older age is often associated with TST positivity either due to increased age as a risk factor of contracting latent

tuberculosis infection or due to increased cumulative exposure to *Mycobacterium tuberculosis* as people grow older[62-64].

Infectiousness of the tuberculosis Index case

The concentration of bacilli in the sputum from a tuberculosis case is positively correlated with the infectivity of the tuberculosis patient[65]. Several studies have shown that sputum smear-positive pulmonary tuberculosis cases are more likely to infect their contacts than sputum smear-negative tuberculosis cases[65-67]. This is relationship is further modified by cavitary disease and HIV status[68].

2.5 Diagnosis of tuberculosis disease

The procedures for diagnosing tuberculosis are usually carried out on patients who have symptoms suggestive of active disease. Different methods are used to confirm the diagnosis of tuberculosis disease that is; microscopy, culture, chest radiograph and gene expert.

Acid- Fast Bacilli smear

The detection of acid-fast bacilli in stained smears examined microscopically is the first bacteriologic evidence of the presence of mycobacteria in sputum. It is the easiest and quickest procedure that can be performed, and it provides quantitative estimation of the number of bacilli seen in the sputum[69]. This is reported as either 'No acid-fast bacilli seen' or 'scanty' when very few acid-fast bacilli are seen. When any red bacilli are seen, the smear results are report as 'acid-fast bacilli positive' and an indication of the number of bacteria present is given. The presence of acid-fast-bacilli on a sputum smear often indicates tuberculosis disease, but it does not confirm a

diagnosis of tuberculosis because some acid-fast-bacilli are not *M. tuberculosis*. We performed cultures on all initial samples to confirm the diagnosis.

Acid- Fast Bacilli culture

The acid-fast bacilli culture is the inoculation of a sputum specimen onto culture media Becton-Dickinson Mycobacteria Growth Indicator Tube and Lowenstein-Jensen media slant, incubation at 37°C for up to six weeks. Detection is usually a binary measure of growth or no growth during this incubation period[70].

Chest Radiograph

A posterior-anterior chest radiograph is used to detect chest abnormalities. Features of chest radiograph consistent with tuberculosis disease include miliary picture, pleural effusion and mediastinal lymph gland enlargement with lung infiltration. Lesions may appear anywhere in the lungs and may differ in size, shape, density, and cavitation. These abnormalities may suggest tuberculosis disease, but cannot be used to definitively diagnose active tuberculosis. Chest radiograph may be used to rule out the possibility of pulmonary tuberculosis in a person who has had a positive reaction to a tuberculin skin test or tuberculosis blood test with no symptoms of disease. Because of these limitations and cost chest x-ray was never used in this study for diagnosis.

GeneXpert

The GeneXpert *M. Tuberculosis* assay is a nucleic acid amplification test that uses a disposable cartridge with the GeneXpert Instrument System[70]. A sputum sample is collected from the patient with suspected tuberculosis disease and is mixed with the reagent that is provided with the assay, and a cartridge containing the mixture is placed in the GeneXpert machine. GeneXpert

purifies and concentrates *M. Tuberculosis* from sputum samples, isolates genomic material from the captured bacteria by sonication and subsequently amplifies the genomic DNA by PCR[70].

2.6 Diagnosis of latent tuberculosis infection

The screening for latent tuberculosis infection is done either using the TST or IGRAs, which detect the presence of infection.

Tuberculin Skin Test

The skin test utilizes a substance called PPD that has a standard chemical composition belonging to the bacillus antigens[71]. In an infected individual, intradermal injection of tuberculin provokes the liberation of lymphokines that cause a delayed-type hypersensitivity reaction, demonstrated by the appearance 24–72 hours later of localized inflammatory cells into the skin, causing a swelling at the site of injection[72]. BCG, bacilli and environmental mycobacteria also induce the delayed-type hypersensitivity reaction induced by microbial antigens of *M. tuberculosis*[72].

Interpretation of Skin Test Reactions

In any population, the likelihood that a positive test represents a true infection is influenced by the prevalence of infection with *M. tuberculosis*[72]. The tuberculin skin test has a specificity of approximately 99% in populations that have no other mycobacterial exposures or BCG[73]. Based on the sensitivity, specificity, and the prevalence of tuberculosis in different populations, three cut points are recommended for defining a positive tuberculin reaction[74]. Among individuals who are at great risk of developing tuberculosis disease if they become infected with *M. tuberculosis*[75] such as persons who have had recent close contact with persons who have active tuberculosis, HIV and cancer patients, a cut-off point of 5 or more millimeters is recommended[74]. Persons who are

immunosuppressed because of disease or those on treatment with steroids are more likely to progress to tuberculosis disease if they are infected with *M. tuberculosis* and therefore a lower cut off of 5 or more millimeters for separating positive from negative reactions is in these groups[74]. An induration of 10 or more millimeters is considered positive in exposed individuals in endemic areas like Uganda. In low risk countries such as the US, an induration of 10 or more mm in recent immigrants (< 5 years) from high-prevalence countries is considered positive[74]. Other categories that are ordinarily categorized as positive at the 10mm cut off include injection drug users, residents and employees of high-risk congregate settings, mycobacteriology laboratory personnel. Then again, an induration of 15 or more millimeters is considered positive among persons with no known risk factors for tuberculosis[74].

Gamma Release Assays (IGRAs)

Gamma Release Assays are used to determine if a person is infected with *M. tuberculosis* by measuring the immune response to tuberculosis proteins in whole blood[76]. There are two IGRAs available i.e. QuantiFERON®-TB Gold-in-Tube test and the T-SPOT® *TB* test. The QuantiFERON-TB Gold In-Tube and the T-SPOT.TB are in vitro assays that measure interferon-γ released by effector T cells following stimulation with *M. tuberculosis* antigens namely ESAT-6, CFP-10, and TB7.7[48]. Specimens are mixed with peptides that simulate antigens derived from *M. tuberculosis* and controls. In a person infected with *M. tuberculosis*, the white blood cells recognize the simulated antigens and release interferon-gamma (IFN- γ).

Interpretation of Gamma Release Assays

Results are based on the amount of IFN- γ released. IGRAs tend to detect recent as well as remote T-cell responses although the sensitivity of IGRAs decreases with higher degrees of

immunosuppression[77],[78]. The interpretation of IGRAs is based on the amount of IFN-γ released, in QFT, or on the number of cells that release IFN-γ, in T-SPOT®.TB. and therefore provide both qualitative and quantitative results[76]. Qualitative results are reported as positive, negative, indeterminate or borderline while quantitative results are reported as numerical values that include a response to the tuberculosis antigen and 2 controls, nil and mitogen. Quantitative results may be useful for clinical decision making in individual cases, in combination with risk factors[76].

The ability of QFT-GIT and T-SPOT® to identify persons at the highest risk of progressing to active tuberculosis is poor. Both tests cannot predict future disease among persons with positive tests, and strong positive tests do not suggest a higher risk[79]. IGRAs and the TST are immune based, and have limited ability to predict disease or to identify which individuals with tuberculosis infection are likely to progress to active tuberculosis disease. They both have limited sensitivity in people with HIV infection and cannot differentiate between recent and remote infection, and past infection when a person is exposed multiple times[80].

2.7 Special considerations in testing for latent tuberculosis infection

Boosted Tuberculin Skin Test Reactions

Some individuals with latent tuberculosis infection may have a negative reaction to the TST if many years have passed since they became infected with tuberculosis[81]. They may have a positive reaction to a subsequent TST because the initial test stimulates their ability to react to the TST. This is reaction is referred to as the "booster phenomenon" and may incorrectly be interpreted as a skin test conversion[81,82]. Boosted reactions are common in individuals exposed to other mycobacteria or who have been vaccinated with BCG[75]. Because of this reason, the "two-step

method" is suggested at the time of initial testing for individuals who may be tested periodically such as health care workers[83].

In this method, persons who have a negative initial PPD skin test undergo a second tuberculin test 1–3 weeks after the first. The results from the second test are considered to be the "correct" result, i.e., those individuals with a positive reaction on the second test are considered to be previously infected, and those with a negative reaction on the second test should be considered uninfected[84]. In uninfected persons, a positive result on any future PPD skin test is interpreted as a skin test conversion[84]. Repeated skin testing with tuberculin does not induce a positive skin test reaction in individuals who have no cellular immunity to the antigens in PPD[75,84]. A cut off point of $\geqslant$ 10 mm is recommended for individuals who have normal immunity and are without other risk factors that would increase their likelihood of developing active disease. The rate of boosting varies widely between different studies, ranging from 0% to 31%[85,86]. Variation in the rate of boosting among different studies may be a reflection of differences in the populations studied in terms of age, BCG vaccination status, and previous exposure to *M. tuberculosis* or environmental mycobacteria[89].

HIV Infection and Latent Tuberculosis Infection

Interpretation of TST is different in HIV infected individuals. A negative TST or IGRA result does not exclude latent tuberculosis infection as HIV infected individuals may have a compromised ability to react to tests for tuberculosis infection[87]. Similarly, after the initiation of antiretroviral therapy, HIV-infected individuals previously known to have negative TST or IGRA results may test positive because the immune response may be restored by antiretroviral therapy. After the initiation of antiretroviral therapy, repeat testing for latent tuberculosis infection is

recommended for HIV-infected persons previously known to have negative TST or IGRA results[76].

2.8 Latent tuberculosis infection prevention and control strategies

Prevention strategies are designed to avoid infection with *Mycobacterium tuberculosis* or to stop active disease from developing in previously infected persons. Three approaches namely treatment of active tuberculosis disease to prevent infection, vaccine and prevention therapy are used to reduce the future burden of latent tuberculosis infection.

Treatment of active tuberculosis disease

Identification and treatment of individuals with active disease is critical to breaking the replacement principle of transmission[88]. One approach to preventing the spread of tuberculosis is to keep it controlled in people who have it. People with active tuberculosis if quickly diagnosed and treated, are less likely to pass it on to others in their household or communities. Infectiousness of tuberculosis patients diminishes rapidly once effective treatment is initiated with. Community studies suggest that transmission greatly diminishes or ceases within 2-4 weeks of effective treatment initiation[89].

Bacillus Calmette–Guérin (BCG) vaccine

WHO recommends BCG vaccination during infancy in tuberculosis endemic countries[90]. Introduction of bacilli into the body provokes the same immunological reactions as primary infection with tubercle bacilli without leading to disease. BCG vaccination provides partial immunity against the degree of primary infection and particularly against the acute forms of tuberculosis in children[90]. In many parts of the world where tuberculosis is common, BCG vaccine is used to protect infants and young children from serious, life threatening tuberculosis[1]

Immunization of infants less than five years old with BCG protects them against tuberculosis meningitis and other severe forms of tuberculosis. Because the protection provided by the vaccine is variable and less certain in older children, BCG vaccine is not recommended after 12 months of age[91]. There is a great variation in the protection conferred by BCG vaccination, from 0% efficacy in some endemic areas such as Southern India to 75% protection in the United Kingdom[92]. For this reason the policy on BCG vaccination varies especially in low burden countries with some countries like the US electing not to administer it routinely[93].

Isoniazid prevention therapy

The third strategy for reduction in the prevalence of latent tuberculosis infection is the detection and treatment of people with latent tuberculosis. This approach is aimed at individuals with the highest risk of progression to active tuberculosis and once identified are treated. Tuberculosis preventive therapy aims at eliminating latent infection with Mycobacterium tuberculosis and prevents progression to active disease. Research has shown conflicting benefits of preventive therapy for instance Leung and colleagues showed treatment of latent tuberculosis infection with isoniazid for up to 12 months with a protective efficacy of up to 90 percent[94], and a combination of rifampicin and isoniazid for 3 months has protective efficacy of 60 percent[95]. Two other studies from South Africa and Botswana suggested benefits with isoniazid treatment particularly for participants with positive TST with 74% reduction in tuberculosis incidence in this subgroup[96,97]. An isoniazid prevention trial in South African miners in the 'Thibela study', incidence of tuberculosis decreased to 63%, showing that isoniazid prevention therapy for latent tuberculosis infection worked at the individual level[96].

In contrast, however, some studies have shown otherwise. In a trial from India which enrolled both TST-positive and negative individuals, no benefit from isoniazid preventive therapy was reported[98]. Similarly no benefit was shown in a large community-wide trial to reduce tuberculosis incidence in high HIV settings in South Africa[99]. In yet another study, no long-term effect on tuberculosis prevalence was noted and incidence returned to baseline levels after the isoniazid prevention trial treatment period ended[99]. This study demonstrated that isoniazid prevention therapy failed to have a population-level effect and this was thought to be due to the absence of durable effect of preventive therapy beyond the on-treatment period. Similarly, the Botswana trial where 36 months of isoniazid preventive therapy reduced tuberculosis incidence by 43%, an increase in tuberculosis incidence was observed after termination of preventive therapy even in the presence of ART[100]. Although the degree of protection varies from study to study the benefit of preventive treatment cannot be denied.

Implementing prevention therapy is quite challenging because ensuring completion of treatment is difficult. This is because many patients are reluctant to take medication when they are asymptomatic and maintaining treatment adherence for 9 months is particularly challenging. Although WHO recommends the use of isoniazid preventive therapy as one of the strategies to reduce the tuberculosis burden there is still debate about the benefits of preventive therapy.

2.9 Effect of gender on latent tuberculosis infection

Observed in many parts of the world is that tuberculosis is more a disease of men than of women[1]. The higher rates in men are consistent across different sources of data, including case detection, notification and prevalence surveys[27,101]. Current prevalence data suggests that men are more likely to be diagnosed with active tuberculosis than women, with a male-to-female ratio of 2:1 to 3:1

globally[1,27]. The existence of gender differences in tuberculosis disease is well documented with variations observed across geographical regions and age groups[27,101]. Biological function provides some explanation for the differences although there is also an interactive effect with risk and exposure[102,103].

Once infected, individuals with *M. tuberculosis* move through a two-stage process. Following exposure and subsequent infection, individuals move through a latent stage which could progress to a second stage, active disease stage[3,4]. Not all individuals exposed to *M. tuberculosis* will be infected and neither all who are infected will progress to active disease. In an attempt to understand why differences in tuberculosis rates occur, we will be seeking to clarify whether sex differences in tuberculosis infections are only a result of reporting bias or accurately describe the epidemic.

2.10 Cross sectional studies that have shown sex differences in latent tuberculosis infection

Cross sectional studies have shown sex differences in the prevalence of latent tuberculosis infection[3,4]. A study in Eastern China evaluated the prevalence of latent tuberculosis infection and associated risk factors and found prevalence of latent tuberculosis infection was higher among men than women and this increased with increasing age[104]. In yet another study conducted in Taiwan, Ting showed a similar trend of men having a higher prevalence of latent tuberculosis infection[105]. In a study comparing the efficacy of the tuberculin skin test and QuantiFERON®-TB GOLD to diagnose latent tuberculosis infection, Seung found that the occurrence of positive TST was significantly higher in males than in females[106]. In a cross sectional study among health care workers in a South African hospital, a positive TST was more likely in males than female healthcare workers[107]. Another cross-sectional survey among adolescents in rural eastern Uganda,

Mwanja and colleagues showed that being a male was a significant risk factors for latent tuberculosis infection among the adolescents[108]. Although these studies consistently show that prevalence of latent tuberculosis infection is higher among males than women, they do not answer the question why and how men are more likely than women to become infected. This dissertation proposal offers the opportunity to answer this question.

CHAPTER 3

TUBERCULOSIS INFECTION AMONG HOUSEHOLD AND NON-HOUSEHOLD CONTACTS PERSONS WITH AND WITHOUT TB DISEASE IN KAMPALA, UGANDA

Robert Kakaire[1], Juliet Sekandi[1], Hanwen Huang[2], Fred Quin[3] and Christopher C. Whalen[1]. To be submitted to *International Journal of Tuberculosis and Lung disease*.

3.1 Abstract

Objective. To investigate differences in risk for latent tuberculosis infection among household and non-household social contacts of *M. Tuberculosis* cases and their matched controls in an urban African setting.

Methods. We performed a cross-sectional study from 2012 to 2016 of social contacts of tuberculosis patients and their controls from Lubaga division, Kampala Uganda. We estimated prevalence risk for latent tuberculosis infection by household status of case and control contacts. The prevalence of infection in contacts of controls represents community infection. The difference in infection between household case and control contacts represents the additional risk associated with being a household member of a TB case.

Results. Of 1050 contacts of TB cases enrolled, 410 (39%) were household contacts while 640 (61%) were non-household contacts and of 1035 contacts of controls, 225 (22%) were household contacts while 810 (78%) were non-household contacts. Prevalence of latent tuberculosis infection was 56% among household contacts of cases and 42.5% among non-household contacts. In control contacts, prevalence of infection was 31.1% among household contacts and 40.7% among non-household contacts.

Conclusion. Prevalence of infection was very high even in households without a TB case. The high prevalence in households without a TB case calls for improved non-household contact investigations to detect infection with the goal of targeted preventive interventions.

3.2 Intoduction

Tuberculosis (TB) kills nearly two million people annually mainly in low and middle-income countries, particularly in sub-Saharan Africa and Asia[109]. Approximately 1.7 billion individuals are latently infected with *Mycobacterium tuberculosis (M. tuberculosis)*, the causative microorganism of TB[9]. People with latent tuberculosis infection are not infectious, but are at risk of developing active disease and becoming infectious[5]. Individuals with latent tuberculosis infection are a reservoir from which new cases of active TB arise, propagating the global epidemic[8].

In settings with high TB burden, control programs typically define high-risk contacts mainly by virtue of them sharing a household with a person with active TB disease (the index case). Household contact investigation is the principal method used to detect additional TB patients and recently exposed persons with latent infection at risk for progression to active TB. The focus on household has been justified by studies that have shown high infection rates in households of TB cases[110-112] and genetic clustering of identical strains in infections within families[113]. Children are at higher risk of acquiring infection from the household index case[17,54,114]. Household contact investigations do not identify and evaluate all contacts of TB patients. The current contact tracing approach has limited success because it does not consider non-household contacts of the index who may equally be at risk of infection.

There is increasing evidence that non-household interaction is an equally important source of community TB transmission. Molecular epidemiologic studies[115-117] and community studies in South Africa[118-120] found high rates of non-household transmission ranging between 60 and 80% [17,121,122] Similarly, a meta-analysis by Martinez and colleagues, determined that although

households represent an area of intense transmission of *M. tuberculosis,* less than 20% of transmission in a population occurs in households[123].

In this investigation of social networks of tuberculosis, we measured the prevalence of infection among social contacts of the TB cases and social contacts of matched controls, and examined the risk for latent tuberculosis infection among household and non-household social contacts in Kampala city, Uganda.

3.3 Methods

Study design and Setting

We performed a cross sectional study using a two stage sampling approach. We identified index TB cases through Lubaga hospital and Kampala capital city authority (KCCA) clinics in Lubaga division and matched them with controls. In the second stage, the index participants enumerated their social contacts who were then approached and asked to participate in the study.

Study Population

Index TB cases and Index controls

TB Index cases were enrolled from Lubaga hospital, Kitebi and Kawaala outpatient clinics all located in Lubaga division. TB cases were diagnosed registered patients of the Uganda National Tuberculosis Control Program (NTLP). We restricted enrollment to TB patients 15 years or older because children are unlikely to transmit disease and are less likely to have independent social contacts outside the home. We excluded TB cases treated for more than a week with appropriate anti-tuberculosis medication and short-term residents. A case of tuberculosis was defined as a patient with signs and symptoms consistent with pulmonary TB who had at least one positive

sputum smear for acid-fast bacilli. We enrolled smear-positive cases of TB because they are most likely to transmit infection to their contacts. Community controls, or comparison group without disease, were residents of Lubaga division matched with the index cases. Matching was based on age, sex and residency.

Social contacts

Social contacts were sampled through index participants. Social contacts are individuals who spend at least four hours of physical interaction with an index participant on a typical day. Contacts of TB cases are herein also referred to as case contacts and contacts of controls referred to as control contacts. These were further categorized as either household or non-household contacts. Household contacts were individuals who had resided within the household of an index participant for the previous 3 months and had eaten meals in the household regardless of their age. Non-household social contacts were identified based on personal relations with the index enumerated through a combination of name and location generators or standard prompts and recent time-frames which helped participants recall their non-household social contacts.

Study Procedures

Study teams consisting of clinical officers, nurses and counselors enrolled participants into the study and administered a census to enumerate both household and non-household social contacts. After obtaining informed consent, study staff administered structured questionnaires and collected demographic and TB risk and symptom information according to standardized study protocol. All participants were offered a TST plus free HIV counseling and testing and results were confidentially returned to the participant and for those found HIV positive they were referred to points of care.

Diagnosis of latent tuberculosis infection was by tuberculin skin test (TST) performed using the Mantoux method 5 Tuberculin Units (TU) of purified protein derivative (Lederle, Tubersol) injected intradermally into the left forearm[72]. The diameter of induration was measured in millimeters using the "ball-point" technique between 48 and 72 hours by trained Ugandan technicians. A pair of digital calipers was used to measure induration and on-going in-servicing was done to reduce digit bias. TST reactivity was recorded as continuous variable in millimeters independently by two readers. A criterion for a positive test of 10mm was used to minimize misclassification due to BCG vaccination[124]. Diagnosis of TB disease was by microscopy with Auramine staining. Culture was performed using the BAC-TEC MGIT 960 Detection System (Becton Dickinson, Sparks, MD). The main outcome in this study was presence of latent tuberculosis infection defined as a TST reaction of 10mm or greater for HIV negative individuals and 5mm or greater for HIV positive individuals.

Statistical Analysis

Data were entered using Teleform optical capture software and analyzed using SAS, version 9.3 (SAS Institute, Cary, NC). The estimated prevalence of infection in different groups was compared using chi-square test. Descriptive analysis using proportions for index participants and their social contacts was conducted. Stratified analysis was done to assess for confounding and effect modification, risk ratios estimated with 95% confidence intervals (CI). We used univariate analysis to calculate unadjusted prevalence ratios (PR) and 95% CI for characteristics associated with household status and latent tuberculosis infection. We then performed a multivariate analysis using Log binomial regression[125] using forward selection model building approach to determine the association between the exposure and the outcome while controlling for

confounders and interaction terms. We selected a log binomial model instead of a logistic regression model because the primary outcome was highly prevalent (>10%).

3.4 Results

Index participants reported 2,303 social contacts of whom 2,085 (90.5%) were located and participated in the study. Those who were not located did not differ by socio-demographic characteristics from those located. Index participants (n=246) were similar regarding the matched characteristics, formal education, and religious affiliation. HIV prevalence was significantly higher among TB cases (Table 3.1). Of the 2,085 social contacts enrolled, 1,050 were case contacts and 1,035 were control contacts. The mean age for case contacts was 23 (SD=13.9) while for control contacts was 25 (SD=11.5). Case contacts were similar to control contacts with regard to HIV prevalence (p=0.55), marital status (p=0.06) and income level (p=0.18). Differences were observed in household status between case contacts and control contacts. Majority of case contacts were female household members. More case contacts worked in enclosed work environments compared to control contacts (Table 3.2).

Prevalence of latent tuberculosis

Prevalence of latent tuberculosis infection in Lubaga division was 50.4% (95% CI; 47.3, 53.5). Infection was highest among household case contacts and lowest in household contacts of controls. Prevalent infection among non-household case contacts was similar to that of non-household control contacts (Figure 3.1). Because sex and age may confound the relationship between household exposure and infection, we performed gender and age stratified analysis. The prevalence of infection among household case contacts was similar in men and women. Among household control contacts, men had higher prevalence of infection than women. Infection among

non-household case contacts was higher in men than women and in non-household contacts of controls infection in men was even higher (Figure 3.1). In household contacts of TB cases, prevalence of latent tuberculosis infection was highest in the 0-5 age category and decreased with increasing age. (Figure 3.3). Interestingly in the non-household contacts of TB cases, prevalence is highest in the 16-25 age category. The trend is however different with contacts of controls with infection being highest in the 16-25 age category for both household and non-household contacts (Figure 3.4).

Univariate and multivariate analysis

In multivariate analysis, two separate models (Table 3.4) for case contacts and for control contacts (Table 3.5) were done to estimate adjusted prevalence ratios (PR) using a log binomial regression model. This was to maintain comparison of risk for infection between the two groups. All significant variables at the univariate level were included in the multivariate analysis. BCG and HIV status were included because of their biological association with latent tuberculosis infection [126]. In contacts of cases, being a household contact was associated with a higher risk for infection compared to non-household contacts after controlling for other factors PR, 1.1.6,95% CI 1.4,2.0; $p<.01$. The effect of age remained significant both as a continuous variable and when analyzed as a categorical variable. Across different age groups, risk for infection increased with age. The effect of HIV status as a risk factor was in the expected direction although remained non-significant at PR 0.7, (95% CI 0.5, 1.2; 0.19 in the final model.

In contacts of controls, as expected, there was no differences in risk for infection by household status since there was no TB case in the household. However, being a male was associated with a higher risk for infection compared to women PR1.1.4, 95% CI 1.2, 1.7; $p<.01$.

As was with case contacts, the effect of age remained significant both as a continuous variable and when analyzed as a categorical variable and across different age groups, risk for infection increased with age. The effect of HIV status as a risk factor was in the expected direction although remained non-significant at PR 0.6, (95% CI 0.4, 1.0; 0.07 in the final model.

3.5 Discussion

To our knowledge, this is the first study to use a non-traditional approach for studying prevalence of latent tuberculosis infection among contacts of TB cases that goes beyond household contact investigations. Our study evaluated the prevalence of latent tuberculosis infection among household and non-household contacts of TB cases and matched controls. As predicted, we observed high prevalence of latent tuberculosis infection (50.4%) in Lubaga division similar to prevalence of 49% previously reported by a previous study done in this same setting in 2015[127]. In households of TB cases, we found even higher prevalence at (56%) and lower in households of control contacts prevalence at (31%). Our study shows that residing in the household of a TB case is significantly associated with latent tuberculosis infection as previously reported by other studies in this same region[67,128,129]. What is striking though is that we found high prevalence of infection in households without a TB case. Among contacts of cases, after adjusting for other risk factors, being a household contact of a TB case was associated with a 70% increased risk compared to non-household contacts.

Infection among non-household contacts of cases was 42.5% (95% CI 38.7, 46.4) while that of non-household contacts of controls was 40.7% (95% CI 7.4, 44.2) suggesting that being a non-household contact of a TB case does not appear to confer much additional risk beyond that

encountered in the community rendering yet more evidence of transmission occurring outside households of TB cases as previously reported[118,119,130].

The high prevalence of infection in households without TB cases may be TB attributed to a weighted average of prevalence from children with adults in the household. Non-household prevalence of infections was comparable in both case and control contacts at 42.5% and 40.7% respectively. Exposure to a household TB case is well established risk factor for latent tuberculosis infection [67,128,129] and our results are consistent with other household contact investigations which have shown the household as an area of intense transmission of *M. tuberculosis*[131,111].

Although prevalence among household contacts may vary between communities in the same setting, similar high burden was found in Lubaga division Uganda by Kizza and colleagues[18]. Exposure within and outside the household is complex and may be confounded by sex, age as well as other factors[111,132,133]. However, in our study, even after controlling for age and sex, a statistically significant association between household status and prevalent infection was still observed among case contacts.

Published studies on male gender and increased risk for latent tuberculosis infection are conflicting. A higher latent tuberculosis infection prevalence among males was observed in a rural area of Ethiopia[134] and a Peruvian peri-urban shantytown[128] but not in Kawempe, Uganda[127] nor in South Africa studies[64,135,136]. In our study being male was associated with a 40 percent increased risk for TB infection compared to women among control contacts. The high risk for TB infection among men could be partly explained by social mixing outside the household. Men are usually mobile away from their homes and depending on the nature of their daily activities may offer opportunities for frequent contact with infectious TB cases who may not be in their households. Moreover because tuberculosis disease is prevalent in less than 1% of households in a community

at any time, exposure opportunities between a person with tuberculosis and their social network outside the household are more numerous, influencing the proportion of *M. tuberculosis* transmission that occurs in the community[123]. Although household contact status increases the risk for infection, it appears that its effect on the overall burden of disease in a community may be limited.

Risk for latent tuberculosis infection increased with age in both case and control contacts but the magnitude was higher among control contacts. The increasing prevalence of latent tuberculosis infection with age observed in both contacts of cases and controls reflects the cumulative exposure to TB in high TB burden settings [16,137] and is consistent with findings of other studies in urban populations[64,127,138,139].

Limitations

Results of this study should be interpreted in light of some limitations. The use of TST alone for diagnosis of latent tuberculosis infection could have led to misinterpretation because of false positive and false negative results due to wide BCG vaccination coverage, HIV infection and malnutrition[124]. To minimize misclassification due to BCG vaccination, we used 10mm as a criterion for positive TST[140] and for the HIV positives, we used a cut off of >= 5mm. Boosting of the TST, as the result of repeated testing, has been found to be a potential cause of false-positive TST conversion, especially in BCG-populations[81]. We assessed this possibility in this community and found that boosting occurred in only 2% of participants[86]. Although there is good correlation between TST and IGRA in low TB endemic settings[141,142], high levels of discordance have been reported in some high TB endemic areas including Uganda[143]. It is therefore unlikely that the use of IGRAs would have changed the results we obtained in this study population.

We limited our investigation to first degree contacts and moreover due to recall bias it is likely that the list of social contacts that were enumerated was not exhaustive and this may have under estimated the prevalence of infection especially among non-household social contacts. Interaction outside the household is complex and due to social mixing, there is heterogeneous mixing of household and non-household contacts. Non-household case contacts could double as household or non-household control contacts and thus it's not feasible to exclusively classify social contacts as being case or control contacts especially when they are all from the same community. Contacts of controls may have had prior household exposure that was not captured at the time of interview moreover some household infections may have resulted from infection outside the household and there potentially may have been misclassified.

3.6 Conclusion

In conclusion, this is yet another evidence of risk for infection outside of the TB home in high burden settings. Therefor more studies are needed to guide the use of contact investigations outside of the household by specifying which particular locations offer the highest risk for TB infection.

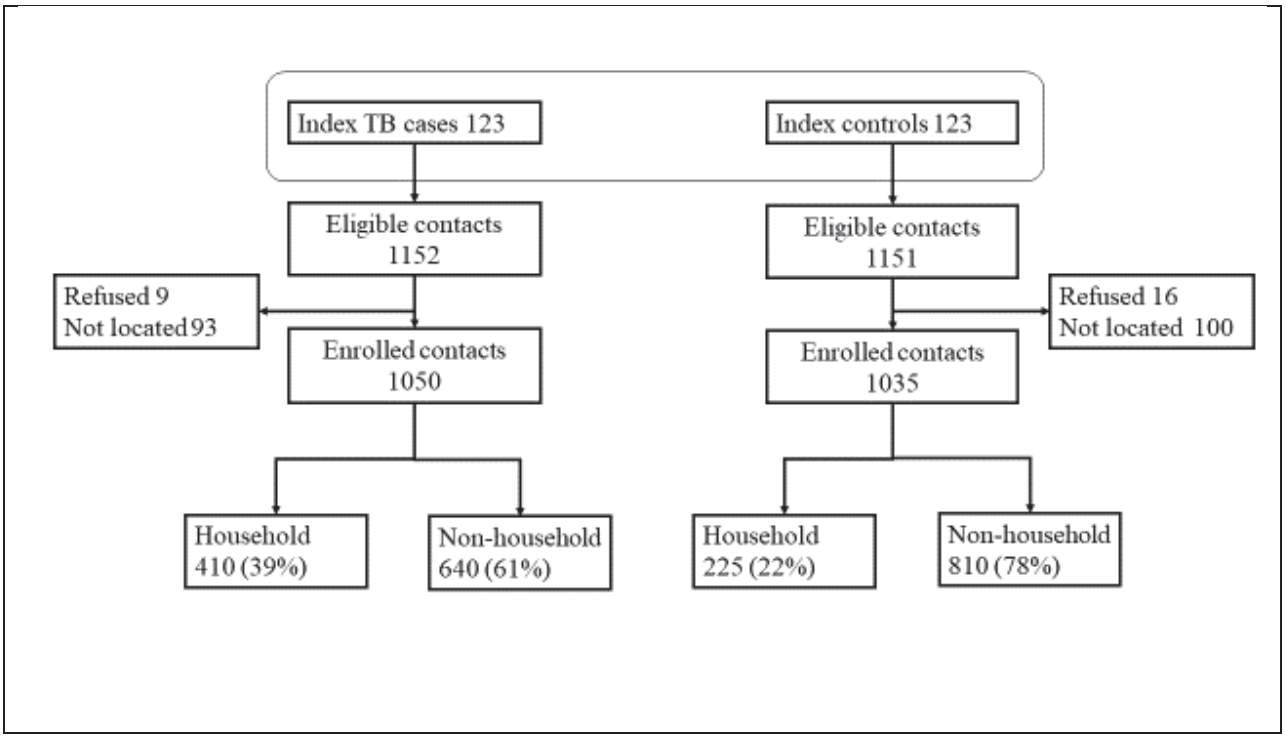

Figure 3. 1. Study profile showing number of index participants and social contacts enrolled

Table 3.1: Demographic Characteristics of Index TB Cases and Matched Controls

Characteristics		**TB Cases N=123 (%)**	**Matched Controls N=123 (%)**
Sex	Male	84	84
	Female	39	39
Age Category	15-19	07 (5.69)	03 (2.44)
	20-29	56 (45.53)	55 (44.72)
	30-39	36 (29.27)	36 (29.27)
	40+	24 (19.51)	29 (23.57)
HIV Status	Positive	21 (17.08)	06 (4.88)
	Negative	98 (79.67)	115 (93.50)
	Refused	04 (3.25)	02 (1.63)
Marital status	Married	41 (33.33)	66 (53.66)
	Not married	82 (66.67)	57 (46.34)
	Religion Christians	82 (66.66)	89 (72.36)
	Muslim	26 (21.14)	25 (20.33)
	Other	15 (12.20)	08 (7.31)
Education	None	08 (6.50)	06 (4.88)
	Primary level	42 (34.15)	49 (39.84)
	Post primary level	73 (59.35)	68 (55.28)
Income (UG SHS / US $)			
	<200,000 / (<80)	92 (74.80)	75 (60.98)
	≥ 200,000 / (≥ 80)	31 (25.20)	48 (39.02)
Work environment			
	Indoors	50 (40.65)	28 (22.76)
	Semi-Indoors	38 (30.89)	48 (39.02)
	Outdoor	29 (23.58)	39 (31.71)
	No response	06 (4.88)	08 (6.5)

Table 3.2: Demographic Characteristics of Social Contacts of TB Cases and Matched Controls

Characteristics	**Case contacts N=1050**	**Control contacts N= 1035**
Sex		
Male	519 (49.43)	582 (56.23)
Female	531 (50.57)	453 (43.77)
Age category		
0-5	195 (18.57)	133 (12.85)
6-15	188 (17.90)	151 (14.59)
16-25	347 (33.05)	428 (41.35)
26-35	175 (16.67)	219 (21.16)
36-45	95 (9.05)	83 (8.02)
46+	50 (4.76)	21 (2.03)
HIV Status		
Positive	77 (8.45)	65 (6.64)
Negative	822 (90.23)	902 (92.13)
Refused	12 (1.32)	12 (1.23)
Marital status		
Married	368 (35.05)	405 (39.13)
Not married	682 (64.95)	630 (60.87)
Religion		
Christians	638 (60.76)	701 (67.73)
Muslim	267 (25.43)	244 (23.57)
Other	145 (13.81)	90 (8.70)
Education		
None	167 (15.90)	123 (11.88)
Primary	374 (35.62)	424 (40.97)
Post primary	509 (48.48)	488 (47.15)
Household status		
Household	410 (39.05)	225 (21.74)
Non-household	640 (60.95)	810 (78.26)
Income (UG SHS/ US $)		
<200,000 / (<80)	772 (73.52)	728 (70.34)
≥ 200,000 / (≥ 80)	267 (25.43)	299 (28.89)
No response	11 (1.05)	08 (0.77)
Work environment		
Indoors	519 (49.43)	379 (36.62)
Semi-Indoors	274 (26.10)	302 (29.18)
Outdoor	176 (16.76)	283 (27.34)
No response	81 (7.71)	71 (6.86)

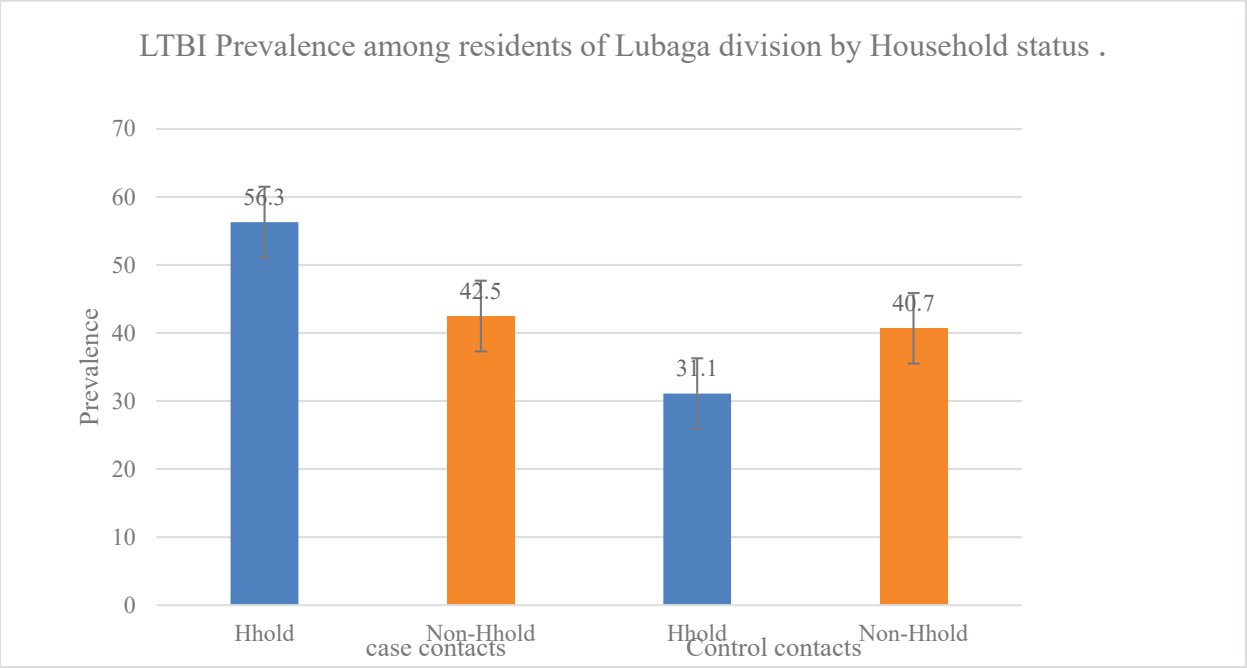

Figure 3.2: Latent Tuberculosis Infection Prevalence among Case and Control Contacts by Household Status

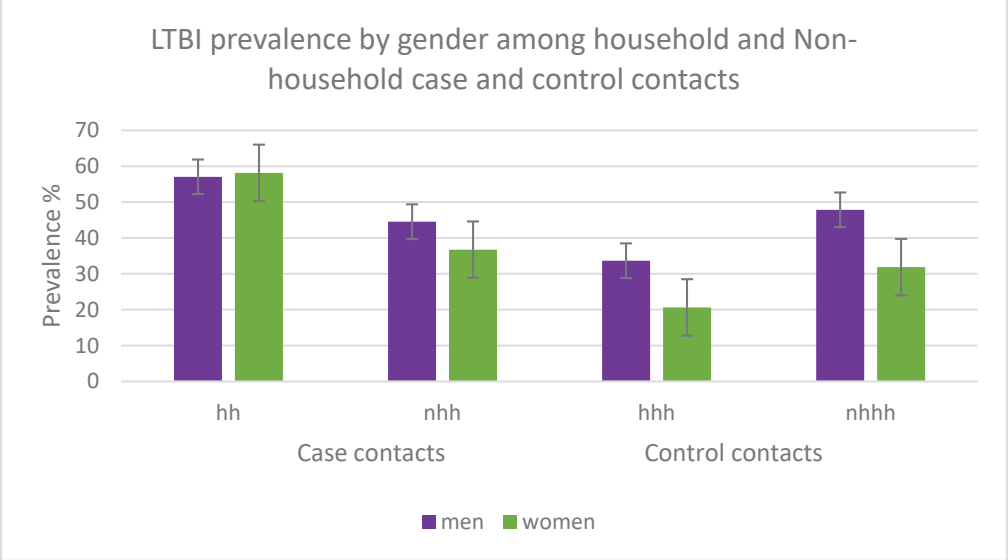

Figure 3.3: Prevalence of Latent Tuberculosis Infection by Sex among Household and Non-household Case and Control Contacts

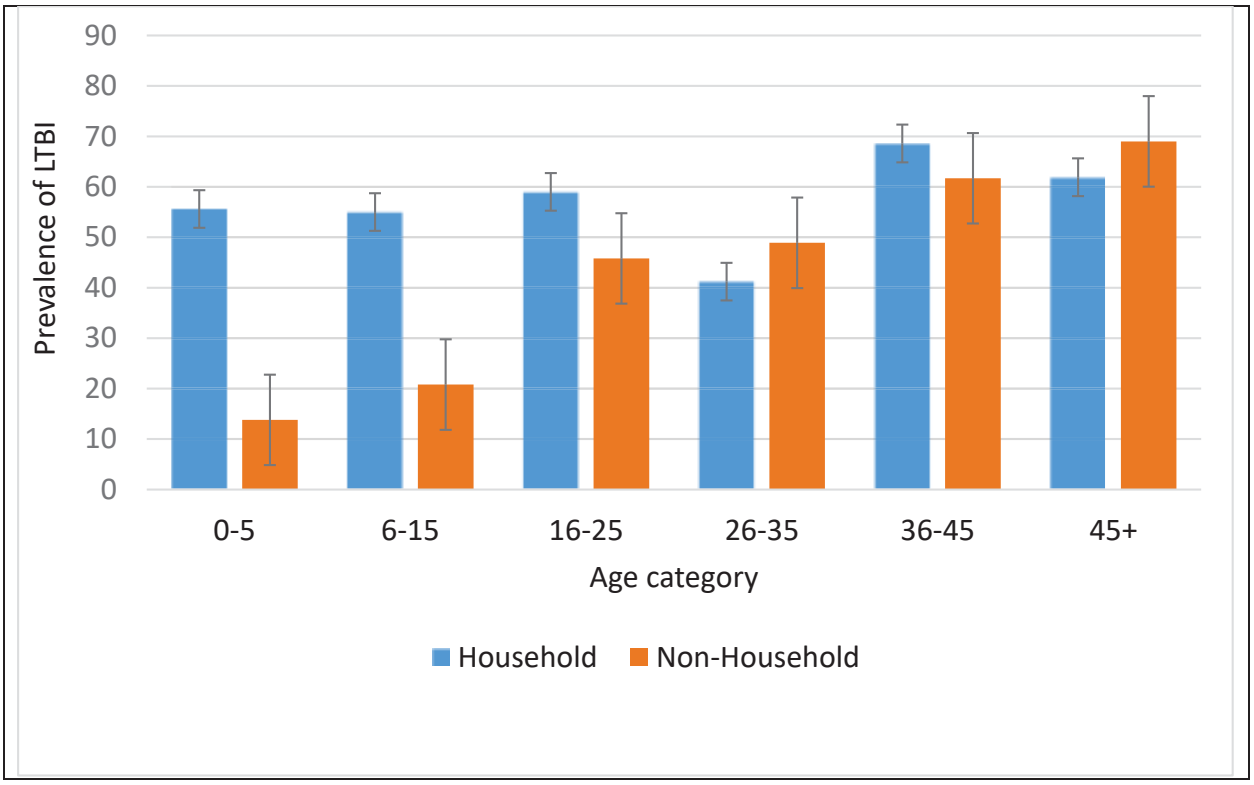

Figure 3.4: Prevalence of latent tuberculosis infection by household status according to age category among contacts of TB cases

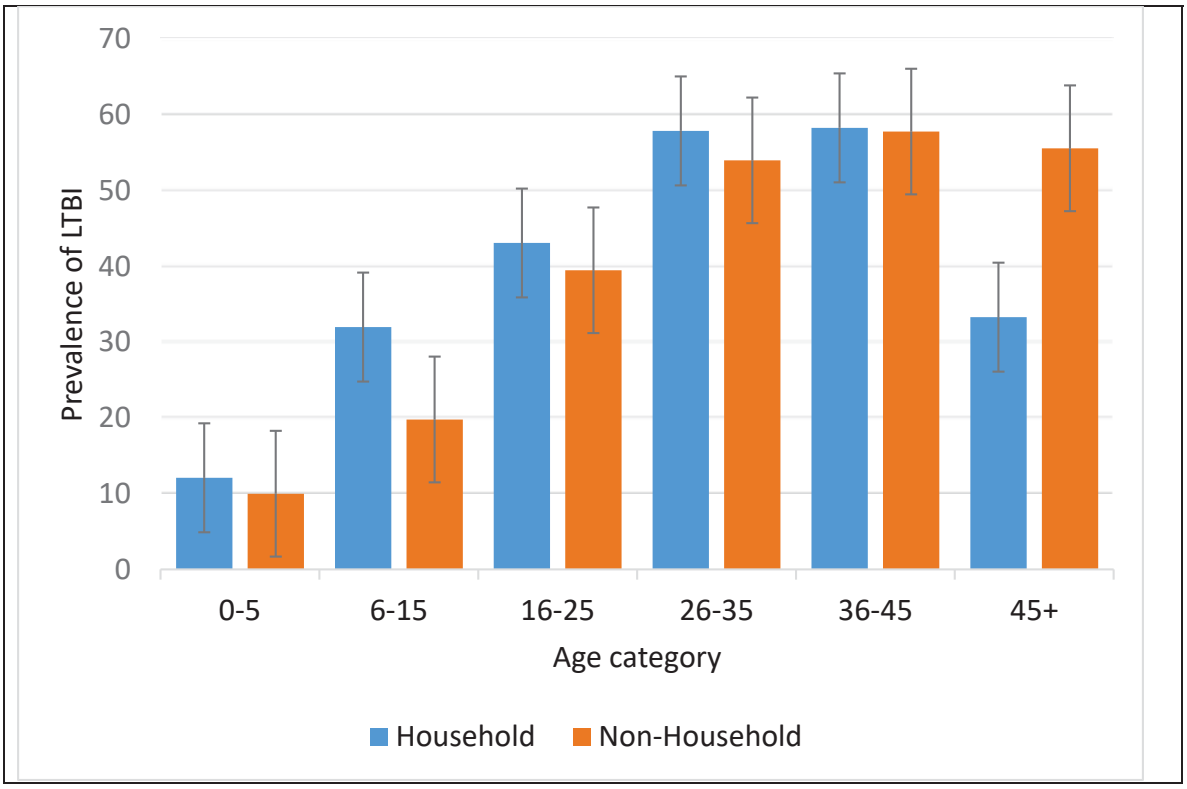

Figure 3.5: Prevalence of Latent tuberculosis infection by household status according to age category among contacts of controls

Table 3.3: Univariate Log Binomial analysis of Latent Tuberculosis Infection According Household status

Category	N	Prevalence tuberculosis infection		Prevalence ratio (log-binomial)
		n	% (95% CI)	
Household Status				
Non household	1406	602	42.8 (40.3,45.4)	1
Household	606	301	49.7 (45.7,53.6)	**1.2 (1.1,13**)
Contact status				
Control	1014	400	39.5 (36.5,42.5)	1
Case contact	998	503	50.4 (47.3,53.5)	**1.2 (1.1,1.3)**
Sex				
Female	984	378	38.4 (35.4,45.5)	1
Male	1101	525	47.7 (44.8,50.6)	**1.2 (1.1,1.4)**
Age category				
0-5	328	94	28.7 (24.0,33.8)	1
6-15	339	113	33.3 (28.5,38.5)	1.2 (0.9,1.5)
16-25	775	341	44 (40.5,47.5)	**1.6 (1.3,1.9)**
26-35	394	202	51.2 (46.4,56.2)	**1.8 (1.5,2.2)**
36-45	178	109	61.2 (53.9,68.1)	**2.1 (1.7,2.6)**
45+	71	44	62 (50.3,72.4)	**2.2 (1.7,2.9)**
HIV Status				
Negative	1860	1003	53.9 (51.7,56.2)	1
Positive	169	66	39.1 (32.0,46.6)	0.6 (0.1,3.4)
Refused Test	56	06	10.7 (4.7,21.8)	0.9 (0.6,1.5)
Alcohol Use				
No	1435	597	41.6 (39.1,44.2)	1
Yes	581	289	49.7 (45.7,53.8)	**1.2 (1.1,1.3)**
Smoking status				
Never Smoked	1827	774	42.4 (40.1,44.6)	1
Past Smoker	76	47	61.8 (50.6,72.0)	**1.4 (1.1,1.7)**
Current smoker	116	67	57.8 (48.7,66.4)	**1.4 (1.1,1.6)**
BCG status				
Vaccinated	1729	749	43.3 (41.0,45.7)	1
Not vaccinated	187	87	46.5 (39.5,53.7)	1.0 (0.8,1.3)
Don't know	111	54	48.7 (39.5,57.8)	1.0 (0.8,1.3)
Ever stayed in a home with a TB case				
No	1262	532	42.2 (39.5,44.9)	1
Yes	435	256	58.9 (54.2,63.4)	**1.4 (1.2,1.5)**
Uncertain	37	14	37.8 (24.0,53.9)	0.9 (0.6,1.4)
Lived in same home with TB case in last 12 months				

Category	N	Prevalence tuberculosis infection		Prevalence ratio (log-binomial)
		n	% (95% CI)	
No	1057	439	41.5 (38.6,44.5)	1
yes	648	352	54.3 (50.5,58.1)	**1.3 (1.1,1.4)**
Uncertain	28	12	42.7 (26.5, 61.0)	1.0 (0.7,1.6)
Occupation				
Out doors	423	218	51.5 (46.8,56.3)	1
Semi-Indoors	532	237	44.6 (40.4,48.8)	**0.9 (0.8,1.0)**
Indoors	843	326	38.7 (35.4,42.0)	**0.8 (0.6,1.0)**

Table 3.4: Multivariate Log Binomial Analysis of Latent Tuberculosis Infection Among Case Contacts According to Household Status Controlling for Other Factors

variable	N (n)	Multivariate	
		PR (95%CI)	P-value
Household Status			
Non household	410 (212)	**1**	
Household	640 (236)	**1.6 (1.4,2.0)**	**<.01**
Age			
0-5	195 (76)	1	
6-15	188 (73)	1.2 (0.8,1.8)	0.32
16-25	347 (156)	**1.9 (1.4,2.7)**	**<.01**
26-35	175 (69)	**1.8 (1.2,2.7)**	**<.01**
36-45	95 (45)	**2.3 (1.6,3.3)**	**<.01**
45+	50 (29)	**2.4 (1.6,3.5)**	**<.01**
HIV Status			
Positive	37 (12)	1	
Not positive	1013 (436)	0.7 (0.5,1.2)	0.19

n=number with latent tuberculosis infection

Table 3.5: Multivariate Log Binomial Analysis of Latent Tuberculosis Infection Among Control Contacts According to Household status Controlling for Other Factors

variable	N (n)	Multivariate	
		PR (95%CI)	P-value
Household Status			
Non household	225(66)	1	
Household	810(314)	1.0 (0.7,1.5)	0.80
Sex			
Female	582(251)	**1**	
Male	453(129)	**1.4 (1.2,1.7)**	**<.01**
Age			
0-5	133(13)	**1**	
6-15	151(32)	**2.4 (1.3,4.3)**	**<.01**
16-25	428(167)	**4.0 (2.3,7.0)**	**<.01**
26-35	219(144)	**5.7 ()3.3,10.0**	**<.01**
36-45	83(43)	**5.5 (3.0,10.0)**	**<.01**
45+	21(11)	**4.5 (1.7,11.8)**	**<.01**
HIV Status			
Positive	39(11)	1	
Not Positive	1046(369)	0.6 (0.4,1.0)	0.07

CHAPTER 4

DIFFERENCES IN INCIDENCE RATE OF LATENT TUBERCULOSIS INFECTION BETWEEN MEN AND WOMEN AND ASSOCIATED RISK FACTORS. RESULTS FORM AN INCIDENT COHORT IN URBAN AFRICA, UGANADA.

Robert Kakaire[1], Juliet Sekandi[1], Hanwen Huang[2], Fred Quin[3] and Christopher C. Whalen[1]. To be submitted to *International Journal of Tuberculosis and Lung disease*.

4.1 Abstract

Objective. To investigate differences in incidence rate of latent tuberculosis infection between men and women among community residents of Lubaga division, Kampala Uganda.

Methods. We performed an incident cohort study from 2014 to 2016 of community residents of from Lubaga division, Kampala Uganda. We estimated incidence rate by sex and assessed risk factors for incident latent tuberculosis infection. The difference in incident rate between men and women may partly explain the observed high prevalence of latent tuberculosis infection in this high endemic African setting.

Results. We enrolled 1600 volunteers (40.3% men Vs 59.7% women) at baseline. A total of 1249 (78%) participants were located at 12 months for follow up evaluations of whom 514 (41.2%) were men and 735(59.8%) women. One hundred ninety-two (12%) had converted their TST. Of these 192 converters, 99 (51.6% were men representing an incidence rate of 16.2 per 100PYO while 93 (48.4%) were women representing an incidence rate of 10.6 per 100PYO.

Conclusion. Incidence of latent tuberculosis infection was consistently higher among men across all age groups. We think men may have different social networks that lead to a greater exposure to the mycobacteria. These results have implications for contact investigations, whereas not ignoring women, should be greatly strengthened for men.

4.2 Introduction

Tuberculosis is among the leading causes of death worldwide (WHO Global Tuberculosis report 2017). It is estimated that about 25 percent of the world's population has latent tuberculosis infection [2], 10% of whom will develop active disease in their lifetime[144]. Sex differences in the epidemiology of active disease have been well-described in different studies[26,145-147]. Current prevalence data suggests that men are more likely to be diagnosed with active disease than women[27], with a male-to-female ratio of 2:1 to 3:1 globally[1]. While the association between sex

and active disease has been well studied, the relationship between sex and latent tuberculosis infection is less clear[148]. Male sex has been identified as an independent risk factor associated with latent tuberculosis infection in some studies[132,149], but other studies also report insignificant correlation between sex and latent tuberculosis infection [150-152].

Following exposure and subsequent infection, individuals with *M. tuberculosis* move through a two-stage process; the latent stage which could progress to a second stage of active disease. Differences in active disease observed between men and women may be due to differences in the risk of infection once exposed or differences in the risk for progression to active disease once infected, or a combination of both. We conducted an incident cohort study between 2015 and 2017 to investigate whether differences in the prevalence of latent tuberculosis infection among men and women observed in aim one of this dissertation is as a result of differences in the incidence of latent tuberculosis infection.

4.3 Methods

Study Design and Setting

This was a prospective cohort study designed to estimate the incidence of latent tuberculosis infection. Participants with a TST readining of <5mm at baseline were enrolled and later a repeat TST was placed to determine tuberculin skin test conversion. A research team comprising a medical doctor, clinical officers, nurses and social scientists based at Lubaga hospital was responsible for the implementation of the study protocol. Community mobilization and health education was done by study home visitors who had received in-service about the study protocol. Baseline evaluation verified the absence of tuberculosis disease through a symptom survey and review of medical records. If tuberculosis disease was ruled out, then risk assessment

for tuberculosis infection and disease was done. Home visitors collected demographic and risk factors from study participants using structured questionnaires. Participants were screened for clinical symptoms of TB through a limited medical history after which tuberculin was placed for a Tuberculin skin test. If a participant had suspicious symptoms of TB disease, they were referred for further TB evaluation at Lubaga hospital.

Study Population

Residents of Lubaga division of Kampala without evidence of previous infection with *M. tuberculosis* were eligible for enrollment. Study participants were screened for LTBI and recruited to the cohort in two ways. One source was a concurrent, cross-sectional study of contacts of tuberculosis cases. A second source was a community survey screening for latent tuberculosis infection using the TST. Individuals from these studies that met the inclusion criteria, and none of the exclusion criteria were approached to give consent for the study. Details in study profile Figure 4.1 below.

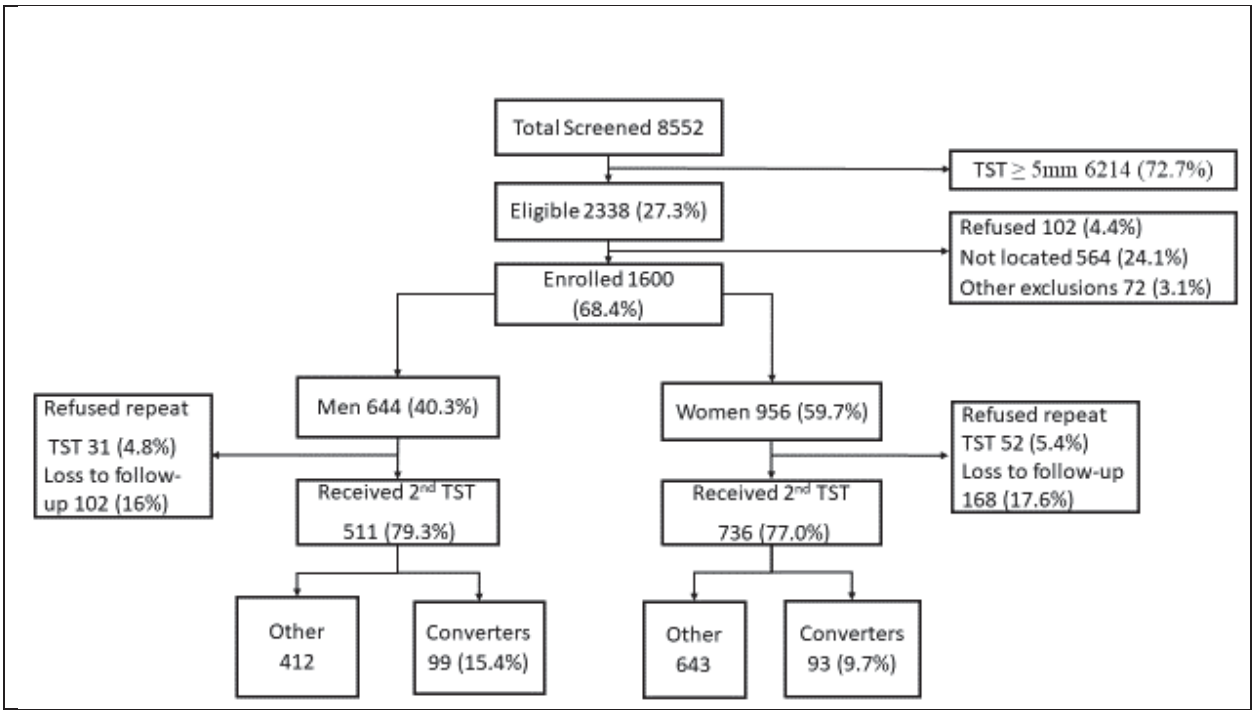

Figure 4.1. Cohort study participant flow diagram

Inclusion criteria were residency in Lubaga, TST < 5mm, age 15-65 years and willingness to take an HIV test and exclusion criteria were refusal to consent and transients. We considered TST in-duration of < 5mm at baseline as negative for HIV negative individuals[81]. For HIV positive individuals in whom immunosuppression may be present, we used a cut off of 3mm[81]. Absence of infection was defined using this tuberculin skin cut off criteria because this criterion is highly sensitive for latent tuberculosis infection, even in HIV infected persons[81]. A resident was defined as a person who had stayed in Lubaga division for more than 1 week in the previous 3 months and had intentions to stay for the next 12 months. The age range 15-65 was chosen because the prevalence of latent tuberculosis infection increases beginning in late adolescence and also this marks a risk period in the lifetime of urban Africans as younger individuals begin to spend more time out of the household or out of school[153].

Study Outcome

Latent tuberculosis infection was diagnosed using the tuberculin skin test performed using the mantoux method 5 of purified protein derivative placed intradermal on the left forearm [154]. The diameter of induration was measured in millimeters using a pair of digital calipers between 48 and 72 hours by trained technicians. Two readers independently recorded the tuberculin skin test reactivity as continuous variable in millimeters. Two criteria, one for HIV positive and the other for HIV negative individuals was used for determining conversion at follow-up. A converter was defined as one with a follow up TST induration of >= 10mm AND a 6mm increase over the baseline among HIV negative individuals[155]. For HIV positive individuals, whose baseline TST was < 3mm, a converter was defined as one with a follow up TST induration of >= 6mm AND a 3mm increase over the baseline. latent tuberculosis infection incidence rate (IR) was defined as the number of new cases of latent tuberculosis infection occurring per 100 person-years of observation.

Study exposure, confounders, and covariates

The exposure variable was biological sex which is a binary variable defined as male or female. Individuals enrolled in the cohort study were evaluated at baseline and quarterly over a period of two years. Individual risk assessment for tuberculosis infection and disease plus a risk site questionnaire to determine the frequency of exposure to high transmission settings was administered.

Statistical Analysis

Baseline demographic characteristics in the study participants were compared using chi-square tests to assess differences between men and women. Stratified analysis of demographic and risks factors was done and assessed for confounding and effect modification (PROC FREQ, SAS v 9.1

SAS Institute, Cary NC). We performed univariate log binomial regression analysis to identify independent predictors of infection. We used SAS PROC GENMOD'S log binomial regression because of its efficient maximum-likelihood-based parameters and moreover the frequency of our outcome measure was observed to be higher thn 10%. All predictors with statistically significant associations at the univariate level were included in the multivariate analysis one at a time dropping those with non-statistically significant associations. No interaction was found between variables. The final model was fitted based on one with the smallest AIC. Statistical analysis was performed using SAS, version 9.1 (SAS Institute, Cary, NC, USA).

Ethical considerations

The study was approved by the University of Georgia institutional review board, the research ethics committee at Makerere University School of Public Health and approved by the Uganda National Council for Science and Technology. All study participants gave individual written informed consent and for minors aged 15-17 they gave assent after obtaining consent from their guardians.

4.4 Results

Description of baseline characteristics

We enrolled 1,600 volunteers (40.3% men Vs 59.7% women) all with a baseline TST reading of less than 5mm. The mean age was 25 years with majority being in the 25-34 age category. The majority of participants were not married, worked indoors or semi indoors, had completed post-primary education, and earned between 50,000 and 200,000 Uganda shillings per month. HIV prevalence in the study population was 5.8% (Appendix Table.1). A total of 1,249 (78%) participants were located for a repeat TST and follow up evaluations of whom 514 (41.2%) were

men and 735(59.8%) women. Those who were lost to follow up did not differ by gender from those located. During follow up, a total 1503.4 person-years observation were recorded.

Final outcome

Among the 1,600 participants, 192 TST converted over 1503.4 PYO leading to an overall incidence rate (IR) of 12.8 per 100 PYO. Of those who converted, 99 (51.6%) were men, representing an incidence rate of 16.0 per 100 PYO, whereas 93 (48.4%) were women representing an incidence rate of 10.6 per 100 PYO. One thousand eighteen individuals, 1018 (64%) were persistent negatives while 39 (2.4%) were indeterminate. 351 (21.9%) study participants were lost to follow-up.

Latent tuberculosis infection incidence rate among men and women was stratified by demographic characteristics and a rate ratio calculated for each demographic variable (Table.4.1). Overall, men were 1.6, 95% CI 1.2,2.0 times more likely to be infected than women. Risk for incident infection was consistently higher among men across most demographic variables and this risk was comparable to the overall risk. Across all age categories, the incidence rate increased with age for both men and women but the magnitude was higher for men. Being married was a risk factor for infection for both men and women although this risk was marginally higher among men.

Because the probability for infection is influenced by degree of contact, we performed a stratified analysis of infection and exposure to a TB index case. Individuals who reported to have been in contact with a TB case i.e.: knew someone with TB, ever lived in the same house as TB case, lived in the same house with TB case in the last 12 months, shared the same bedroom with a TB case in the last 12 months, and spent ≥4hrs with a TB case in the last 12 months all had a higher risk for infection compared to those who did not report contact and moreover the risk was comparable to

the overall risk (Table 4.3). However, comparison by risk factor between men and women appear to be at a higher risk for infection than men (Table 4.4). Multivariate analysis revealed that the relative risk of new infection was 60% higher among men compared to women (RR1.6, 95% CI 1.2, 2.0) and this risk increased with age (Table.4.7). The other covariate associated with infection was sharing a bedroom with a TB case.

4.5 Discussion

We hypothesized that following exposure, infection was greater among men compared to women because surveys show higher prevalence of latent tuberculosis infection in men than women[27,101]. In this study, we found that men were 60% more likely to acquire new *M. tuberculosis* infection than women, and this risk was not confounded by age or HIV infection. The higher incidence of infection was consistently higher across most demographic groups. We did find, however, that the risk of infection was similar between men and women when participants describe previous or recent contact with a known case of tuberculosis.

But another hypothesis may be true where the probability for infection once exposed is equal among men and women but progression to disease once infected may be greater in men than women. This hypothesis is valid but harder to evaluate. Prospective cohort studies would be the best to validate this hypothesis but need thousands of volunteers to be observed for a long time. There are other possible theories but for purposes of this dissertation we focus on the first one which was validated scientifically.

Two major, but not mutually exclusive hypotheses have been suggested to explain differences in infection by sex: the physiological and the behavioral factors[156]. The physiological hypothesis suggests that biological differences between the sexes render one more susceptible to

a given disease, while the behavioral hypothesis relates to sex-specific exposure to infection. This study offers evidence for behavioral theory in that gender may affect *M. tuberculosis* exposure because of differences in social roles, risk behaviors, and activities[144]. Males may travel more frequently, may have more social contacts than women; spend more time in settings that may be conducive to transmission, such as bars[144,157,158].

Our results showed the incidence rate for infection increased with age for both men and women a trend similar to that observed in prevalence studies[111,127,159,160]. Conversely, the incidence rate ratio was consistently higher for men compared to women across all age groups. This trend though not surprising could be explained by within sex preferential mixing[161]. A study of social assortment patterns in sub-Sahara Africa by Todd et al, showed evidence of preferential mixing of close contacts within age and sexes[161]. Because men are more likely to be diagnosed with TB disease[1,27] and therefore more likely to be infectious, the preferential mixing could explain the higher incidence rate among men than women.

Risk for infection was higher among those exposed to the index TB case compared to the non-exposed but the risk was not different between men and women. Close contact is a well described risk factor for TB infection[58,113,162-164]. In a systematic review by Fox and colleagues, the prevalence latent TB infection among TB contacts was 51.5% [162]. Close contact was estimated to account for 9-13% of the TB cases in Malawi[113].

A relative risk of 1.7 (1.3, 2.2) compared to those not exposed represents a substantial risk. It follows therefore that TB prevalence among contacts is very high further highlighting the importance of close TB contact investigations[164]. This factor however, is not well studied among

non-household contacts in most TB programs. We believe that inquiring about one's exposure to a TB case would refine targeting of non-household TB screening strategies.

Annual risk of tuberculosis infection (ARTI) estimated from repeated tuberculin surveys of children are commonly used to infer community transmission across age groups[165-167]. Using a direct measure of incidence rate, we have shown that incidence rates are considerably higher in men than women. Male gender is a risk factor for incident infection across all age groups 15 years and older in this high endemic African population. In Africa, a number of tuberculin surveys have been published in the recent past with ARTIs between 0.3% and 3.1% reported[165-167]. However, it is very likely that no recent direct measure of incidence estimates are available for Uganda for comparison moreover estimates in adults are often higher than those in children partly because of life time exposure and waning protection from BCG[161].

Men may be more susceptible to infection because of differences in their social networks and risk behaviors[168]. Males may be having different social networks from those of women and moreover due to sex assortment men may be spending more time in settings that may be conducive for transmission to occur, such as bars[169]. Adult males particularly above the age of 35 years had the highest risk for latent tuberculosis infection. This is consistent with a number of prevalence studies that have consistently shown higher prevalence of latent tuberculosis infection especially among older men[132,170-172]. We also found that a history of smoking, alcohol use, and HIV positive status were associated with incidence of latent tuberculosis infection. Smoking has previously been shown to be a risk factor for TB infection[173,174].

Strengths and limitations of our study

This is the second study (the other being in Malawi) to estimate incidence of *M. Tuberculosis* infection using direct methods that rely on following up of initially uninfected individuals over time. Moreover, this incident cohort is located in an endemic area with limited preventive therapy routinely offered. One limitation of this study though, relates to the TST, a non-specific test that may cross-react with mycobacterial antigens from BCG or environmental bacteria. In theory, IGRAs address this concern by using antigens that are more specific to *M. tuberculosis*, but performance of IGRAs in Africa is variable[48].

4.6 Conclusion

Our study observed differences in incidence rates between men and women. This may explain, in part, why men have a higher prevalence of latent tuberculosis infection than women. We propose that men may have different social networks that lead to a greater exposure to the *M.tuberculosis*. Differences in social contacts may be even more impactful as individuals grow older because as individuals grow they spend more time in the community interacting with community individuals. These results have implications for contact investigations, whereas not ignoring women, should be greatly strengthened for men.

Table 4.1: Baseline Demographic Characteristics of Cohort Participants by Sex

Characteristics	Males N=644 (40.3%)	Females N =956 (59.7%)
Age category		
15-24	347 (53.9)	515 (53.9)
25-34	200 (31.1)	324 (33.9)
35-44	71 (11.0)	92 (9.6)
45+	26 (4.0)	25 (2.6)
HIV Status		
Positive	28 (4.4)	65 (6.8)
Negative	569 (88.3)	822 (86.0)
Not done	47 (7.3)	69 (7.2)
Marital status		
Married	246 (38.2)	507 (53)
Not married	398 (61.8)	449 (47)
Religion		
Catholics	269 (41.8)	375 (39.2)
Anglican	166 (25.8)	250 (26.2)
Muslim	156 (24.2)	212 (22.2)
Other	53 (8.2)	119 (12.4)
Education level		
Primary	210 (32.7)	286(30.0)
Post primary	434 (67.3)	640 (70.0)
Income Uganda Shs		
≤ 99,999	118 (18.3)	460 (84.1)
100,000-199,999	194 (30.1)	289 (30.2)
≥ 200,000	332 (51.6)	207 (21.7)
Work environment		
Indoors	75 (11.7)	512 (53.6)
Semi-Indoors	168 (26.1)	286 (30.0)
Out-doors	401 (62.2)	158 (16.4)
Parish		
Lubaga North	166 (25.8)	248 (25.9)
Lubaga South	478 (74.2)	708 (74.1)

Table 4.2: Stratified Analysis of Incidence Rate Ratios by Demographic Characteristics among Men and Women

	MEN				WOMEN				
Characteristic	**(N) Baseline**	**Follow-up Person-year**	**(N) Converters**	**LATENT TUBERCULOSIS INFECTION Incidence Rate/100PYO**	**(N) Baseline**	**Follow-up Person-year**	**(N) Converters**	**LATENT TUBERCULOSIS INFECTION density rate/100PYO**	**Incidence Rate Ratio (CI)**
Overall	644	617.3	99	16.0	956	886.1	93	10.6	**1.6 (1.2,2.0)**
Age category									
15-24	347 (53.9)	332.7	43	13	515 (53.9)	464.9	42	9.0	1.5 (1.0,2.2)
25-34	200 (31.1)	192.6	33	17.1	324 (33.9)	307.1	35	11.4	1.5 (0.9,2.4)
35-44	71 (11.0)	61.9	16	25.8	92 (9.6)	89.1	10	11.7	2.1 (1.0,4.2)
45+	26 (4.0)	30.1	07	23.3	25 (2.6)	25.0	06	24.0	1.1 (0.4,2.9)
HIV Status									
Positive	28 (4.4)	23.2	06	25.9	65 (6.8)	47.7	08	16.8	1.7 (0.7,4.6)
Negative	569 (88.3)	553.7	86	15.5	822 (86.0)	777.1	80	10.3	**1.6 (1.2,2.1)**
Refused	47 (7.3)	40.4	07	17.3	69 (7.2)	61.3	05	8.2	2.1 (0.7,6.1)
Marital status									
Married	246 (38.2)	238.5	44	18.8	507 (53)	474.4	54	11.5	**1.7 (1.2,2.4)**
Not married	398 (61.8)	378.7	55	13.8	449 (47)	411.7	39	9.5	1.6 (1.1,2.3)
Religion									
Catholics	269 (41.8)	261.7	31	12.1	375 (39.2)	349.1	27	7.8	1.6 (1.0,2.9)

Anglican	166 (25.8)	148.5	27	18.2	250 (26.2)	226.9	30	13.2	1.4 (0.8,2.2)
Muslim	156 (24.2)	157.6	28	17.8	212 (22.2)	200.4	24	12.1	1.6 (1.0,2.6)
Other	53 (8.2)	49.4	13	26.3	119 (12.4)	109.7	12	10.9	**2.4 (1.2,5.0)**
Education level									
Primary	210 (32.7)	193.4	28	14.5	286(30.0)	286.1	37	12.9	1.0(0.6,1.6)
Post-primary	434 (67.3)	423.8	71	16.9	640 (70.0)	600.0	56	9.4	**2.0 (1.4,2.7)**
Income Uganda Shs									
≤ 99,999	118 (18.3)	113.7	08	7.2	460 (48.1)	428.0	46	8.7	0.7 (0.3,1.4)
100,000-199,999	194 (30.1)	182.3	32	17.5	289 (30.2)	267.1	28	10.6	1.7 (1.1,2.7)
≥ 200,000	332 (51.6)	321.2	59	18.4	207 (21.7)	191.0	19	10.0	**1.9 (1.2,3.2)**
Work environment									
Indoors	75 (11.7)	66.8	12	18.0	512 (53.6)	459.6	49	10.7	1.7 (0.9,3.0)
Semi-Indoors	168 (26.1)	158.7	29	18.7	286 (30.0)	275.8	29	10.7	1.7 (1.0,2.7)
Out doors	401 (62.2)	391.8	58	14.9	158 (16.4)	150.7	15	10.0	1.5 (0.9,2.6)
Parish of residency									
Lubaga North	166(25.8)	147.6	24	46.3	248 (25.9)	223.3	23	46.1	1.6 (0.9,2.7)
Lubaga South	478 (74.2)	469.6	75	16.0	708 (74.1)	662.7	70	10.6	**1.6 (1.2,2.2)**

Table 4.3: Stratified Analysis of Incidence Rate Ratios by Risk Factors Among Men and Women

	MEN				WOMEN				
Characteristic	**(N) Baseline**	**Follow-up Person-year**	**(N) Converters**	**LATENT TUBERCULOSIS INFECTION Incidence rate/100PYO**	**(N) Baseline**	**Follow-up Person-year**	**(N) Converters**	**LATENT TUBERCULOSIS INFECTION Incidence rate/100PYO**	**Rate Ratio**
Knew someone with TB									
Yes	121 (18.8)	114.7	19	16.6	197 (20.6)	194.1	29	14.9	1.1 (0.6,2.0)
No	523 (81.1)	502.5	80	15.9	759 (79.4)	692	64	9.2	**1.7 (1.2,2.4)**
Smoking									
Ever smoked	25 (3.9)	19.8	06	30.3	18 (1.9)	14.0	02	14.3	2.1 (0.4,10.5)
Never smoked	619 (96.1)	597.4	93	15.6	938 (98.1)	872.1	91	10.4	**1.5 (1.1,2.0)**
Alcohol use									
Yes	156 (24.2)	148.4	32	21.6	184 (19.3)	166.4	16	9.6	**2.2 (1.2,4.0)**
No	488 (75.8)	468.8	67	14.3	772 (80.2)	719.7	77	10.7	1.3 (1.0,1.8)
Works from home									
Yes	67 (10.4)	60.8	13	21.4	308 (32.2)	286.1	31	10.8	2.0 (1.0,3.7)
No	577 (89.6)	556.5	86	15.5	648 (67.9)	560.0	62	11.1	1.4 (1.0,1.9)

Frequency of Taxi rides									
Daily Use	326 (50.6)	311.3	52	16.7	432 (45.2)	421.4	42	10	**1.7 (1.1,2.5)**
Sometimes	124 (19.3)	131.5	15	11.4	195 (20.4)	177.6	20	11.3	1.0 (0.5,2.0)
Never	194 (30.1)	174.4	32	18.3	329 (34.4)	287.1	31	10.8	1.7 (1.0,2.8)
Ever lived with TB case									
Yes	59 (9.2)	60.6	07	11.6	117 (12.2)	96.3	14	14.5	0.8 (0.3,2.0)
No	585 (90.8)	556.6	92	16.5	839 (87.8)	789.8	79	10.0	**1.7 (1.2,2.2)**
Been in contact of TB case in 12mo									
Yes	55 (8.5)	48.4	09	18.6	92 (9.6)	88.0	15	17.0	1.1 (0.5,2.5)
No	589 (91.5)	568.9	90	15.8	864 (90.4)	798.1	78	10.2	**1.6 (1.2,2.2)**
Lived with TB case in 12mo.									
Yes	14 (2.2)	14.4	02	13.9	34 (3.6)	31.2	07	22.4	0.6 (0.1,3.0)
No	630 (97.8)	602.8	97	16.1	922 (96.4)	854.8	86	10.1	**1.6 (1.2,2.1)**
Shared bedroom with TB case									
Yes	06 (0.9)	5.3	01	18.9	19 (2.0)	18.7	06	32.1	0.6 (0.1,4.9)

No	638 (99.1)	612	98	16.0	937 (98.0)	867.3	87	10.0	**1.6 (1.2,2.1)**
Spent ≥ 4hrs with case									
Yes	23 (3.6)	19.5	02	10.2	48 (5.0)	46	07	15.2	0.7 (0.1,3.2)
No	621 (96.4)	597.8	97	16.2	908 (95.0)	840.1	86	10.2	**1.6 (1.2,2.1)**
HIV status									
Positive	28	23.2	06	25.9	65	47.7	08	16.8	1.7 (0.7,4.6)
Negative	569	553.7	86	15.5	822	777.1	80	10.3	**1.6 (1.2,2.1)**
Refused	47	40.4	07	17.3	69	61.3	05	8.2	2.1 (0.7,6.1)

Table 4.4. Stratified Analysis of Risk Factors Among Men and Women

Category	Person Years OB	Incidence rate of Latent tuberculosis infection		Incidence Rate Ratio (log-binomial)
		n	/100YOB	
Knew someone with TB				
No	1194.5	144	12.1	1
Yes	308.8	48	15.5	**1.8 (1.3,2.5)**
Smoking				
Never smoked	1469.5	184	12.5	1
Ever smoked	33.8	08	23.7	**1.5 (1.2,2.0)**
Alcohol use				
No	1188.5	144	12.1	**1**
Yes	314.8	48	15.2	1.4 (1.0,1.9)
Works from home				
No	1116.5	148	13.3	1
Yes	346.9	44	12.7	**1.6 (1.1,2.1)**
Frequency of Taxi rides				
Never	461.5	63	13.7	1
Daily Use	732.7	94	12.8	1.0 (0.8,1.3)
Sometimes	309.1	35	11.3	0.9 (0.6,1.3)
Ever lived with TB case				
No	1346.4	171	12.7	1
Yes	156.9	21	13.4	**1.7 (1.3,2.2)**
Been in contact of TB case in 12mo				
No	1367	168	12.3	1
Yes	136.4	24	17.5	**1.7 (1.3,2.2)**
Lived with TB case in 12mo.				
No	1457.7	183	12.6	1
Yes	45.6	09	19.7	**1.7 (1.3,2.2)**
Shared bedroom with TB case				
No	1479.3	185	12.5	1
Yes	24	07	29.2	**1.7 (1.3,2.2)**
Spent 4hrs with case				
No	1437.9	180	12.5	1
Yes	65.5	12	18.3	**1.7 (1.3,2.2)**
HIV status				

Category	Person Years OB	Incidence rate of Latent tuberculosis infection		Incidence Rate Ratio (log-binomial)
		n	/100YOB	
Negative	777.1	166	21.4	1
Positive	47.7	14	29.4	1.1 (0.7,1.8)
Refused	61.3	12	19.6	2.0 (0.,1.5)

Table 4.5 Univariate Analysis of Demographic Factors Among Men and Women

Category	Person Years OB	Incidence rate of Latent tuberculosis infection		Relative risk
		n	/100YOB	
Sex				
Female	886.1	93	9.7	**1**
Male	617.3	99	16.0	**1.6 (1.2,2.1)**
Age Category				
15-24	797.6	85	10.7	1
25-34	499.7	68	13.6	1.3 (0.9,1.7)
35-44	151	26	17.2	**1.6 (1.1,2.4)**
45+	55.1	13	23.6	**2.6 (1.6,4.3)**
Marital status				
Not married	790.4	94	11.9	1
Married	712.9	98	13.7	1.6 (1.1,2.3)
Education level				
Post-Secondary	1023.8	127	12.4	1
Pre-secondary	479.5	65	13.6	1.1 (0.9,1.5)
***Income Uganda Shs**				
≥ 200,000	512.2	78	15.2	1
100,000-199,999	449.4	60	13.4	**1.3 (1.1,2.1)**
≤ 99,999	541.7	54	10	**1.5 (1.1,2.1)**
***Occupation**				
Indoors	526.4	61	11.6	1.0 (0.7,1.3)
Semi-Indoors	434.5	58	13.3	0.8 (0.6,1.1)
Out-doors	542.5	73	13.5	1
Parish				
Lubaga North	337.1	36	10.7	1
Lubaga South	1132.3	145	12.8	1.3 (0.9,1.8)

Table 4.6: Univariate Analysis of Risk Factors among Men and Women

Category	Person Years OB	Incidence rate of Latent tuberculosis infection		Incidence Rate Ratio (log-binomial)
		n	/100YOB	
Sex				
Female	886.1	93	9.7	1
Male	617.3	99	16.0	**1.6 (1.2,2.1)**
Age Category				
15-24	797.6	85	10.7	1
25-34	499.7	68	13.6	1.3 (1.0,1.8)
35-44	151	26	17.2	**1.6 (1.1,2.4)**
45+	55.1	13	23.6	**2.6 (1.6,4.3)**
Knew someone with TB				
No	1194.5	144	12.1	1
Yes	308.8	48	15.5	1.3 (1.01.8)
Smoking				
Never smoked	1469.5	184	12.5	1
Ever smoked	33.8	08	23.7	1.6 (0.8,3.0)
Works from home				
No	1116.5	148	13.3	1
Yes	346.9	44	12.7	1.0 (0.7,1.3)
Ever lived with TB case				
No	1346.4	171	12.7	1
Yes	156.9	21	13.4	1.0 (0.6,1.5)
Been in contact of TB case in 12mo				
No	1367	168	12.3	1
Yes	136.4	24	17.5	1.4 (1.0,2.1)
Lived with TB case in 12mo.				
No	1457.7	183	12.6	1
Yes	45.6	09	19.7	1.6 (0.9,2.9)
Shared bedroom with TB case				
No	1479.3	185	12.5	**1**
Yes	24	07	29.2	**2.4 (1.3,4.5)**
Spent 4hrs with case				
No	1437.9	180	12.5	1
Yes	65.5	12	18.3	1.4 (0.8,2.4)
HIV status				

Category	Person Years OB	Incidence rate of Latent tuberculosis infection		Incidence Rate Ratio (log-binomial)
		n	/100YOB	
Negative	777.1	166	21.4	1
Positive	47.7	14	29.4	1.2 (0.7,2.1)
Refused	61.3	12	19.6	0.9 (0.5,1.5)

Table 4.7: Final Model of Multivariate Analysis

Category	Person Years OB	Incidence rate of Latent tuberculosis infection		Relative risk
		n	/100YOB	
Sex				
Female	886.1	93	9.7	1
Male	617.3	99	16.0	**1.6 (1.2,2.0)**
Age Category				
15-24	797.6	85	10.7	1
25-34	499.7	68	13.6	1.3 (1.0,1.8)
35-44	151	26	17.2	1.6 (1.0,2.3)
45+	55.1	13	23.6	**2.3 (1.4,3.9)**
Shared bedroom with TB case				
No	1479.3	185	12.5	**1**
Yes	24	07	29.2	**2.2 (1.2,4.2)**
HIV status				
Negative	777.1	166	21.4	1
Positive	47.7	14	29.4	1.2 (0.7,1.9)
Refused	61.3	12	19.6	0.9 (0.5,1.5)

CHAPTER 5

SYNTHESIS OF RESULTS, BROADER PUBLIC HEALTH IMPLICATIONS AND CONCLUSIONS

In this final chapter of the dissertation, I review and synthesize some of the important findings from this work in the broader context of public health.

Globally, nearly 25% of the population is estimated to be latently infected with *M*. tuberculosis [175]. People with latent tuberculosis infection are at risk of developing active disease and becoming infectious[5]. Studies show that risk for progression to disease once infected may range from 5% in children[6] who are recently exposed to as high as 16% among individuals infected with HIV[7].

Treatment of high-risk individuals with latent tuberculosis infection to prevent active disease has been a core strategy of TB control programs in high-income countries[176]. In these settings, preventive therapy has led to the reduction of TB disease[177]. There is need to address the burden of latent infection particularly in high burden areas to achieve the reductions in incidence necessary to reach the End TB Strategy targets or TB elimination by 2050[178]. This synthesis reiterates the rationale for contact investigations, latent tuberculosis infection screening, preventive therapy for latent tuberculosis infection and current implication for public health in light of our research findings.

5.1 TB prevention and control strategies

The Global control of TB is currently based on three strategies[179]: vaccination with BCG, case finding and treatment of active disease, plus treatment of latent TB infection. BCG vaccine is given to infants to reduce the risk of TB disease. There is no consensus on the protective efficacy of BCG vaccines, and estimates of protection vary.[92,180] Meta-analysis by Roy found that BCG provides on average 50% protection against TB and is effective for 10–20 years, although efficacy varies by geographic region and is much lower in adults than in children[181], epidemiologic data suggest that the vaccine is more useful in protecting children from disseminated extra-pulmonary TB including TB meningitis, than in protecting adults from primary pulmonary infection[180]. Although the efficacy of the BCG vaccine continues to be controversial, live attenuated BCG is still the only vaccine in use for the prevention of TB in humans.

Passive case finding (PCF), is the detection of active TB disease among symptomatic persons voluntarily presenting to the health system and is the standard approach adopted by most national TB programs[179] and widely promoted in developing countries as part of the WHO recommended DOTS strategy. Active case finding (ACF) is an alternative approach for case detection. It refers to provider-initiated efforts to find, evaluate and diagnose active TB among asymptomatic and symptomatic individuals who have not sought care so as to interrupt transmission of TB through early detection and prompt initiation of effective treatment[14]. Case detection and treatment of TB is the principal means of controlling transmission and reducing TB Incidence. Case detection entails systematic identification of people with suspected active TB, in a predetermined target group. In this context, such groups include household contacts and other close contacts of TB cases and people living with HIV.

5.2 Household contact investigation

Our study revealed, as expected, a high prevalence of latent tuberculosis infection among household contacts of TB cases at 56%. However, what was surprising was the high prevalence of latent tuberculosis infection among household contacts of controls at 31.1%. We had predicted a gradient in the prevalence as one moved from the household contacts of cases to non-household contacts of controls. However, the prevalence of latent tuberculosis infection among non-household contacts of cases was similar to that of non-household controls meaning that non-household contacts of cases had more or less the same contact or exposure with TB patients as non-household contacts of control.

Investigation of household contacts exposed to infectious tuberculosis is widely recommended by international guidelines mainly to identify secondary cases of TB because household contacts of patients with active tuberculosis are at high risk for the disease[182]. Although contact investigation has been a component of tuberculosis control in high-income, low-prevalence countries like the US for a longtime[15], its implementation in high burden low income settings remains limited. In low- and middle-income countries, where the burden of TB is often high, the strategy is mainly concentrated on intensifying active case finding[183].

Household contact investigation if done in a systematic way helps to identify TB infected individuals among household contacts of confirmed TB patients[184] but policy makers have tended to neglected other many non-household contacts who may be equally at risk for infection.

5.3 Community contact investigation

Knowledge of prevalent latent tuberculosis infection is relevant to public health programs because individuals with latent tuberculosis infection represent a pool from which future TB cases arise. It

has been shown that infection with TB in high burden settings is often due to exposure outside of the home environment[168]. Majority of TB transmission in high-burden settings occurs outside of the home environment after early childhood[185]. Despite strong evidence that most TB infections occur outside of the home in TB high burden settings[16,20], there is little data to guide the use of contact investigations outside of the household. Results from this study add to the importance of non-household contact investigations in TB prevention approaches because of the high prevalence reported in the community.

Frequently used definitions of 'household' in TB contact investigations in high-burden TB settings include 'all dwellings on the same plot of land that share the same residential address'[186] and or 'a person or a group of related or unrelated persons, who live together in the same dwelling unit, who share the same housekeeping arrangements, and who have the same eating arrangements[187]. In the context of TB contact investigations, the term 'household' should be expanded to capture not only a group of people who function as a unit, or a physical housing structure/residential address but rather to include some aspect of non-household social contacts in order to increase the sensitivity of community investigations.

5.4 Guidelines on latent tuberculosis infection

WHO recommends contact investigation for household and close contacts to be performed for children aged < 5 years, people living with/or have high-risk of HIV infection and contacts of index cases with multidrug resistant[179]. Current WHO guidelines on latent tuberculosis infection are based on the likelihood of progression to active TB disease in specific risk groups, the underlying epidemiology and burden of TB, the availability of resources and the likelihood of a broader public health impact[188].

5.5 The importance of treating latent tuberculosis infection in high burden settings

To achieve the broad public health goal of TB elimination requires detection and treatment of latent tuberculosis infection [11]. This is critically important because even in the absence of new infections, new cases of TB disease can emerge from the pool of latently infected individuals due to TB reactivation. We now know that HIV infection is the strongest risk factor for reactivation TB[111]. In high burden settings like urban Uganda where the prevalence of latent tuberculosis infection is 50%[127] and that of adult HIV infection at 6.5%[189], it is imperative to put concerted efforts into screening and treating for latent tuberculosis infection.

5.6 Policy Implications of our findings in the context of TB control

We need to interrupt the path to disease in order to supplement case detection strategies that identify new cases. Within the broader guidelines offered by WHO[190] country TB control programs prioritize contact investigation usually on the basis of the local epidemiology of TB, operational capacity and resources[183]. In endemic areas, many of which have high prevalence of HIV, contact investigations that go beyond the traditional household setting is likely to be beneficial.

To optimize the current policy for contact investigations, we recommend a comprehensive TB elimination program that exploits the synergies of different TB control strategies including BCG vaccination, passive and active case finding, household and non-household contact investigations, infection control, preventive therapy for latent tuberculosis infection and antiretroviral therapy.

5.7 Recommendation for future Research

Our findings highlight the critical need for expanding the scope of contact investigation beyond the traditional confines of “household” to non-household community contacts. Similarly,

the higher probability for infection among men proves the gender differences in TB infection are real. Many questions remain unanswered though and present opportunities for future research.

First, research is needed to explore whether the observed gender differences in infection are due to behavioral factors such as differences in social networks of or differences are due to biological factors. Second, non-household interactions are complex and prioritizing non-household contacts is still a challenge. In order to make strong recommendations for the implementation of specific community interventions, more research is needed to identify hot spots or locations of high transmission to allow for prioritizing screening outside the household of the index case. Understanding where in the community most TB transmissions occurs, will broaden the understanding of how best to prioritize non-household contact investigations and increase the sensitivity of non-household contact evaluations.

5.8 Conclusions

In summary, this work provides consistent evidence of non-household transmission in endemic areas is high. Equally Important, the second aim of the study has shown that men are more susceptible to latent tuberculosis infection that women. Therefore, policy makers ought to take into account non-household contacts and men in particular when designing TB prevention interventions. Implementing contact investigations, whereas not ignoring women, should be greatly strengthened for men.

REFERENCES

1. WHO. Global tuberculosis report 2016. 2016.
2. Houben RMGJ, Dodd PJ. The Global Burden of Latent Tuberculosis Infection: A Re-estimation Using Mathematical Modelling. *PLoS Medicine.* 2016;13(10):e1002152.
3. WHO. The Global Plan to Stop TB. 20016-2020.
4. Dye C, Scheele S, Dolin P, Pathania V, Raviglione MC. Consensus statement. Global burden of tuberculosis: estimated incidence, prevalence, and mortality by country. WHO Global Surveillance and Monitoring Project. *Jama.* 1999;282(7):677-686.
5. World Health Organization. *WHO TB REPORT,Global Tuberculosis Control, World Health Organization,.* WHO 2013,.
6. Whalen CC, Zalwango S, Chiunda A, et al. *Secondary Attack Rate of Tuberculosis in Urban Households in Kampala, Uganda.*
7. Trieu L, Li J, Hanna DB, Harris TG. Tuberculosis Rates Among HIV-Infected Persons in New York City, 2001--2005. *American Journal of Public Health.* 2010;100(6):1031-1034.
8. Rangaka MX, Cavalcante SC, Marais BJ, et al. Controlling the seedbeds of tuberculosis: diagnosis and treatment of tuberculosis infection. *Lancet (London, England).* 2015;386(10010):2344-2353.
9. Houben RMGJ, Dodd PJ. The Global Burden of Latent Tuberculosis Infection: A Re-estimation Using Mathematical Modelling. *Plos Medicine.* 2016;13(10):e1002152-e1002152.
10. Lorent N, Choun K, Thai S, et al. Community-based active tuberculosis case finding in poor urban settlements of Phnom Penh, Cambodia: a feasible and effective strategy. *PLoS One.* 2014;9(3):e92754.
11. Dye C, Glaziou P, Floyd K, Raviglione M. Prospects for tuberculosis elimination. *Annual review of public health.* 2013;34:271-286.
12. Abu-Raddad LJ, Sabatelli L, Achterberg JT, et al. Epidemiological benefits of more-effective tuberculosis vaccines, drugs, and diagnostics. *Proceedings of the National Academy of Sciences of the United States of America.* 2009;106(33):13980-13985.
13. Mack U, Migliori GB, Sester M, et al. LTBI: latent tuberculosis infection or lasting immune responses to M. tuberculosis? A TBNET consensus statement. *European Respiratory Journal.* 2009;33(5):956-973.
14. De Cock KM, Chaisson RE. Will DOTS do it? A reappraisal of tuberculosis control in countries with high rates of HIV infection. *The international journal of tuberculosis and lung disease : the official journal of the International Union against Tuberculosis and Lung Disease.* 1999;3(6):457-465.
15. Guidelines for the investigation of contacts of persons with infectious tuberculosis. Recommendations from the National Tuberculosis Controllers Association and CDC.

MMWR Recommendations and reports : Morbidity and mortality weekly report Recommendations and reports. 2005;54(Rr-15):1-47.
16. Andrews JR, Morrow C, Walensky RP, Wood R. Integrating social contact and environmental data in evaluating tuberculosis transmission in a South African township. *J Infect Dis.* 2014;210(4):597-603.
17. Andrews JR, Morrow C, Walensky RP, Wood R. Integrating social contact and environmental data in evaluating tuberculosis transmission in a South African township. *Journal of Infectious Diseases.* 2014;210(4):597-603.
18. Kizza FN, List J, Nkwata AK, et al. Prevalence of latent tuberculosis infection and associated risk factors in an urban African setting. *BMC Infectious Diseases.* 2015;15:165-165.
19. Chamie G, Wandera B, Marquez C, et al. Identifying locations of recent TB transmission in rural Uganda: a multidisciplinary approach. *Tropical Medicine & International Health.* 2015;20(4):537-545.
20. Zelner JL, Murray MB, Becerra MC, et al. *Age-Specific Risks of Tuberculosis Infection From Household and Community Exposures and Opportunities for Interventions in a High-Burden Setting.*
21. .
22. Holmes CB, Hausler H, Nunn P. A review of sex differences in the epidemiology of tuberculosis. *The international journal of tuberculosis and lung disease : the official journal of the International Union against Tuberculosis and Lung Disease.* 1998;2(2):96-104.
23. Mancuso JD, Diffenderfer JM, Ghassemieh BJ, Horne DJ, Kao TC. The Prevalence of Latent Tuberculosis Infection in the United States. *Am J Respir Crit Care Med.* 2016;194(4):501-509.
24. Hudelson P. Gender differentials in tuberculosis: the role of socio-economic and cultural factors. *Tubercle and lung disease : the official journal of the International Union against Tuberculosis and Lung Disease.* 1996;77(5):391-400.
25. Horton KC, MacPherson P, Houben RM, White RG, Corbett EL. Sex Differences in Tuberculosis Burden and Notifications in Low- and Middle-Income Countries: A Systematic Review and Meta-analysis. *PLoS Med.* 2016;13(9):e1002119.
26. Borgdorff MW, Nagelkerke NJ, Dye C, Nunn P. Gender and tuberculosis: a comparison of prevalence surveys with notification data to explore sex differences in case detection. *The international journal of tuberculosis and lung disease : the official journal of the International Union against Tuberculosis and Lung Disease.* 2000;4(2):123-132.
27. Borgdorff MW, Nagelkerke NJD, Dye C, Nunn P. Gender and tuberculosis: a comparison of prevalence surveys with notification data to explore sex differences in case detection. *International Journal of Tuberculosis and Lung Disease.* 2000;4(2):123-132.
28. Chen C, Yang CG, Gao X, et al. Community-based active case finding for tuberculosis in rural western China: a cross-sectional study. *The international journal of tuberculosis and lung disease : the official journal of the International Union against Tuberculosis and Lung Disease.* 2017;21(11):1134-1139.
29. Chheng P, Nsereko M, Malone LL, et al. Tuberculosis case finding in first-degree relative contacts not living with index tuberculosis cases in Kampala, Uganda. *Clinical epidemiology.* 2015;7:411-419.

30. Martinez AN, Rhee JT, Small PM, Behr MA. Sex differences in the epidemiology of tuberculosis in San Francisco. *International Journal of Tuberculosis and Lung Disease.* 2000;4(1):26-31.
31. Mumpe-Mwanja D, Verver S, Yeka A, et al. Prevalence and risk factors of latent Tuberculosis among adolescents in rural Eastern Uganda. *African health sciences.* 2015;15(3):851-860.
32. Guwatudde D, Nakakeeto M, Jones-Lopez EC, et al. Tuberculosis in household contacts of infectious cases in Kampala, Uganda. *Am J Epidemiol.* 2003;158(9):887-898.
33. Kayanja HK, Debanne S, King C, Whalen CC. Tuberculosis infection among health care workers in Kampala, Uganda. *Int J Tuberc Lung Dis.* 2005;9(6):686-688.
34. Pai M, Denkinger CM, Kik SV, et al. Gamma interferon release assays for detection of Mycobacterium tuberculosis infection. *Clinical microbiology reviews.* 2014;27(1):3-20.
35. Gagneux S. Host–pathogen coevolution in human tuberculosis. *Philosophical Transactions of the Royal Society B: Biological Sciences.* 2012;367(1590):850-859.
36. Sinha P, Gupta A, Prakash P, Anupurba S, Tripathi R, Srivastava GN. Differentiation of Mycobacterium tuberculosis complex from non-tubercular mycobacteria by nested multiplex PCR targeting IS6110, MTP40 and 32kD alpha antigen encoding gene fragments. *BMC Infect Dis.* 2016;16.
37. Dunlap NE, Bass J, Fujiwara P, et al. American Thoracic Society. Diagnostic standards and classification of tuberculosis in adults and children. *American Journal of Respiratory & Critical Care Medicine.* 2000;161(4):1376-1395.
38. Dheda K, Schwander SK, Zhu B, van Zyl-Smit RN, Zhang Y. The immunology of tuberculosis: from bench to bedside. *Respirology (Carlton, Vic).* 2010;15(3):433-450.
39. Edwards D, Kirkpatrick CH. The immunology of mycobacterial diseases. *The American review of respiratory disease.* 1986;134(5):1062-1071.
40. Behar SM, Divangahi M, Remold HG. Evasion of innate immunity by Mycobacterium tuberculosis: is death an exit strategy? *Nature Reviews Microbiology.* 2010;8(9):668-674.
41. Barry CE, 3rd, Boshoff HI, Dartois V, et al. The spectrum of latent tuberculosis: rethinking the biology and intervention strategies. *Nature Reviews Microbiology.* 2009;7(12):845-855.
42. Comas I, Chakravartti J, Small PM, et al. Human T cell epitopes of Mycobacterium tuberculosis are evolutionarily hyperconserved. *Nature Genetics.* 2010;42(6):498-503.
43. Pai M, Behr MA, Dowdy D, et al. Tuberculosis. *Nature Reviews Disease Primers.* 2016;2:16076.
44. Walzl G, Ronacher K, Hanekom W, Scriba TJ, Zumla A. Immunological biomarkers of tuberculosis. *Nature Reviews Immunology.* 2011;11(5):343-354.
45. Esmail H, Barry CE, Young DB, Wilkinson RJ. The ongoing challenge of latent tuberculosis. *Philosophical Transactions of the Royal Society B: Biological Sciences.* 2014;369(1645).
46. Wallgren A. The time-table of tuberculosis. *Tubercle.* 1948;29(11):245-251.
47. Ioachimescu OC. Tuberculosis. 2010.
48. Sharma SK, Mohanan S, Sharma A. Relevance of Latent TB Infection in Areas of High TB Prevalence. *Chest.* 2012;142(3):761-773.
49. Ai JW, Ruan QL, Liu QH, Zhang WH. Updates on the risk factors for latent tuberculosis reactivation and their managements. *Emerging Microbes & Infections.* 2016;5(2):e10-.

50. Padmanesan N, James W, Chandini Raina M, Dilip M. Risk Factors for Tuberculosis. *Pulmonary Medicine, Vol 2013 (2013).* 2013.
51. Lienhardt C. *From exposure to disease: The role of environmental factors in susceptibility to and development of tuberculosis.*
52. Ingrosso L, Vescio F, Giuliani M, et al. Risk factors for tuberculosis in foreign-born people (FBP) in Italy: a systematic review and meta-analysis. *PLoS One.* 2014;9(4):e94728.
53. WHO. WHO Guidelines on the management of latent tuberculosis infection. 2015.
54. Ferrarini MAG, Spina FG, Weckx LY, Lederman HM, Moraes-Pinto MId. Rate of tuberculosis infection in children and adolescents with household contact with adults with active pulmonary tuberculosis as assessed by tuberculin skin test and interferon-gamma release assays. *Epidemiology and Infection.* 2016;144(4):712-723.
55. Kenyon TA, Valway SE, Onorato IM. Transmission of Tuberculosis during a Long Airplane Flight. In. Vol 3351996:675-676.
56. Andersen S, Geser A. The distribution of tuberculous infection among households in African communities. *Bull World Health Organ.* 1960;22:39-60.
57. Stead WW, Senner JW, Reddick WT, Lofgren JP. Racial differences in susceptibility to infection by Mycobacterium tuberculosis. *N Engl J Med.* 1990;322(7):422-427.
58. Lienhardt C. From exposure to disease: the role of environmental factors in susceptibility to and development of tuberculosis. *Epidemiologic reviews.* 2001;23(2):288-301.
59. Schmidt CW. Linking TB and the Environment: An Overlooked Mitigation Strategy. *Environmental Health Perspectives.* 2008;116(11):A478-485.
60. Zeligs BJ, Nerurkar LS, Bellanti JA. Chemotactic and candidacidal responses of rabbit alveolar macrophages during postnatal development and the modulating roles of surfactant in these responses. *Infection and immunity.* 1984;44(2):379-385.
61. Jackson JC, Palmer S, Wilson CB, et al. Postnatal changes in lung phospholipids and alveolar macrophages in term newborn monkeys. *Respiration physiology.* 1988;73(3):289-300.
62. Li J, Munsiff SS, Agerton TB. Prevalence of tuberculin skin test positivity in clinical population in New York City. *J Immigr Minor Health.* 2010;12(6):816-822.
63. Pareek M, Bond M, Shorey J, et al. Community-based evaluation of immigrant tuberculosis screening using interferon gamma release assays and tuberculin skin testing: observational study and economic analysis. *Thorax.* 2013;68(3):230-239.
64. Shanaube K, Hargreaves J, Fielding K, et al. Risk factors associated with positive QuantiFERON-TB Gold In-Tube and tuberculin skin tests results in Zambia and South Africa. *PLoS One.* 2011;6(4):e18206.
65. Shaw JB, Wynn-Williams N. Infectivity of pulmonary tuberculosis in relation to sputum status. *American review of tuberculosis.* 1954;69(5):724-732.
66. Loudon RG, Spohn SK. Cough frequency and infectivity in patients with pulmonary tuberculosis. *The American review of respiratory disease.* 1969;99(1):109-111.
67. Lienhardt C, Fielding K, Sillah J, et al. Risk factors for tuberculosis infection in sub-Saharan Africa: a contact study in The Gambia. *Am J Respir Crit Care Med.* 2003;168(4):448-455.
68. Martinez L, Sekandi JN, Castellanos ME, Zalwango S, Whalen CC. Infectiousness of HIV-Seropositive Patients with Tuberculosis in a High-Burden African Setting. *Am J Respir Crit Care Med.* 2016;194(9):1152-1163.

69. van Ingen J. 185 - Mycobacteria A2 - Cohen, Jonathan. In: Powderly WG, Opal SM, eds. *Infectious Diseases (Fourth Edition).* Elsevier; 2017:1645-1659.e1642.
70. Talbot EA, Raffa BJ. Chapter 92 - Mycobacterium tuberculosis A2 - Tang, Yi-Wei. In: Sussman M, Liu D, Poxton I, Schwartzman J, eds. *Molecular Medical Microbiology (Second Edition).* Boston: Academic Press; 2015:1637-1653.
71. Yang HL, Kruh-Garcia NA, Dobos KM. *Purified protein derivatives of tuberculin - past, present, and future.*
72. Azimi S, Tebianian M, Mosavari N, et al. *Evaluation of Immunological Parameters in Purified Protein Derivative Positive Tuberculin Workers.*
73. Abdel-Samea SA, Ismail YM, Fayed SMA, Mohammad AA. Comparative study between using QuantiFERON and tuberculin skin test in diagnosis of Mycobacterium tuberculosis infection. *Egyptian Journal of Chest Diseases and Tuberculosis.* 2013;62(1):137-143.
74. Nayak S, Acharjya B. Mantoux test and its interpretation. *Indian dermatology online journal.* 2012;3(1):2-6.
75. Diagnostic Standards and Classification of Tuberculosis in Adults and Children. This official statement of the American Thoracic Society and the Centers for Disease Control and Prevention was adopted by the ATS Board of Directors, July 1999. This statement was endorsed by the Council of the Infectious Disease Society of America, September 1999. *Am J Respir Crit Care Med.* 2000;161(4 Pt 1):1376-1395.
76. CDC. LATENT TUBERCULOSIS INFECTION: A GUIDE FOR PRIMARY HEALTH CARE PROVIDERS. 2013.
77. Aabye MG, Ravn P, PrayGod G, et al. The Impact of HIV Infection and CD4 Cell Count on the Performance of an Interferon Gamma Release Assay in Patients with Pulmonary Tuberculosis. *PLoS ONE.* 2009;4(1):1-8.
78. Metcalfe JZ, Everett CK, Steingart KR, et al. Interferon-γ release assays for active pulmonary tuberculosis diagnosis in adults in low- and middle-income countries: systematic review and meta-analysis. *The Journal Of Infectious Diseases.* 2011;204 Suppl 4:S1120-S1129.
79. Campion EW, Getahun H, Matteelli A, Chaisson RE, Raviglione M. Review Article: Latent Mycobacterium tuberculosis Infection. *The New England Journal of Medicine.* 2015;372:2127-2135.
80. Tincati C, Cappione Iii AJ, Snyder-Cappione JE. Distinguishing Latent from Active Mycobacterium tuberculosis Infection Using Elispot Assays: Looking Beyond Interferon-gamma. *Cells.* 2012;1(2):89-99.
81. Menzies D. Interpretation of repeated tuberculin tests. Boosting, conversion, and reversion. *American journal of respiratory and critical care medicine.* 1999;159(1):15-21.
82. Farhat M, Greenaway C, Pai M, Menzies D. False-positive tuberculin skin tests: what is the absolute effect of BCG and non-tuberculous mycobacteria? *The international journal of tuberculosis and lung disease : the official journal of the International Union against Tuberculosis and Lung Disease.* 2006;10(11):1192-1204.
83. Teixeira EG, Kritski A, Ruffino-Netto A, et al. Two-step tuberculin skin test and booster phenomenon prevalence among Brazilian medical students. *The international journal of tuberculosis and lung disease : the official journal of the International Union against Tuberculosis and Lung Disease.* 2008;12(12):1407-1413.

84. Wormser GP, Colebunders RL. Control of Communicable Diseases Manual, 19th Edition Edited by David L. Heymann Washington, DC: American Public Health Association, 2008. 746 pp. $45.00 (hardcover). *Clinical Infectious Diseases.* 2009;49(8):1292-1293.
85. Al Mazrou AM. Booster Effect of Two‐Step Tuberculin Skin Testing Among Hospital Employees From Areas With a High Prevalence of Tuberculosis. *Infection Control and Hospital Epidemiology.* 2004;25(12):1117-1120.
86. Sekandi JN, Zalwango S, Nkwata AK, et al. Low Prevalence of Tuberculin Skin Test Boosting among Community Residents in Uganda. *Am J Trop Med Hyg.* 2018;98(2):379-381.
87. Kerkhoff AD, Ankur G, Samandari T, Lawn SD. The proportions of people living with HIV in low and middle-income countries who test tuberculin skin test positive using either a 5 mm or a 10 mm cut-off: a systematic review. *BMC Infectious Diseases.* 2013;13(307):(8 July 2013)-(2018 July 2013).
88. Whalen CC. The Replacement Principle of Tuberculosis. Why Prevention Matters. *American journal of respiratory and critical care medicine.* 2016;194(4):400-401.
89. Schwartzman K, Menzies D. *How long are TB patients infectious?* : CMAJ. 2000 Jul 25;163(2):157-8.
90. BCG vaccines: WHO position paper - February 2018. *Releve epidemiologique hebdomadaire.* 2018;93(8):73-96.
91. WHO Expert Committee on Biological Standardization. *World Health Organization technical report series.* 2016(999):1-267.
92. Colditz GA, Brewer TF, Berkey CS, et al. Efficacy of BCG vaccine in the prevention of tuberculosis. Meta-analysis of the published literature. *Jama.* 1994;271(9):698-702.
93. CDC. The Role of BCG Vaccine in the Prevention and Control of Tuberculosis in the United States A Joint Statement by the Advisory Council for the Elimination of Tuberculosis and the Advisory Committee on Immunization Practices 1995.
94. Leung CC, Rieder HL, Lange C, Yew WW. Treatment of latent infection with Mycobacterium tuberculosis: update 2010. *European Respiratory Journal.* 2011;37(3):690-711.
95. Sharma SK, Sharma A, Kadhiravan T, Tharyan P. *Rifamycins (rifampicin, rifabutin and rifapentine) compared to isoniazid for preventing tuberculosis in HIV-negative people at risk of active TB.*
96. van Halsema CL, Fielding KL, Chihota VN, et al. *Tuberculosis outcomes and drug susceptibility in individuals exposed to isoniazid preventive therapy in a high HIV prevalence setting.*
97. Samandari T, Agizew TB, Nyirenda S, et al. *6-month versus 36-month isoniazid preventive treatment for tuberculosis in adults with HIV infection in Botswana: a randomised, double-blind, placebo-controlled trial.*
98. *BARRON, Gladys India.* Oxford University Press; 2006.
99. Fielding KL, Grant AD, Hayes RJ, Chaisson RE, Corbett EL, Churchyard GJ. Thibela TB: Design and methods of a cluster randomised trial of the effect of community-wide isoniazid preventive therapy on tuberculosis amongst gold miners in South Africa. *Contemporary Clinical Trials.* 2011;32:382-392.
100. Samandari T, Agizew TB, Nyirenda S, et al. Articles: 6-month versus 36-month isoniazid preventive treatment for tuberculosis in adults with HIV infection in Botswana: a randomised, double-blind, placebo-controlled trial. *The Lancet.* 2011;377:1588-1598.

101. Holmes CB, Hausler H, Nunn P. A review of sex differences in the epidemiology of tuberculosis. *The International Journal Of Tuberculosis And Lung Disease: The Official Journal Of The International Union Against Tuberculosis And Lung Disease.* 1998;2(2):96-104.
102. Horton KC, MacPherson P, Houben RMGJ, White RG, Corbett EL. Sex Differences in Tuberculosis Burden and Notifications in Low- and Middle-Income Countries: A Systematic Review and Meta-analysis. *Plos Medicine.* 2016;13(9):e1002119-e1002119.
103. Neyrolles O, Quintana-Murci L. Sexual Inequality in Tuberculosis. *PLoS Medicine.* 2009;6(12):1-6.
104. Chen C, Zhu T, Wang Z, et al. High Latent TB Infection Rate and Associated Risk Factors in the Eastern China of Low TB Incidence. *PLoS One.* 2015;10(10).
105. Wen-Ying T, Shiang-Fen H, Ming-Che L, et al. Gender disparities in latent tuberculosis infection in high-risk individuals: a cross-sectional study. *PLoS ONE, Vol 9, Iss 11, p e110104 (2014).* (11):e110104.
106. Lee SJ, Lee SH, Kim YE, et al. Risk factors for latent tuberculosis infection in close contacts of active tuberculosis patients in South Korea: a prospective cohort study. *BMC Infectious Diseases.* 2014;14(1):566.
107. Nassaji M, Ghorbani R. Risk factors for latent tuberculosis infection among healthcare workers in a university-affliated hospital. *Southern African Journal of Epidemiology and Infection.* 2012;27(1):30.
108. Mumpe-Mwanja D, Verver S, Yeka A, et al. Prevalence and risk factors of latent Tuberculosis among adolescents in rural Eastern Uganda. *African Health Sciences.* 2015;15(3):851-860.
109. Syhre M, Chambers ST. The scent of Mycobacterium tuberculosis. *Tuberculosis.* 2008;88:317-323.
110. Hector J, Anderson ST, Banda G, et al. TST positivity in household contacts of tuberculosis patients: a case-contact study in Malawi. *BMC Infect Dis.* 2017;17(1):259.
111. Whalen CC, Zalwango S, Chiunda A, et al. Secondary attack rate of tuberculosis in urban households in Kampala, Uganda. *PLoS One.* 2011;6(2):e16137.
112. Kozinska M, Augustynowicz-Kopec E. The incidence of tuberculosis transmission among family members and outside households. *Pneumonologia i alergologia polska.* 2016;84(5):271-277.
113. Crampin AC, Glynn JR, Traore H, et al. Tuberculosis transmission attributable to close contacts and HIV status, Malawi. *Emerg Infect Dis.* 2006;12(5):729-735.
114. Andersen S, Geser A. The Distribution of Tuberculous Infection among Households in African Communities. *Bulletin of the World Health Organization.* 1960;22(1/2):39-60.
115. Shapiro AE, Variava E, Rakgokong MH, et al. Community-based targeted case finding for tuberculosis and HIV in household contacts of patients with tuberculosis in South Africa. *American Journal of Respiratory and Critical Care Medicine.* 2012;185(10):1110-1116.
116. Wilkinson D, Pillay M, Crump J, Lombard C, Davies GR, Sturm AW. Molecular epidemiology and transmission dynamics of Mycobacterium tuberculosis in rural Africa. *Tropical Medicine & International Health: TM & IH.* 1997;2(8):747-753.
117. Whalen CC, Zalwango S, Chiunda A, et al. Secondary Attack Rate of Tuberculosis in Urban Households in Kampala, Uganda. *PLoS ONE.* 2011;6(2):1-7.

118. Verver S, Warren RM, Munch Z, et al. Proportion of tuberculosis transmission that takes place in households in a high-incidence area. *Lancet (London, England).* 2004;363(9404):212-214.
119. Glynn JR, Guerra-Assunção JA, Houben RMGJ, et al. Whole Genome Sequencing Shows a Low Proportion of Tuberculosis Disease Is Attributable to Known Close Contacts in Rural Malawi. *PLoS ONE.* 2015;10(7):1-12.
120. Middelkoop K, Mathema B, Myer L, et al. Transmission of tuberculosis in a South African community with a high prevalence of HIV infection. *Journal of Infectious Diseases.* 2015;211(1):53-61.
121. Chamie G, Marquez C, Havlir DV, et al. *Identifying locations of recent TB transmission in rural Uganda: a multidisciplinary approach.*
122. Beyers N. Case finding in children in contact with adults in the house with TB. *The International Journal Of Tuberculosis And Lung Disease: The Official Journal Of The International Union Against Tuberculosis And Lung Disease.* 2003;7(11):1013-1014.
123. Martinez L, Ye S, Mupere E, Kizza A, Hill PC, Whalen CC. Transmission of Mycobacterium tuberculosis in Households and the Community: A Systematic Review and Meta-Analysis. *American Journal of Epidemiology.* 2017;185(12):1327-1339.
124. Goletti D, Sanduzzi A, Delogu G. Performance of the tuberculin skin test and interferon-γ release assays: an update on the accuracy, cutoff stratification, and new potential immune-based approaches. *Journal of Rheumatology.* 2014;41(Suppl. 91):24-31.
125. Spiegelman D, Hertzmark E. Easy SAS calculations for risk or prevalence ratios and differences. *American journal of epidemiology.* 2005;162(3):199-200.
126. Lawn SD, Wood R, Wilkinson RJ. Changing concepts of "latent tuberculosis infection" in patients living with HIV infection. *Clinical & developmental immunology.* 2011;2011.
127. Kizza FN, List J, Nkwata AK, et al. Prevalence of latent tuberculosis infection and associated risk factors in an urban African setting. *BMC infectious diseases.* 2015;15:165.
128. Martinez L, Arman A, Haveman N, et al. Changes in tuberculin skin test positivity over 20 years in periurban shantytowns in Lima, Peru. *Am J Trop Med Hyg.* 2013;89(3):507-515.
129. Gustafson P, Lisse I, Gomes V, et al. Risk factors for positive tuberculin skin test in Guinea-Bissau. *Epidemiology.* 2007;18(3):340-347.
130. Middelkoop K, Mathema B, Myer L, et al. Transmission of tuberculosis in a South African community with a high prevalence of HIV infection. *The Journal Of Infectious Diseases.* 2015;211(1):53-61.
131. Shapiro AE, Variava E, Rakgokong MH, et al. Community-based targeted case finding for tuberculosis and HIV in household contacts of patients with tuberculosis in South Africa. *Am J Respir Crit Care Med.* 2012;185(10):1110-1116.
132. Ting WY, Huang SF, Lee MC, et al. Gender disparities in latent tuberculosis infection in high-risk individuals: a cross-sectional study. *PLoS One.* 2014;9(11):e110104.
133. Xu J, Hu Y, Jiang W, et al. [Prevalence and risk factors of latent tuberculosis infection in close contacts of tuberculosis patients among non-resident populations in Shanghai, China]. *Zhonghua jie he he hu xi za zhi = Zhonghua jiehe he huxi zazhi = Chinese journal of tuberculosis and respiratory diseases.* 2016;39(1):25-29.
134. Legesse M, Ameni G, Mamo G, Medhin G, Bjune G, Abebe F. Community-based cross-sectional survey of latent tuberculosis infection in Afar pastoralists, Ethiopia, using QuantiFERON-TB Gold In-Tube and tuberculin skin test. *BMC Infect Dis.* 2011;11:89.

135. Mahomed H, Hawkridge T, Verver S, et al. Predictive factors for latent tuberculosis infection among adolescents in a high-burden area in South Africa. *The international journal of tuberculosis and lung disease : the official journal of the International Union against Tuberculosis and Lung Disease.* 2011;15(3):331-336.
136. den Boon S, van Lill SW, Borgdorff MW, et al. Association between smoking and tuberculosis infection: a population survey in a high tuberculosis incidence area. *Thorax.* 2005;60(7):555-557.
137. Middelkoop K, Bekker LG, Morrow C, Lee N, Wood R. Decreasing household contribution to TB transmission with age: a retrospective geographic analysis of young people in a South African township. *BMC Infect Dis.* 2014;14:221.
138. Ncayiyana JR, Bassett J, West N, et al. Prevalence of latent tuberculosis infection and predictive factors in an urban informal settlement in Johannesburg, South Africa: a cross-sectional study. *BMC Infect Dis.* 2016;16(1):661.
139. Middelkoop K, Bekker LG, Myer L, Dawson R, Wood R. Rates of tuberculosis transmission to children and adolescents in a community with a high prevalence of HIV infection among adults. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America.* 2008;47(3):349-355.
140. Mudido PM, Guwatudde D, Nakakeeto MK, et al. The effect of bacille Calmette-Guérin vaccination at birth on tuberculin skin test reactivity in Ugandan children. *International Journal of Tuberculosis and Lung Disease.* 1999;3(10):891-895.
141. Mori T, Sakatani M, Yamagishi F, et al. Specific detection of tuberculosis infection: an interferon-gamma-based assay using new antigens. *Am J Respir Crit Care Med.* 2004;170(1):59-64.
142. Ewer K, Deeks J, Alvarez L, et al. Comparison of T-cell-based assay with tuberculin skin test for diagnosis of Mycobacterium tuberculosis infection in a school tuberculosis outbreak. *Lancet.* 2003;361(9364):1168-1173.
143. Nkurunungi G, Lutangira JE, Lule SA, et al. Determining Mycobacterium tuberculosis infection among BCG-immunised Ugandan children by T-SPOT.TB and tuberculin skin testing. *PLoS One.* 2012;7(10):e47340.
144. Nhamoyebonde S, Leslie A. Biological differences between the sexes and susceptibility to tuberculosis. *The Journal of infectious diseases.* 2014;209 Suppl 3:S100-106.
145. Onifade DA, Bayer AM, Montoya R, et al. Gender-related factors influencing tuberculosis control in shantytowns: a qualitative study. *BMC Public Health.* 2010;10:381.
146. Jimenez-Corona ME, Garcia-Garcia L, DeRiemer K, et al. Gender differentials of pulmonary tuberculosis transmission and reactivation in an endemic area. *Thorax.* 2006;61(4):348-353.
147. Hamid Salim MA, Declercq E, Van Deun A, Saki KA. Gender differences in tuberculosis: a prevalence survey done in Bangladesh. *The international journal of tuberculosis and lung disease : the official journal of the International Union against Tuberculosis and Lung Disease.* 2004;8(8):952-957.
148. Yates TA, Atkinson SH. Ironing out sex differences in tuberculosis prevalence. *The international journal of tuberculosis and lung disease : the official journal of the International Union against Tuberculosis and Lung Disease.* 2017;21(5):483-484.

149. He G, Li Y, Zhao F, et al. The Prevalence and Incidence of Latent Tuberculosis Infection and Its Associated Factors among Village Doctors in China. *PLoS One.* 2015;10(5):e0124097.
150. Zhang X, Jia H, Liu F, et al. Prevalence and Risk Factors for Latent Tuberculosis Infection among Health Care Workers in China: A Cross-Sectional Study. *PLoS One.* 2013;8(6):e66412.
151. Yen YF, Hu BS, Lin YS, et al. Latent tuberculosis among injection drug users in a methadone maintenance treatment program, Taipei, Taiwan: TSPOT.TB versus tuberculin skin test. *Scand J Infect Dis.* 2013;45(7):504-511.
152. Adams S, Ehrlich R, Baatjies R, et al. Incidence of occupational latent tuberculosis infection in South African healthcare workers. *Eur Respir J.* 2015;45(5):1364-1373.
153. Frost WH. *Risk of Persons in Familial Contact with Pulmonary Tuberculosis.*
154. Cohn DL, O'Brien RJ, Geiter LJ, et al. *Supplement - American Thoracic Society Centers for Disease Control and Prevention - Targeted tuberculin testing and treatment of latent tuberculosis infection.*
155. Lewinsohn DM, Leonard MK, LoBue PA, et al. Official American Thoracic Society/Infectious Diseases Society of America/Centers for Disease Control and Prevention Clinical Practice Guidelines: Diagnosis of Tuberculosis in Adults and Children. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America.* 2017;64(2):111-115.
156. Guerra-Silveira F, Abad-Franch F. Sex bias in infectious disease epidemiology: patterns and processes. *PLoS One.* 2013;8(4):e62390.
157. Narasimhan P, Wood J, Macintyre CR, Mathai D. Risk factors for tuberculosis. *Pulmonary medicine.* 2013;2013:828939.
158. Oni T, Gideon HP, Bangani N, et al. Smoking, BCG and employment and the risk of tuberculosis infection in HIV-infected persons in South Africa. *PLoS One.* 2012;7(10):e47072.
159. Pelly TF, Santillan CF, Gilman RH, et al. Tuberculosis skin testing, anergy and protein malnutrition in Peru. *The international journal of tuberculosis and lung disease : the official journal of the International Union against Tuberculosis and Lung Disease.* 2005;9(9):977-984.
160. Wood R, Liang H, Wu H, et al. Changing prevalence of tuberculosis infection with increasing age in high-burden townships in South Africa. *The international journal of tuberculosis and lung disease : the official journal of the International Union against Tuberculosis and Lung Disease.* 2010;14(4):406-412.
161. Dodd PJ, Looker C, Plumb ID, et al. Age- and Sex-Specific Social Contact Patterns and Incidence of Mycobacterium tuberculosis Infection. *American journal of epidemiology.* 2016;183(2):156-166.
162. Fox GJ, Barry SE, Britton WJ, Marks GB. Contact investigation for tuberculosis: a systematic review and meta-analysis. *The European respiratory journal.* 2013;41(1):140-156.
163. Singh J, Sankar MM, Kumar S, et al. Incidence and prevalence of tuberculosis among household contacts of pulmonary tuberculosis patients in a peri-urban population of South Delhi, India. *PLoS One.* 2013;8(7):e69730.

164. Kirenga BJ, Ssengooba W, Muwonge C, et al. Tuberculosis risk factors among tuberculosis patients in Kampala, Uganda: implications for tuberculosis control. *BMC public health.* 2015;15:13.
165. Bosman MC, Swai OB, Kwamanga DO, Agwanda R, Idukitta G, Misljenovic O. National tuberculin survey of Kenya, 1986-1990. *The international journal of tuberculosis and lung disease : the official journal of the International Union against Tuberculosis and Lung Disease.* 1998;2(4):272-280.
166. Odhiambo JA, Borgdorff MW, Kiambih FM, et al. Tuberculosis and the HIV epidemic: increasing annual risk of tuberculous infection in Kenya, 1986-1996. *American journal of public health.* 1999;89(7):1078-1082.
167. Harries AD, Libamba E, Schouten EJ, Mwansambo A, Salaniponi FM, Mpazanje R. Expanding antiretroviral therapy in Malawi: drawing on the country's experience with tuberculosis. *BMJ (Clinical research ed).* 2004;329(7475):1163-1166.
168. Verver S, Warren RM, Munch Z, et al. Proportion of tuberculosis transmission that takes place in households in a high-incidence area. *Lancet.* 2004;363(9404):212-214.
169. Chamie G, Kato-Maeda M, Emperador DM, et al. Spatial overlap links seemingly unconnected genotype-matched TB cases in rural Uganda. *PLoS ONE.* 2018;13(2):e0192666.
170. Chen C, Zhu T, Wang Z, et al. High Latent TB Infection Rate and Associated Risk Factors in the Eastern China of Low TB Incidence. *PLoS One.* 2015;10(10):e0141511.
171. Lee SJ, Lee SH, Kim YE, et al. Risk factors for latent tuberculosis infection in close contacts of active tuberculosis patients in South Korea: a prospective cohort study. *BMC Infect Dis.* 2014;14:566.
172. Mumpe-Mwanja D, Verver S, Yeka A, et al. Prevalence and risk factors of latent Tuberculosis among adolescents in rural Eastern Uganda. *African health sciences.* 2015;15(3):851-860.
173. Tagawa H, Sugita H, Nakazono T, Takayanagi K, Yamaguchi T, Shimao T. [Association between smoking and tuberculosis infection]. *Kekkaku : [Tuberculosis].* 2014;89(11):803-806.
174. Bates MN, Khalakdina A, Pai M, Chang L, Lessa F, Smith KR. Risk of tuberculosis from exposure to tobacco smoke: a systematic review and meta-analysis. *Archives of internal medicine.* 2007;167(4):335-342.
175. Houben RM, Dodd PJ. The Global Burden of Latent Tuberculosis Infection: A Re-estimation Using Mathematical Modelling. *PLoS Med.* 2016;13(10):e1002152.
176. Getahun H, Matteelli A. Tailoring Treatment of Latent Tuberculosis to the Needs of Patients and Families. *Annals of internal medicine.* 2017;167(10):742-743.
177. Brewer TF, Heymann SJ. To control and beyond: moving towards eliminating the global tuberculosis threat. *Journal of epidemiology and community health.* 2004;58(10):822-825.
178. WHO. END TB STRATEGY 2015. 2015.
179. Golub JE, Mohan CI, Comstock GW, Chaisson RE. Active case finding of tuberculosis: historical perspective and future prospects. *The international journal of tuberculosis and lung disease : the official journal of the International Union against Tuberculosis and Lung Disease.* 2005;9(11):1183-1203.

180. Favorov M, Ali M, Tursunbayeva A, et al. Comparative Tuberculosis (TB) Prevention Effectiveness in Children of Bacillus Calmette-Guérin (BCG) Vaccines from Different Sources, Kazakhstan. *PLoS ONE.* 2012;7(3):e32567.
181. Roy A, Eisenhut M, Harris RJ, et al. Effect of BCG vaccination against Mycobacterium tuberculosis infection in children: systematic review and meta-analysis. *BMJ (Clinical research ed).* 2014;349:g4643.
182. Fox GJ, Nhung NV, Sy DN, et al. Household-Contact Investigation for Detection of Tuberculosis in Vietnam. *The New England journal of medicine.* 2018;378(3):221-229.
183. Pang YK. Close contact investigation of TB in high-burden, low- and middle-income countries. *Malaysian Family Physician : the Official Journal of the Academy of Family Physicians of Malaysia.* 2014;9(2):11-17.
184. Hsu KH. CONTACT INVESTIGATION: A PRACTICAL APPROACH TO TUBERCULOSIS ERADICATION. *American journal of public health and the nation's health.* 1963;53:1761-1769.
185. Ustero PA, Kay AW, Ngo K, et al. School and household tuberculosis contact investigations in Swaziland: Active TB case finding in a high HIV/TB burden setting. *PLoS One.* 2017;12(6):e0178873.
186. Beyers N, Gie RP, Schaaf HS, et al. A prospective evaluation of children under the age of 5 years living in the same household as adults with recently diagnosed pulmonary tuberculosis. *The international journal of tuberculosis and lung disease : the official journal of the International Union against Tuberculosis and Lung Disease.* 1997;1(1):38-43.
187. International I. Demographic and Health Survey Sampling and Household Listing Manual
. . *MEASURE DHS, Calverton, Maryland, USA: ICF International* 2012.
188. WHO. Latent tuberculosis infection: updated and consolidated guidelines for programmatic management. 2018.
189. Uganda Bureau of Statistics Kampala U. Demographic and Health Survey 2016.
190. WHO Guidelines Approved by the Guidelines Review Committee. In: *Recommendations for Investigating Contacts of Persons with Infectious Tuberculosis in Low- and Middle-Income Countries.* Geneva: World Health Organization

APPENDIX A

SAS PROC GENMOD'S ANALYSIS CODE FOR AIM ONE

```
Libname Libref 'P:\\MANUSCRIPTS DATASETS\\LTBI_Robert\\Dissertation
datasets\\Analysis2\\';

/*Droping unwanted variables*/
data Data2   (drop=Enroll_dob Enroll_lastname enroll_firstname enroll_othname
Enroll_int enroll_bidabst enroll_pic enroll_picabst enroll_init enroll_rev
enroll_dmo enroll_cd enroll_compdate
enroll_revdate enroll_dmodate enroll_cddate enroll_time_stamp enroll_x32
enroll_x33 X enroll_x1 enroll_csnt enroll_live enroll_treat enroll_tbtreat
enroll_daystreat enroll_title enroll_lnkid
 enroll_formid enroll_serial enroll_bioid);set Data2;run;
/*Checking distribution of remaing variables*/
ods rtf file='P:\MANUSCRIPTS DATASETS\LTBI_Robert\Dissertation
datasets\Analysis2\Completeness Assessment.rtf';
proc freq data=Data2;table enroll_cat enroll_sex census_cat DM_Income DM_read
DM_write DM_married DM_religion TST_PPDGiven;run;;Title 'Gender by HHold
Status';run;ods rtf close;run;

/*Creating New dataset for Case contacts only and control contacts only.*/
Data casecontacts controlcontacts;set contactsdataset;if Index_contact = 1
then output casecontacts;else if index_contact =2 then output
controlcontacts;run;
proc means data=casecontacts;
variable Enroll_age;run;
proc means data=controlcontacts;
variable Enroll_age;run;

/*Creating new variable TSTResult*/
data Data2;set Data2;TSTResult=(TST_PPD_1+TST_PPD_2)/2;RUN;
/*Creating new variable "outcome" combining PPDgiven and TSTresult*/
data Data2;set Data2;Outcome='.';if TST_PPDGiven= '3' then outcome='3';Else
if TST PPDGiven ='2' then outcome='4';Else if TST PPDGiven ='4' then
outcome='6';if Tstresult GE 10 and Tstresult LT 30 then outcome= 1;Else if
Tstresult GE 77 then outcome=5;Else if Tstresult GE 0 and TSTresult LT 10then
outcome=2;run;run;;Title 'Outcome Variable created';run;ods rtf close;run;
proc freq data=Data2;table Outcome;run;ods rtf close;run;
/*Creating new variable "HIVStatus"*/
data Data2;set Data2;HIVStatus='.';if HIVt= '2' then HIVstatus='4';Else if
HIVt ='3' then HIVstatus='3';Else if HIVt ='7' then HIVStatus='7';if hivr =
'1' then HivStatus= '1';Else if Hivr=2 then HivStatus='2';run;
proc freq data=Data2;table HIVStatus;run;
/*Creating data set to check which HIVStatus values are missing*/
proc freq data=Data2;table HIVStatus;run;
Data HIVmissing;set Data2;if HIVStatus='.';run;
```

```
proc freq data=Data2;table HIVStatus;run;Title 'Hivstatus Variable
created';run;
/*Creating data set to check which outcome values are missing*/
proc freq data=Data2;table outcome;run;
Data missing;set Dataa;if outcome='.';run;
proc freq data=Dataa;table outcome;run;Title 'Outcome Variable created';run;
/*Creating new variable "Education"*/
data Data2;set Data2;
EducationLevel='.';
if education='1' then EducationLevel='1'; if education='4' then
EducationLevel='2'; if education='2' then EducationLevel='3';if education='3'
then EducationLevel='3';Else if education ='5' then EducationLevel='3';run;
proc freq data=Data2;table EducationLevel;run;
/*Creating new variable "AgeCategory", AgeCDC, AgeUGA*/
data Data2;set Data2;
AgeCategory='.';
If Enroll_age GE 0 and Enroll_age LT 10 then AgeCategory=1;Else if Enroll_age
GE 10 and Enroll_age LT 20 then AgeCategory=2;Else if Enroll_age GE 20 and
Enroll_age LT 30 then AgeCategory=3;Else if Enroll_age GE 30 and Enroll_age
LT 40 then AgeCategory=4;Else if Enroll_age GE 40 and Enroll_age LT 50 then
AgeCategory=5;Else if Enroll_age GE 50 then AgeCategory=6;run;
proc freq data=Data2;table AgeCategory;run;
proc freq data=Data2;table AgeCategory;run;Title 'AgeCategory created';run;
data data2;set Data2;
AgeCDC='.';
If Enroll_age GE 0 and Enroll_age LT 6 then AgeCDC=1;Else if Enroll_age GE 6
and Enroll_age LT 16 then AgeCDC=2;Else if Enroll_age GE 16 and Enroll_age LT
26 then AgeCDC=3;Else if Enroll_age GE 26 and Enroll_age LT 36 then
AgeCDC=4;Else if Enroll_age GE 36 and Enroll_age LT 46 then AgeCDC=5;Else if
Enroll_age GE 46 then AgeCDC=6;run;
proc freq data=Data2;table AgeCDC;run;
data Data2;set Data2;
AgeUGA='.';
If Enroll_age GE 0 and Enroll_age LT 6 then AgeUGA=1;Else if Enroll_age GE 6
and Enroll_age LT 16 then AgeUGA=2;Else if Enroll_age GE 16 and Enroll_age LT
26 then AgeUGA=3;Else if Enroll_age GE 26 then AgeUGA=4;run;
proc freq data=Data2;table AgeUGA;run;
data Data2;set Data2;
AgeBinary='.';
If Enroll_age GE 0 and Enroll_age LT 15 then AgeBinary=1;Else if Enroll_age
GE 15 then AgeBinary=2;run;
proc freq data=Data2;table AgeBinary;run;
*/Creating new variables MaritalStatus, SES, EducFinal, Occupation */;
data Data2;set Data2;
MaritalStatus='.';
If DM_married = 2 then MaritalStatus=1;Else if DM_married = 3 then
MaritalStatus=1; Else if DM_married = 1 then MaritalStatus=2;Else if
DM_married = 4 then MaritalStatus=2; Else if DM_married = 5 then
MaritalStatus=2; Else if DM_married = 7 then MaritalStatus=2;run;
proc freq data=Data2;table MaritalStatus;run;
data Data2;set Data2;
SES='.';
If DM_Income = 1 then SES=1;Else if DM_Income = 2 then SES=1; Else if
DM_Income = 3 then SES=1;Else if DM_Income = 4 then SES=1; Else if DM_Income
= 5 then SES =2;run;
proc freq data=Data2;table SES;run;
```

```
data Data2;set Data2;
Occupation='.';
if DM_occup1 in (3,4,6,8,14,16,18,19) then Occupation=1; *If you have more
than one value (3,4,6,8...) should have to use in instead of '=';
Else if DM_occup1 in  (5,9,10,11,13,12) then Occupation=2;
else if DM_occup1 in  (1,2,7,17,20,21,22) then Occupation=3;run;
proc freq data=Data2;table Occupation;run;
/*Creating new variable Contacts and Indexes*/
Data Data2;set Data2;Contacts='.';if True_index = '0' then contacts='1';run;
proc freq data=Data2;table contacts;run;
Data Data2;set Data2;INDEXES='.';if True_index = 1 then INDEXES=1; else if
True_index = 2 then INDEXES=2;run;
proc freq data=Data2;table INDEXES;run;
/*Creating new variable HIV*/
data Data2;set Data2;
HIV='.';
If Hivstatus='.' then HIV=2; if Hivstatus = 1 then HIV=1;Else if HIVstatus GE
2 then HIV=2;run;
proc freq data=Data2;table HIV;run;
data contactsdataset;set contactsdataset;
HIV='.';
If Hivstatus='.' then HIV=2; if Hivstatus = 1 then HIV=1;Else if HIVstatus GE
2 then HIV=2;run;
proc freq data=Data2;table HIV;run;
/*Creating New dataset for contacts ONLY.There are 2087 contacts*/
Data ContactsDataset;set Data2;if contacts = 1 then output
ContactsDataset;run;
/*Creating New dataset for INDEXES ONLY.*/
Data INDEXESDataset;set Data2;if Indexes = 1 then output INDEXESDataset;else
if indexes =2 then output INDEXESDataset;run;
Data TBInfections;set Data2;if outcome = 3 then output TBInfections;run;
/*Creating New dataset for those reporting Past TB infection.*/

ods rtf file='P:\MANUSCRIPTS DATASETS\LTBI_Robert\Dissertation
datasets\analysis2\Tables\Table 1(A).rtf';
/*Completing Table 1(a)*/
proc freq data=Indexesdataset;tables indexes*Enroll_sex/cmh;
proc freq data=Indexesdataset;tables indexes*AgeCategory/cmh;
proc freq data=Indexesdataset;tables indexes*HIVstatus/cmh;
proc freq data=Indexesdataset;tables indexes*MaritalStatus/cmh;
proc freq data=Indexesdataset;tables indexes*DM_religion/cmh;
proc freq data=Indexesdataset;tables indexes*EducationLevel/cmh;
proc freq data=Indexesdataset;tables indexes*SES/cmh;run;
proc freq data=Indexesdataset;tables indexes*Occupation/cmh;
run;ods rtf close;run;

/*Completing Table 1(b)*/
ods rtf file='P:\MANUSCRIPTS DATASETS\LTBI_Robert\Dissertation
datasets\analysis2\Tables\Table 1(B).rtf';
proc freq data=Contactsdataset;tables index_contact*Enroll_sex/cmh;
proc freq data=Contactsdataset;tables index_contact*AgeCategory/cmh;
proc freq data=Contactsdataset;tables index_contact*Agebinary/cmh;run;
proc freq data=Casecontacts;tables Index_contacts*Agebinary/cmh;
proc freq data=Contactsdataset;tables index_contact*census_cat/cmh;
proc freq data=Contactsdataset;tables index_contact*HIVstatus/cmh;
proc freq data=Contactsdataset;tables index_contact*MaritalStatus/cmh;
```

```
proc freq data=Contactsdataset;tables index_contact*DM_religion/cmh;
proc freq data=Contactsdataset;tables index_contact*Educationlevel/cmh;
proc freq data=Contactsdataset;tables index_contact*DM_Income/chisq;run;
proc freq data=Contactsdataset;tables index_contact*SES/cmh;
proc freq data=Contactsdataset;tables index_contact*Occupation/cmh;
run;ods rtf close;run;

proc univariate data=Contactsdataset plot;var Enroll_age;histogram
Enroll_age;run; *Examining the Distribution of continous variables;
proc univariate data=Contactsdataset plot;var DM_Income;histogram
DM_Income;run;

proc freq data=data2;
where (True_Index=0) and (index_contact in (1,2)); *Only contacts, not index.
We have contacts of cases and controls;
Tables Enroll_sex*census_cat/chisq exact;
title 'Proportion males and females by Household status';
title2 ' In the contacts';
by index_contact;run;
proc freq data=data2;
where index_contact in (1,2); *Just including in the proc freq, the
index_contact that are either 1 or 2;
Tables True_index*index_contact;run;
proc sort data=data2;
by index_contact;
run;
proc freq data=data2;
tables index_contact;
run;
proc freq data=data2;
tables index_contact*True_Index;
run;
proc freq data=data2;
tables index_contact*census_cat;
run;
proc freq data=data2;
tables index_contact*Enroll_Sex;
run;
proc contents data=data2;
run;

ods rtf file='P:\MANUSCRIPTS DATASETS\LTBI_Robert\Dissertation
datasets\analysis2\Tables\Main Results.rtf';
proc freq data=data2;
where (True_Index=0) and (index_contact in (1,2)); *MAIN RESULT Proportion of
Infection by hhold status among contacts. We have contacts of cases and
controls;
Tables census_cat*outcome/chisq;
title 'Proportion Infection by Household status';
title2 ' In the contacts';
by index_contact;run;ods rtf close;run;
 ods rtf file='P:\MANUSCRIPTS DATASETS\LTBI_Robert\Dissertation
datasets\analysis2\Tables\Main Results by Gender.rtf';
proc freq data=data2;
where (True_Index=0) and (index_contact in (1,2)); *gender distribution among
contacts. We have contacts of cases and controls;
```

```
Tables enroll_sex*census_cat*outcome/chisq;
title 'Proportion Infection by sex';
title2 ' In the contacts';
by index_contact;run;ods rtf close;run;
ods rtf file='P:\MANUSCRIPTS DATASETS\LTBI_Robert\Dissertation
datasets\analysis2\Tables\Main Results by AGEBinary.rtf';
proc freq data=data2;
where (True_Index=0) and (index_contact in (1,2)); *Binary age distribution
among contacts. We have contacts of cases and controls;
Tables Agebinary*census_cat*outcome/chisq;
title 'Proportion Infection by AgeUGA';
title2 ' In the contacts';
by index_contact;run;ods rtf close;run;
ods rtf file='P:\MANUSCRIPTS DATASETS\LTBI_Robert\Dissertation
datasets\analysis2\Tables\RiskFactors Table 4.rtf';
proc freq data=data2;
where (True_Index=0) and (index_contact in (1,2));
Tables agecategory*outcome;
title 'Proportion HH/NHH contacts of index controls';
title2 ' By agecategory';
by index_contact;run;
proc freq data=data2;
where (True_Index=0) and (index_contact in (1,2));
Tables agebinary*outcome;
title 'Proportion HH/NHH contacts of index controls';
title2 ' By agebinary';
by index_contact;run;
proc freq data=data2;
where (True_Index=0) and (index_contact in (1,2));
Tables Enroll_sex*outcome;
title 'Proportion HH/NHH contacts of index controls';
title2 ' By Sex';
by index_contact;run;
proc freq data=data2;
where (True_Index=0) and (index_contact in (1,2));
Tables HIVStatus*outcome;
title 'Proportion HH/NHH contacts of index controls';
title2 ' By HIV status';
by index_contact;run;
proc freq data=data2;
where (True_Index=0) and (index_contact in (1,2));
Tables sym_hadtb*outcome;
title 'Proportion males and females by Household status';
title2 ' By HIV had TB variable';
by index_contact;run;
proc freq data=data2;
where (True_Index=0) and (index_contact in (1,2));
Tables sym_TBcontact*outcome;
title 'Proportion males and females by Household status';
title2 ' By TB contact variable';
by index_contact;run;
proc freq data=data2;
where (True_Index=0) and (index_contact in (1,2));
Tables sym_KnewTB*outcome;
title 'Proportion males and females by Household status';
title2 ' By Knew TB variable';
```

```
by index_contact;run;
proc freq data=data2;
where (True_Index=0) and (index_contact in (1,2));
Tables sym_TBhome*outcome;
title 'Proportion males and females by Household status';
title2 ' By TB home';
by index_contact;run;
proc freq data=data2;
where (True_Index=0) and (index_contact in (1,2));
Tables sym_habit*outcome;
title 'Proportion males and females by Household status';
title2 ' By habit variable';
by index_contact;run;
proc freq data=data2;
where (True_Index=0) and (index_contact in (1,2));
Tables sym_drink*outcome;
title 'Proportion males and females by Household status';
title2 ' By drink variable';
by index_contact;run;
proc freq data=data2;
where (True_Index=0) and (index_contact in (1,2));
Tables sym_tbvac*outcome;
title 'Proportion males and females by Household status';
title2 ' By TB vaccine variable';
by index_contact;run;
proc freq data=data2;
where (True_Index=0) and (index_contact in (1,2));
Tables sym_hiv*outcome;
title 'Proportion males and females by Household status';
title2 ' by HIV symptom variable';
by index_contact;run;ods rtf close;run;


ods rtf file='P:\MANUSCRIPTS DATASETS\LTBI_Robert\Dissertation
datasets\analysis2\Tables\RiskFactors Table 5.rtf';
proc freq data=data2;
where (True_Index=0) and (index_contact in (1,2));
Tables agecategory*census_cat*outcome;
title 'Proportion HH/NHH contacts of index controls';
title2 ' By agecategory';
by index_contact;run;
proc freq data=data2;
where (True_Index=0) and (index_contact in (1,2));
Tables agebinary*census_cat*outcome;
title 'Proportion HH/NHH contacts of index controls';
title2 ' By agebinary';
by index_contact;run;
proc freq data=data2;
where (True_Index=0) and (index_contact in (1,2));
Tables Enroll_sex*census_cat*outcome;
title 'Proportion HH/NHH contacts of index controls';
title2 ' By Sex';
by index_contact;run;
proc freq data=data2;
where (True_Index=0) and (index_contact in (1,2));
Tables HIVStatus*census_cat*outcome;
```

```
title 'Proportion HH/NHH contacts of index controls';
title2 ' By HIV status';
by index_contact;run;
ods rtf file='P:\MANUSCRIPTS DATASETS\LTBI_Robert\Dissertation
datasets\analysis2\Tables\RiskFactors5.rtf';
proc freq data=data2;
where (True_Index=0) and (index_contact in (1,2));
Tables sym_hadtb*census_cat*outcome;
title 'Proportion males and females by Household status';
title2 ' By HIV had TB variable';
by index_contact;run;
proc freq data=data2;
where (True_Index=0) and (index_contact in (1,2));
Tables sym_TBcontact*census_cat*outcome;
title 'Proportion males and females by Household status';
title2 ' By TB contact variable';
by index_contact;run;ods rtf close;run;
proc freq data=data2;
where (True_Index=0) and (index_contact in (1,2));
Tables sym_KnewTB*census_cat*outcome;
title 'Proportion males and females by Household status';
title2 ' By Knew TB variable';
by index_contact;run;
proc freq data=data2;
where (True_Index=0) and (index_contact in (1,2));
Tables sym_TBhome*census_cat*outcome;
title 'Proportion males and females by Household status';
title2 ' By TB home';
by index_contact;run;
proc freq data=data2;
where (True_Index=0) and (index_contact in (1,2));
Tables sym_habit*census_cat*outcome;
title 'Proportion males and females by Household status';
title2 ' By habit variable';
by index_contact;run;
proc freq data=data2;
where (True_Index=0) and (index_contact in (1,2));
Tables sym_drink*census_cat*outcome;
title 'Proportion males and females by Household status';
title2 ' By drink variable';
by index_contact;run;
proc freq data=data2;
where (True_Index=0) and (index_contact in (1,2));
Tables sym_tbvac*census_cat*outcome;
title 'Proportion males and females by Household status';
title2 ' By TB vaccine variable';
by index_contact;run;
proc freq data=data2;
where (True_Index=0) and (index_contact in (1,2));
Tables sym_hiv*outcome;
title 'Proportion males and females by Household status';
title2 ' by HIV symptom variable';
by index_contact;run;ods rtf close;run;


Data overlaps; *new data set ;
```

```
Set data2; /*main data set*/
if overlap > 1; /* I used the overlap variable that was created previous */
run;
PROC SORT data= overlaps out= uniqueoverlaps  NODUPKEY;
BY uniqueid;
RUN;

 ods rtf file='P:\MANUSCRIPTS DATASETS\LTBI_Robert\Dissertation
datasets\analysis2\Tables\Table 1(A).rtf';
/*Completing Table 1(a)*/
proc freq data=uniqueoverlaps;tables indexes*Enroll_sex/cmh;
proc freq data=uniqueoverlaps;tables indexes*AgeCategory/cmh;
proc freq data=uniqueoverlaps;tables indexes*HIVstatus/cmh;
proc freq data=uniqueoverlaps;tables indexes*MaritalStatus/cmh;
proc freq data=uniqueoverlaps;tables indexes*DM_religion/cmh;
proc freq data=uniqueoverlaps;tables indexes*EducationLevel/cmh;
proc freq data=uniqueoverlaps;tables indexes*SES/cmh;run;
proc freq data=uniqueoverlaps;tables indexes*Occupation/cmh;
run;ods rtf close;run;


ods rtf file='P:\MANUSCRIPTS DATASETS\LTBI_Robert\Dissertation
datasets\analysis2\Tables\Table 1(A).rtf';
/*Completing Table 1(a)*/
proc freq data=overlaps;tables indexes*Enroll_sex/cmh;
proc freq data=overlaps;tables indexes*AgeCategory/cmh;
proc freq data=overlaps;tables indexes*HIVstatus/cmh;
proc freq data=overlaps;tables indexes*MaritalStatus/cmh;
proc freq data=overlaps;tables indexes*DM_religion/cmh;
proc freq data=overlaps;tables indexes*EducationLevel/cmh;
proc freq data=overlaps;tables indexes*SES/cmh;run;
proc freq data=overlaps;tables indexes*Occupation/cmh;
run;ods rtf close;run;

proc freq data=uniqueoverlaps;table enroll_sex AgeCategory HIVstatus
Maritalstatus DM_religion EducationLevel SES Occupation;run;
ods rtf file='P:\MANUSCRIPTS DATASETS\LTBI_Robert\Dissertation
datasets\analysis2\Tables\PastTBInfected.rtf';
proc freq data=TBinfections;table enroll_sex AgeCategory agebinary census_cat
Index_contact sym_hadtb sym_tbcontact sym_knewtb sym_tbhome sym_tbever;run;
run;ods rtf close;run;

ods rtf file='P:\MANUSCRIPTS DATASETS\LTBI_Robert\Dissertation
datasets\analysis2\Tables\Table4 results.rtf';
proc freq data=contactsdataset;
where (True_Index=0) and (index_contact=1); *Only contacts of index cases;
Tables agecategory*census_cat*outcome; *Chi square and fisher
title 'Proportion HH/NHH contacts of index controls';
title2 ' By agecdc';
proc freq data=contactsdataset;
where (True_Index=0) and (index_contact=1); *Only contacts of index cases;
Tables Enroll_sex*census_cat*outcome; *Chi square and fisher
title 'Proportion HH/NHH contacts of index controls';
title2 ' By Sex';
proc freq data=contactsdataset;
where (True_Index=0) and (index_contact=1); *Only contacts of index cases;
```

```
Tables HIVStatus*census_cat*outcome; *Chi square and fisher
title 'Proportion HH/NHH contacts of index controls';
title2 ' By HIV status';
run;ods rtf close;run;

ods rtf file='P:\MANUSCRIPTS DATASETS\LTBI_Robert\Dissertation datasets\Final
results folder\ Table 3 CONTROL CONTACTS.rtf';
proc freq data=contacts;
where (True_Index=0) and (index_contact=2); *Only control contacts;
Tables agecdc*census_cat*outcome; *Chi square and fisher
title 'Proportion HH/NHH contacts of index controls';
title2 ' By agecdc';
proc freq data=contacts;
where (True_Index=0) and (index_contact=2); *Only control contacts;
Tables Enroll_sex*census_cat*outcome; *Chi square and fisher
title 'Proportion HH/NHH contacts of index controls';
title2 ' By Sex';
proc freq data=contacts;
where (True_Index=0) and (index_contact=2); *Only control contacts;
Tables HIVStatus*census_cat*outcome; *Chi square and fisher
title 'Proportion HH/NHH contacts of index controls';
title2 ' By HIV status';
run;ods rtf close;

ods rtf file='P:\MANUSCRIPTS DATASETS\LTBI_Robert\Dissertation datasets\Final
results folder\ Model Building.rtf';
proc logistic data=contactsdataset;
class outcome enroll_sex/ param=ref ref=first;
model outcome = enroll_sex;
title 'Crude Model';
run;ods rtf close;

ods rtf file='P:\MANUSCRIPTS DATASETS\LTBI_Robert\Dissertation datasets\Final
results folder\ VariableX is related to exposure.rtf';
proc freq data=contactsdataset;
tables Enroll_sex*census_cat/or chisq cmh;run;
tables AgeCategory*census_cat/or chisq cmh;
tables HIVstatus*census_cat/or chisq cmh;
tables MaritalStatus*census_cat/or chisq cmh;
tables DM_religion*census_cat/or chisq cmh;
tables EducationLevel*census_cat/or chisq cmh;
tables SES*census_cat/or chisq cmh;
tables Occupation*census_cat/or chisq cmh;
tables sym_hadtb*census_cat/or chisq cmh;
tables sym_tbcontact*census_cat/or chisq cmh;
tables sym_knewtb*census_cat/or chisq cmh;
tables sym_tbhome*census_cat/or chisq cmh;
tables sym_tbever*census_cat/or chisq cmh;
tables sym_tbvac*census_cat/or chisq cmh;
tables sym_habit*census_cat/or chisq cmh;
tables sym_drink*census_cat/or chisq cmh;
tables Index_contact*census_cat/or chisq cmh;
Title 'Determine if VariableX is related to exposure';
format census_cat census_cat. Enroll_sex Enroll_sex. AgeCategory
$AgeCategory. outcome $outcome. HIVstatus $HIVstatus.SES $SES.Maritalstatus
$Maritalstatus. occupation $occupation.EducationLevel $EducationLevel.
```

```
sym_knewtb $sym_knewtb. sym_drink sym_drink. sym_habit sym_habit. sym_tbvac
sym_tbvac. Index_contact Index_contact.;
run;ods rtf close;

ods rtf file='P:\MANUSCRIPTS DATASETS\LTBI_Robert\Dissertation datasets\Final
results folder\ VariableX is related to Outcome.rtf';
proc freq data=contactsdataset;
tables Enroll_sex*outcome/or chisq cmh;
tables AgeCategory*outcome/or chisq cmh;
tables HIVstatus*outcome/or chisq cmh;
tables MaritalStatus*outcome/or chisq cmh;
tables DM_religion*outcome/or chisq cmh;
tables EducationLevel*outcome/or chisq cmh;
tables SES*outcome/or chisq cmh;
proc freq data=contactsdataset;
tables Occupation*outcome/or chisq cmh;run;
proc freq data=contactsdataset;
tables educationlevel*outcome/or chisq cmh;run;
tables sym_hadtb*outcome/or chisq cmh;
tables sym_tbcontact*outcome/or chisq cmh;
tables sym_knewtb*outcome/or chisq cmh;
tables sym_tbhome*outcome/or chisq cmh;
tables sym_tbever*outcome/or chisq cmh;
tables sym_tbvac*outcome/or chisq cmh;
tables sym_habit*outcome/or chisq cmh;
tables sym_drink*outcome/or chisq cmh;
tables Index_contact*outcome/or chisq cmh;
Title 'Determine if VariableX is related to Outcome';
format census_cat census_cat. Enroll_sex Enroll_sex. AgeCategory
$AgeCategory. outcome $outcome. HIVstatus $HIVstatus.SES $SES.Maritalstatus
$Maritalstatus. occupation $occupation.EducationLevel $EducationLevel.
sym_knewtb $sym_knewtb. sym_drink sym_drink. sym_habit sym_habit. sym_tbvac
sym tbvac. Index contact Index_contact.;
run;ods rtf close;

proc freq data=Controlcontacts;tables census_cat*outcome;run;
proc freq data=Controlcontacts;tables enroll_sex*outcome;run;
proc freq data=Controlcontacts;tables agecategory*outcome;run;
proc freq data=Controlcontacts;tables Hivstatus*outcome;run;

proc freq data=Casecontacts;tables census_cat*outcome;run;
proc freq data=Casecontacts;tables enroll_sex*outcome;run;
proc freq data=Casecontacts;tables agecategory*outcome;run;
proc freq data=Casecontacts;tables Hivstatus*outcome;run;

Title 'Determine if VariableX is related to Outcome';
run;ods rtf close;

proc format;
value census_cat
1='Household'
2='Non-household';
value Enroll_sex
1='Men'
2='Female';
value $AgeCategory
```

```
1='0-5 years'
2='6-15 years'
3='16-25 years'
4='26-35 years'
5='36-45 years'
6='45-+ years';
value $outcome
1='TST Positive'
2='TST Negative'
3='TST not given';
value $HIVstatus
1='Positive'
2='Negative'
3='Refused Test'
4='Other';
value $SES
1='Less than $80'
2='$80 or more';
.= 'missing'
value Maritalstatus
1='Married'
2='Not Married';run;
value $occupation
1='Indoor'
2='Semi-Indoor'
3='Out door';
value $EducationLevel
1='No education'
2='Primary'
3='Post Primary';
value $sym_knewtb
1='Yes'
2='No'
7='Uncertain';
value $sym_tbhome
1='Yes'
2='No'
7='Uncertain';
value sym_drink
1='Yes'
2='No'
.='Missing';
value sym_habit
1='Yes'
2='No'
3='Uncertain';
value sym_tbvac
1='Yes confirmed by card'
2='Yes verbal report'
3='No vaccine'
77='Idont know';
value Index_contact
1='Case contact'
2='Control contact';
run;
value $BCG
```

```
1='Vaccinated'
2='Not vaccinated'
3='Dont know';run;


proc contents data=contactsdataset;
run;
ods rtf file='P:\MANUSCRIPTS DATASETS\LTBI_Robert\Dissertation
datasets\analysis2\Tables\Univariate Analysis.rtf';
/* Model building using GENMOD */
proc freq data=contactsdataset;
proc genmod data=contactsdataset;
where outcome in ('1','2');
class census_cat (ref='2');
model outcome (event='1')=census cat/dist=bin link=log;
estimate 'PR hh versus nhh' census_cat 1 -1/e exp;
title 'Crude Prevalence ratio by census_cat'
run;
proc freq data=contactsdataset;
where outcome in ('1','2');
tables census_cat*outcome/or;
format outcome $outcome. census_cat census_cat.;
title 'Crude OR Exposure versus outcome';
run;
proc genmod data=contactsdataset;
where outcome in ('1','2');
class Enroll_sex (ref='2');
model outcome (event='1')=Enroll_sex/dist=bin link=log;
estimate 'PR Male versus female' Enroll_sex 1 -1/e exp;
title 'Crude Prevalence ratio by Enroll_sex'
run;
proc genmod data=contactsdataset;
where outcome in ('1','2');
class AgeCategory (ref='1');
model outcome (event='1')=AgeCategory/dist=bin link=log;
estimate 'PR Age cat 2 versus 1' AgeCategory 1 0 0 0 0 -1/e exp;
estimate 'PR Age cat 3 versus 1' AgeCategory 0 1 0 0 0 -1/e exp;
estimate 'PR Age cat 4 versus 1' AgeCategory 0 0 1 0 0 -1/e exp;
estimate 'PR Age cat 5 versus 1' AgeCategory 0 0 0 1 0 -1/e exp;
estimate 'PR Age cat 6 versus 1' AgeCategory 0 0 0 0 1 -1/e exp;
title 'Crude Prevalence ratio by AgeCategory';
run;
proc genmod data=contactsdataset;
where outcome in ('1','2');
class Index_contact (ref='2');
model outcome (event='1')=Index_contact/dist=bin link=log;
estimate 'PR Cases versus controls' Index_contact 1 -1/e exp;
title 'Crude Prevalence ratio by contact status'run;
proc genmod data=contactsdataset;
where outcome in ('1','2');
class SES(ref='2');
model outcome (event='1')=SES/dist=bin link=log;
estimate 'PR >=80 versus <80' SES 1 -1/e exp;
title 'Crude Prevalence ratio by Income'run;
proc genmod data=contactsdataset;
where outcome in ('1','2');
```

```
class Maritalstatus (ref='2');
model outcome (event='1')=maritalstatus/dist=bin link=log;
estimate 'Married versus not married' maritalstatus 1 -1/e exp;
title 'Crude Prevalence ratio by maritalstatus'run;
proc genmod data=contactsdataset;
where outcome in ('1','2');
class Hivstatus (ref='2');
model outcome (event='1')=Hivstatus/dist=bin link=log;
estimate 'PR positives 1 versus 2' Hivstatus 1 0 -1/e exp;
estimate 'PR Refusals 3 versus 2' Hivstatus 0 1 -1/e exp;
estimate 'Other 4 versus 2' Hivstatus 0 0 1 -1/e exp;
title 'Crude Prevalence ratio by Hiv status';
run;
proc genmod data=contactsdataset;
where outcome in ('1','2');
class Educationlevel (ref='3');
model outcome (event='1')=Educationlevel/dist=bin link=log;
estimate 'PR No education 1 versus 3' Educationlevel 1 0 -1/e exp;
estimate 'PR Primary 2 versus 3' Educationlevel 0 1 -1/e exp;
title 'Crude Prevalence ratio by Education Level';
run;
proc genmod data=contactsdataset;
where outcome in ('1','2');
class sym_habit (ref='3');
model outcome (event='1')=sym_habit/dist=bin link=log;
estimate 'PR No current smoker 1 versus 3' sym_habit 1 0 -1/e exp;
estimate 'PR Former smoker 2 versus 3' sym_habit 0 1 -1/e exp;
title 'Crude Prevalence ratio by Smoking status';
run;
proc genmod data=contactsdataset;
where outcome in ('1','2');
class Occupation (ref='3');
model outcome (event='1')=occupation/dist=bin link=log;
estimate 'PR Out doors 1 versus 3' occupation 1 0 -1/e exp;
estimate 'PR Semi-indoors 2 versus 3' occupation 0 1 -1/e exp;
title 'Crude Prevalence ratio by occupation';
run;
proc genmod data=contactsdataset;
where outcome in ('1','2');
class EducationLevel (ref='3');
model outcome (event='1')=Educationlevel/dist=bin link=log;
estimate 'PR Never schooled 1 versus 3' Educationlevel 1 0 -1/e exp;
estimate 'PR Primary 2 versus 3' Educationlevel 0 1 -1/e exp;
title 'Crude Prevalence ratio by EducationLevel';
run;

/*Creating new variable "BCG"*/
data contactsdataset;set contactsdataset;
BCG='.';
if sym_Tbvac='1' then BCG='1'; if sym_Tbvac='2' then BCG='1'; if
sym_Tbvac='3' then BCG='2'; Else if sym_Tbvac ='77' then BCG='3';run;
proc freq data=contactsdataset;table BCG;run;
proc genmod data=contactsdataset;
where outcome in ('1','2');
class BCG (ref='1');
model outcome (event='1')=BCG/dist=bin link=log;
```

```
estimate 'Not vaccinated 2 versus 1' BCG 0 1 -1/e exp;
estimate 'Dont know 3 versus 1' BCG 0 1 -1/e exp;
title 'Crude Prevalence ratio by BCG status';
run;
proc genmod data=contactsdataset;
where outcome in ('1','2');
class sym_Tbhome (ref='2');
model outcome (event='1')=sym_Tbhome/dist=bin link=log;
estimate 'shared home with case 1 versus 2' sym_Tbhome 1 0 -1/e exp;
estimate 'Dont know shared home 7 versus 2' sym_Tbhome 0 1 -1/e exp;
title 'Crude Prevalence ratio by home sharing';
run;
proc genmod data=contactsdataset;
where outcome in ('1','2');
class sym_KnewTb (ref='2');
model outcome (event='1')=sym_KnewTb/dist=bin link=log;
estimate 'Knew someone with TB 1 versus 2' sym_KnewTb 1 0 -1/e exp;
estimate 'Dont know someone with TB 7 versus 2' sym_KnewTb 0 1 -1/e exp;
title 'Crude Prevalence ratio by knoew someone with TB';
run;
proc genmod data=contactsdataset;
where outcome in ('1','2');
class sym_Tbever (ref='2');
model outcome (event='1')=sym_Tbever/dist=bin link=log;
estimate 'Ever lived with TB case 1 versus 2' sym_Tbever 1 0 -1/e exp;
estimate 'Dont remember living with case 77 versus 2' sym_Tbever 0 1 -1/e
exp;
title 'Crude Prevalence ratio by ever lived with TB';
run;
proc genmod data=contactsdataset;
where outcome in ('1','2');
class sym_habit(ref='2');
model outcome (event='1')=sym habit/dist=bin link=log;
estimate 'PR Smokers) versus Non smoker' sym_habit 1 -1/e exp;
title 'Crude Prevalence ratio by Smoking status'run;
proc genmod data=contactsdataset;
where outcome in ('1','2');
class sym_drink(ref='2');
model outcome (event='1')=sym drink/dist=bin link=log;
estimate 'PR alcohol use versus No alcohol' sym_drink 1 -1/e exp;
title 'Crude Prevalence ratio by drinking status'run;
proc genmod data=contactsdataset;
where outcome in ('1','2');
class sym_tbhome(ref='2');
model outcome (event='1')=sym_tbhome/dist=bin link=log;
estimate 'PR In home versus None in home' sym_tbhome 1 -1/e exp;
title 'Crude Prevalence ratio by sym_tbhome'run;
proc genmod data=contactsdataset;
where outcome in ('1','2');
class sym_knewtb(ref='2');
model outcome (event='1')=sym_knewtb/dist=bin link=log;
estimate 'PR Knew case versus Never knewCase' sym_knewtb 1 -1/e exp;
title 'Crude Prevalence ratio by sym_knewtb'run;
proc genmod data=contactsdataset;
where outcome in ('1','2');
class occupation (ref='3');
```

```
model outcome (event='1')=occupation/dist=bin link=log;
estimate 'indoors 1 versus 3' occupation 1 0 -1/e exp;
estimate 'semindoors 2 versus 3' occupation 0 1 -1/e exp;
title 'Crude Prevalence ratio by occupation';
run;
proc genmod data=contactsdataset;
where outcome in ('1','2');
class education (ref='3');
model outcome (event='1')=education/dist=bin link=log;
estimate 'No educ 1 versus 3' education 1 0 -1/e exp;
estimate 'Primary 2 versus 3' education 0 1 -1/e exp;
title 'Crude Prevalence ratio by education';
run;ods rtf close;

proc genmod data=contactsdataset;
where census_cat in (1,2);
class Index_contact (ref='2');
model census_cat (event='1')=Index_contact/dist=bin link=log;
estimate 'PR Cases versus controls' Index_contact 1 -1/e exp;
title 'Crude Prevalence ratio by contact status'run;
proc genmod data=contactsdataset;
where census_cat in (1,2);
class Enroll_sex (ref='2');
model census_cat (event='1')=Enroll_sex/dist=bin link=log;
estimate 'PR Male versus female' Enroll_sex 1 -1/e exp;
title 'Crude Prevalence ratio by Enroll_sex'
run;
proc genmod data=contactsdataset;
where census_cat in (1,2);
class AgeCategory (ref='1');
model census_cat (event='1')=AgeCategory/dist=bin link=log;
estimate 'PR Age cat 2 versus 1' AgeCategory 1 0 0 0 0 -1/e exp;
estimate 'PR Age cat 3 versus 1' AgeCategory 0 1 0 0 0 -1/e exp;
estimate 'PR Age cat 4 versus 1' AgeCategory 0 0 1 0 0 -1/e exp;
estimate 'PR Age cat 5 versus 1' AgeCategory 0 0 0 1 0 -1/e exp;
estimate 'PR Age cat 6 versus 1' AgeCategory 0 0 0 0 1 -1/e exp;
title 'Crude Prevalence ratio by AgeCategory';
run;
proc genmod data=contactsdataset;
where census_cat in (1,2);
class Hivstatus (ref='2');
model census_cat (event='1')=Hivstatus/dist=bin link=log;
estimate 'PR positives 1 versus 2' Hivstatus 1 0 -1/e exp;
estimate 'PR Refusals 3 versus 2' Hivstatus 0 1 -1/e exp;
estimate 'Other 4 versus 2' Hivstatus 0 0 1 -1/e exp;
title 'Crude Prevalence ratio by Hiv status';
run;
proc genmod data=contactsdataset;
where census_cat in (1,2);
class Educationlevel (ref='3');
model census_cat (event='1')=Educationlevel/dist=bin link=log;
estimate 'PR No education 1 versus 3' Educationlevel 1 0 -1/e exp;
estimate 'PR Primary 2 versus 3' Educationlevel 0 1 -1/e exp;
title 'Crude Prevalence ratio by Education Level';
run;
proc genmod data=contactsdataset;
```

```
where census_cat in (1,2);
class sym_drink(ref='2');
model census_cat (event='1')=sym_drink/dist=bin link=log;
estimate 'PR alcohol use versus No alcohol' sym_drink 1 -1/e exp;
title 'Crude Prevalence ratio by drinking status'run;
proc genmod data=contactsdataset;
where census_cat in (1,2);
class sym_habit(ref='2');
model census_cat (event='1')=sym_habit/dist=bin link=log;
estimate 'PR Smokers versus Non smoker' sym_habit 1 0 -1/e exp;
estimate 'PR Smokers versus Past smoker' sym_habit 0 1 -1/e exp;
title 'Crude Prevalence ratio by Smoking status'run;
proc genmod data=contactsdataset;
where census_cat in (1,2);
class BCG (ref='1');
model census_cat (event='1')=BCG/dist=bin link=log;
estimate 'Not vaccinated 2 versus 1' BCG 1 0 -1/e exp;
estimate 'Dont know 3 versus 1' BCG 0 1 -1/e exp;
title 'Crude Prevalence ratio by BCG status';
run;
proc genmod data=contactsdataset;
where census_cat in (1,2);
class sym_Tbever (ref='2');
model census_cat (event='1')=sym_Tbever/dist=bin link=log;
estimate 'Ever lived with TB case 1 versus 2' sym_Tbever 1 0 -1/e exp;
estimate 'Dont remember living with case 77 versus 2' sym_Tbever 0 1 -1/e
exp;
title 'Crude Prevalence ratio by ever lived with TB';
run;
proc genmod data=contactsdataset;
where census_cat in (1,2);
class sym_tbhome(ref='2');
model census cat (event='1')=sym tbhome/dist=bin link=log;
estimate 'PR In home versus None in home' sym_tbhome 1 -1/e exp;
title 'Crude Prevalence ratio by sym_tbhome'run;
proc genmod data=contactsdataset;
where census_cat in (1,2);
class sym_knewtb(ref='2');
model census cat (event='1')=sym knewtb/dist=bin link=log;
estimate 'PR Knew case versus Never knewCase' sym_knewtb 1 -1/e exp;
title 'Crude Prevalence ratio by sym_knewtb'run;
proc genmod data=contactsdataset;
where census_cat in (1,2);
class occupation (ref='3');
model census_cat (event='1')=occupation/dist=bin link=log;
estimate 'indoors 1 versus 3' occupation 1 0 -1/e exp;
estimate 'semindoors 2 versus 3' occupation 0 1 -1/e exp;
title 'Crude Prevalence ratio by occupation';
run;

ods rtf file='P:\MANUSCRIPTS DATASETS\LTBI_Robert\Dissertation
datasets\analysis2\Tables\Final multivariate model.rtf';
*FULL MODEL;
*WHO Age category;
proc genmod data=contactsdataset;
where outcome in ('1','2');
```

```
class census_cat (ref='2') Index_contact (ref='2')Enroll_sex (ref='2')
AgeCategory (ref='1') Hivstatus
(ref='2')sym_knewtb(ref='2')sym_habit(ref='2')sym_drink(ref='2')BCG
(ref='1')sym_Tbever (ref='2')occupation (ref='3'); *Include variables that
are categorical.;
model outcome (event='1')= Index_contact census_cat Enroll_sex AgeCategory
Hivstatus sym_knewtb sym_habit sym_drink BCG sym_Tbever occupation/dist=bin
link=log; *variables that in your previous analysis were statistically
significnat, or any variable that you think should be part of your final
model;
estimate 'PR Household VS Non-household' census_cat 1 -1/e exp;
estimate 'PR Male VS Female' Enroll_sex 1 -1/e exp;
estimate 'PR Age 6-15 VS 0-5' AgeCategory 1 0 0 0 0 -1/e exp;
estimate 'PR Age 16-25 VS 0-5' AgeCategory 0 1 0 0 0 -1/e exp;
estimate 'PR Age 26-35 VS 0-5' AgeCategory 0 0 1 0 0 -1/e exp;
estimate 'PR Age 36-45 VS 0-5' AgeCategory 0 0 0 1 0 -1/e exp;
estimate 'PR Age 45 above VS 0-5' AgeCategory 0 0 0 0 1 -1/e exp;
estimate 'PR Ever Known a TBcase versus Never known' sym_knewtb 1 -1/e exp;
estimate 'PR HIV positives VS HIV negatives' Hivstatus 1 0 -1/e exp;
estimate 'PR Refusals VS HIV negtives' Hivstatus 0 1 -1/e exp;
estimate 'PR CaseContacts VS Controlcontacts' Index_contact 1 -1/e exp;
estimate 'PR Current smoker versus No smoking' sym_habit 1 0 -1/e exp;
estimate 'PR Former smoker versus No smoking' sym_habit 0 1 -1/e exp;
estimate 'PR Alcohol use versus No alcohol use' sym_drink 1 -1/e exp;
estimate 'PR  No BCG VS BCG' BCG 1 0 -1/e exp;
estimate 'PR Dont know VS BCG' BCG 0 1 -1/e exp;
estimate 'PR Ever lived in home with TB case VS Never' sym_Tbever 1 0 -1/e
exp;
estimate 'PR Uncertain lived in home with TB case VS Never' sym_Tbever 0 1 -
1/e exp;
estimate 'PR Works indoors VS works outdoors' occupation 1 0 -1/e exp;
estimate 'PR Works semindoors VS works outdoors' occupation 0 1 -1/e exp;
title 'FULL MODEL, EXPOSURE, CONTACT STATUS, HOUSEHOLD STATUS AGE SEX
AGECATEGORY HIVSTATUS ALCOHOL USE SMOKINGSTATUS BCGSTATUS OCCUPATION  LIVED
WITH CASE EVER KNOWN A CASE WORKLOCATION';
run;ods rtf close;

*Model with Age as a continuous variable;
proc genmod data=contactsdataset;
where outcome in ('1','2');
class census_cat (ref='2') Index_contact (ref='2')Enroll_sex (ref='2')
Hivstatus (ref='2')sym_knewtb(ref='2')sym_habit(ref='2')sym_drink(ref='2')BCG
(ref='1')sym_Tbever (ref='2')occupation (ref='3'); *Include variables that
are categorical.;
model outcome (event='1')= Index_contact census_cat Enroll_sex  Hivstatus
sym_knewtb sym_habit sym_drink BCG sym_Tbever occupation enroll_age /dist=bin
link=log; *variables that in your previous analysis were statistically
significnat, or any variable that you think should be part of your final
model;
estimate 'PR hh versus nhh' census_cat 1 -1/e exp;
estimate 'PR MALE VERSUS FEMALE' Enroll_sex 1 -1/e exp;
estimate 'PR Knew case versus Never knewCase' sym_knewtb 1 -1/e exp;
estimate 'PR positives 1 versus 2' Hivstatus 1 0 -1/e exp;
estimate 'PR Refusals 3 versus 2' Hivstatus 0 1 -1/e exp;
estimate 'PR Cases versus controls' Index_contact 1 -1/e exp;
estimate 'PR smoker versus No alcohol' sym_habit 1 -1/e exp;
```

```
estimate 'PR alcohol use versus No alcohol' sym_drink 1 -1/e exp;
estimate 'Not vaccinated 2 versus 1' BCG 1 0 -1/e exp;
estimate 'Dont know 3 versus 1' BCG 0 1 -1/e exp;
estimate 'Ever lived with TB case 1 versus 2' sym_Tbever 1 0 -1/e exp;
estimate 'Dont remember living with case 77 versus 2' sym_Tbever 0 1 -1/e
exp;
estimate 'indoors 1 versus 3' occupation 1 0 -1/e exp;
estimate 'semindoors 2 versus 3' occupation 0 1 -1/e exp;
estimate 'enroll_age' enroll_age 1 -1/e exp;
title 'FULL MODEL, EXPOSURE, AGE AND SEX';
run;
*Age as a binary variable: UGA age categorisation;
proc genmod data=contactsdataset;
where outcome in ('1','2');
class census_cat (ref='2') Index_contact (ref='2')Enroll_sex (ref='2')
Hivstatus (ref='2')sym_knewtb(ref='2')sym_habit(ref='2')sym_drink(ref='2')BCG
(ref='1')sym_Tbever (ref='2')occupation (ref='3')AgeBinary (ref='1');
*Include variables that are categorical.;
model outcome (event='1')= Index_contact census_cat Enroll_sex  Hivstatus
sym_knewtb sym_habit sym_drink BCG sym_Tbever occupation AgeBinary /dist=bin
link=log; *variables that in your previous analysis were statistically
significnat, or any variable that you think should be part of your final
model;
estimate 'PR hh versus nhh' census_cat 1 -1/e exp;
estimate 'PR MALE VERSUS FEMALE' Enroll_sex 1 -1/e exp;
estimate 'PR Knew case versus Never knewCase' sym_knewtb 1 -1/e exp;
estimate 'PR positives 1 versus 2' Hivstatus 1 0 -1/e exp;
estimate 'PR Refusals 3 versus 2' Hivstatus 0 1 -1/e exp;
estimate 'PR Cases versus controls' Index_contact 1 -1/e exp;
estimate 'PR alcohol use versus No alcohol' sym_habit 1 -1/e exp;
estimate 'PR alcohol use versus No alcohol' sym_drink 1 -1/e exp;
estimate 'Not vaccinated 2 versus 1' BCG 1 0 -1/e exp;
estimate 'Dont know 3 versus 1' BCG 0 1 -1/e exp;
estimate 'Ever lived with TB case 1 versus 2' sym_Tbever 1 0 -1/e exp;
estimate 'Dont remember living with case 77 versus 2' sym_Tbever 0 1 -1/e
exp;
estimate 'indoors 1 versus 3' occupation 1 0 -1/e exp;
estimate 'semindoors 2 versus 3' occupation 0 1 -1/e exp;
estimate 'AgeBinary 2 verus 1' AgeBinary 1 -1/e exp;
title 'FULL MODEL, EXPOSURE, AGE AND SEX';
run;
*MOdels with Intercation terms;
proc genmod data=contactsdataset;
where outcome in ('1','2');
class census_cat (ref='2') Index_contact (ref='2'); *Included hh and contact
status;
model outcome (event='1')= Index_contact*census_cat /dist=bin link=log;
*variables that in your previous analysis were statistically significnat, or
any variable that you think should be part of your final model;
estimate 'PR hh versus nhh' census_cat 1 -1/e exp;
estimate 'PR Cases versus controls' Index_contact 1 -1/e exp;
title 'Intercation MODEL Contact status*HHstatus';
run;

/*After Feb 19th Morning discussion with Chris*/
/*Final model One with Household Status, Contact status Age and Sex*/
```

```
ods rtf file='P:\MANUSCRIPTS DATASETS\LTBI_Robert\Dissertation
datasets\analysis2\Tables\Analysis.rtf';
proc genmod data=contactsdataset;
where outcome in ('1','2');
class census_cat (ref='2') Index_contact (ref='2')Enroll_sex (ref='2');
model outcome (event='1')= Index_contact census_cat Enroll_sex enroll_age
/dist=bin link=log; *variables that in your previous analysis were
statistically significnat*;
estimate 'PR hh versus nhh' census_cat 1 -1/e exp;
estimate 'PR MALE VERSUS FEMALE' Enroll_sex 1 -1/e exp;
estimate 'PR Cases versus controls' Index_contact 1 -1/e exp;
estimate 'enroll_age' enroll_age 1 -1/e exp;
title 'FULL MODEL, Household Status, Contact status Age and Sex';
run;ods rtf close;
*Model two with Household Status, Contact status Age and Sex with interaction
term HHold*ContactStatus;
proc genmod data=contactsdataset;
where outcome in ('1','2');
class census_cat (ref='2') Index_contact (ref='2')Enroll_sex (ref='2');
model outcome (event='1')= Index_contact census_cat Enroll_sex enroll_age
Index_contact*census_cat/dist=bin link=log; *variables that in your previous
analysis were statistically significnat, or any variable that you think
should be part of your final model;
estimate 'PR hh versus nhh' census_cat 1 -1/e exp;
estimate 'PR MALE VERSUS FEMALE' Enroll_sex 1 -1/e exp;
estimate 'PR Cases versus controls' Index_contact 1 -1/e exp;
estimate 'enroll_age' enroll_age 1 -1/e exp;
title 'FULL MODEL, Household Status, Contact status Age and Sex and
Interaction Contacts Status* HHold status ';
run;
*Model three with Household Status, Contact status Age and Sex with
interaction term Age*HHold;
proc genmod data=contactsdataset;
where outcome in ('1','2');
class census_cat (ref='2') Index_contact (ref='2')Enroll_sex (ref='2');
model outcome (event='1')= Index_contact census_cat Enroll_sex enroll_age
Enroll_age*census_cat/dist=bin link=log; *variables that in your previous
analysis were statistically significnat, or any variable that you think
should be part of your final model;
estimate 'PR hh versus nhh' census_cat 1 -1/e exp;
estimate 'PR MALE VERSUS FEMALE' Enroll_sex 1 -1/e exp;
estimate 'PR Cases versus controls' Index_contact 1 -1/e exp;
estimate 'enroll_age' enroll_age 1 -1/e exp;
title 'FULL MODEL, Household Status, Contact status Age and Sex and
Interaction age* HHold';
run;
ods rtf file='P:\MANUSCRIPTS DATASETS\LTBI_Robert\Dissertation
datasets\analysis2\Tables\Analysis.rtf';
*Model three with Household Status, Contact status Age and Sex with
interaction terms  Age*HHold and HHold*ContactStatus;
proc genmod data=contactsdataset;
where outcome in ('1','2');
class census_cat (ref='2') Index_contact (ref='2')Enroll_sex (ref='2');
model outcome (event='1')= Index_contact census_cat Enroll_sex enroll_age
Enroll_age*census_cat Index_contact*census_cat/dist=bin link=log; *variables
```

```
that in your previous analysis were statistically significnat, or any
variable that you think should be part of your final model;
estimate 'PR hh versus nhh' census_cat 1 -1/e exp;
estimate 'PR MALE VERSUS FEMALE' Enroll_sex 1 -1/e exp;
estimate 'PR Cases versus controls' Index_contact 1 -1/e exp;
estimate 'enroll_age' enroll_age 1 -1/e exp;
title 'FULL MODEL, Household Status, Contact status Age and Sex AND BOTH
INTERACTION TERMS';
run;ods rtf close;

/*Model One with Household Status, Contact status Age Sex  and HIVSTATUS*/
ods rtf file='P:\MANUSCRIPTS DATASETS\LTBI_Robert\Dissertation
datasets\analysis2\Tables\Analysis.rtf';
proc genmod data=contactsdataset;
where outcome in ('1','2');
class census_cat (ref='2') Index_contact (ref='2')Enroll_sex
(ref='2')Hivstatus(ref='2');
model outcome (event='1')= Index_contact census_cat Enroll_sex enroll_age
Hivstatus /dist=bin link=log; *variables that in your previous analysis were
statistically significnat*;
estimate 'PR hh versus nhh' census_cat 1 -1/e exp;
estimate 'PR MALE VERSUS FEMALE' Enroll_sex 1 -1/e exp;
estimate 'PR Cases versus controls' Index_contact 1 -1/e exp;
estimate 'PR positives Vs Negatives' Hivstatus 1 0 -1/e exp;
estimate 'PR Refusals Vs negatives' Hivstatus 0 1 -1/e exp;
estimate 'enroll_age' enroll_age 1 -1/e exp;
title 'FULL MODEL, Household Status, Contact status Age and Sex';run;
*Model two with Household Status, Contact status Age Sex  and HIVSTATUS with
interaction term HHold*ContactStatus;
proc genmod data=contactsdataset;
where outcome in ('1','2');
class census_cat (ref='2') Index_contact (ref='2')Enroll_sex (ref='2')
Hivstatus (ref='2');
model outcome (event='1')= Index_contact census_cat Enroll_sex enroll_age
Index_contact*census_cat Hivstatus/dist=bin link=log; *variables that in your
previous analysis were statistically significnat, or any variable that you
think should be part of your final model;
estimate 'PR hh versus nhh' census_cat 1 -1/e exp;
estimate 'PR MALE VERSUS FEMALE' Enroll_sex 1 -1/e exp;
estimate 'PR Cases versus controls' Index_contact 1 -1/e exp;
estimate 'PR positives Vs Negatives' Hivstatus 1 0 -1/e exp;
estimate 'PR Refusals Vs negatives' Hivstatus 0 1 -1/e exp;
estimate 'enroll_age' enroll_age 1 -1/e exp;
title 'FULL MODEL, Household Status, Contact status Age and Sex and
Interaction Contacts Status* HHold status ';run;
*Model three with Household Status, Contact status Age Sex  and HIVSTATUS
with interaction term Age*HHold;
proc genmod data=contactsdataset;
where outcome in ('1','2');
class census_cat (ref='2') Index_contact (ref='2')Enroll_sex (ref='2')
Hivstatus (ref='2');
model outcome (event='1')= Index_contact census_cat Enroll_sex enroll_age
Enroll_age*census_cat Hivstatus/dist=bin link=log; *variables that in your
previous analysis were statistically significnat, or any variable that you
think should be part of your final model;
estimate 'PR hh versus nhh' census_cat 1 -1/e exp;
```

```
estimate 'PR MALE VERSUS FEMALE' Enroll_sex 1 -1/e exp;
estimate 'PR Cases versus controls' Index_contact 1 -1/e exp;
estimate 'PR positives Vs Negatives' Hivstatus 1 0 -1/e exp;
estimate 'PR Refusals Vs negatives' Hivstatus 0 1 -1/e exp;
estimate 'enroll_age' enroll_age 1 -1/e exp;
title 'FULL MODEL, Household Status, Contact status Age and Sex and
Interaction age* HHold';run;
*Model three with Household Status, Contact status Age Sex  and HIVSTATUS
with interaction terms  Age*HHold and HHold*ContactStatus;
proc genmod data=contactsdataset;
where outcome in ('1','2');
class census_cat (ref='2') Index_contact (ref='2')Enroll_sex (ref='2')
Hivstatus (ref='2');
model outcome (event='1')= Index_contact census_cat Enroll_sex enroll_age
Enroll_age*census_cat Index_contact*census_cat Hivstatus/dist=bin link=log;
*variables that in your previous analysis were statistically significnat, or
any variable that you think should be part of your final model;
estimate 'PR hh versus nhh' census_cat 1 -1/e exp;
estimate 'PR MALE VERSUS FEMALE' Enroll_sex 1 -1/e exp;
estimate 'PR Cases versus controls' Index_contact 1 -1/e exp;
estimate 'PR positives Vs Negatives' Hivstatus 1 0 -1/e exp;
estimate 'PR Refusals Vs negatives' Hivstatus 0 1 -1/e exp;
estimate 'enroll_age' enroll_age 1 -1/e exp;
title 'FULL MODEL, Household Status, Contact status Age and Sex AND BOTH
INTERACTION TERMS';
run;ods rtf close;

*Model One with Household Status, Contact status Age Sex  HIVSTATUS AND BCG*/
ods rtf file='P:\MANUSCRIPTS DATASETS\LTBI_Robert\Dissertation
datasets\analysis2\Tables\Analysis.rtf';
proc genmod data=contactsdataset;
where outcome in ('1','2');
class census_cat (ref='2') Index_contact (ref='2')Enroll_sex
(ref='2')Hivstatus (ref='2') BCG (ref='1');
model outcome (event='1')= Index_contact census_cat Enroll_sex enroll_age
Hivstatus BCG /dist=bin link=log; *variables that in your previous analysis
were statistically significnat*;
estimate 'PR hh versus nhh' census_cat 1 -1/e exp;
estimate 'PR MALE VERSUS FEMALE' Enroll_sex 1 -1/e exp;
estimate 'PR Cases versus controls' Index_contact 1 -1/e exp;
estimate 'PR positives Vs Negatives' Hivstatus 1 0 -1/e exp;
estimate 'PR Refusals Vs negatives' Hivstatus 0 1 -1/e exp;
estimate 'enroll_age' enroll_age 1 -1/e exp;
estimate 'PR  No BCG VS BCG' BCG 1 0 -1/e exp;
estimate 'PR Dont know VS BCG' BCG 0 1 -1/e exp;
title 'FULL MODEL, Household Status, Contact status Age and Sex';run;
*Model two with Household Status, Contact status Age Sex  and HIVSTATUS with
interaction term HHold*ContactStatus;
proc genmod data=contactsdataset;
where outcome in ('1','2');
class census_cat (ref='2') Index_contact (ref='2')Enroll_sex (ref='2')
Hivstatus (ref='2') BCG (ref='1');
model outcome (event='1')= Index_contact census_cat Enroll_sex enroll_age
Index_contact*census_cat Hivstatus BCG/dist=bin link=log; *variables that in
your previous analysis were statistically significnat, or any variable that
you think should be part of your final model;
```

```
estimate 'PR hh versus nhh' census_cat 1 -1/e exp;
estimate 'PR MALE VERSUS FEMALE' Enroll_sex 1 -1/e exp;
estimate 'PR Cases versus controls' Index_contact 1 -1/e exp;
estimate 'PR positives Vs Negatives' Hivstatus 1 0 -1/e exp;
estimate 'PR Refusals Vs negatives' Hivstatus 0 1 -1/e exp;
estimate 'enroll_age' enroll_age 1 -1/e exp;
estimate 'PR  No BCG VS BCG' BCG 1 0 -1/e exp;
estimate 'PR Dont know VS BCG' BCG 0 1 -1/e exp;
title 'FULL MODEL, Household Status, Contact status Age and Sex and
Interaction Contacts Status* HHold status ';run;
*Model three with Household Status, Contact status Age Sex  and HIVSTATUS
with interaction term Age*HHold;
proc genmod data=contactsdataset;
where outcome in ('1','2');
class census_cat (ref='2') Index_contact (ref='2')Enroll_sex (ref='2')
Hivstatus (ref='2') BCG (ref='1');
model outcome (event='1')= Index_contact census_cat Enroll_sex enroll_age
Enroll_age*census_cat Hivstatus BCG/dist=bin link=log; *variables that in
your previous analysis were statistically significnat, or any variable that
you think should be part of your final model;
estimate 'PR hh versus nhh' census_cat 1 -1/e exp;
estimate 'PR MALE VERSUS FEMALE' Enroll_sex 1 -1/e exp;
estimate 'PR Cases versus controls' Index_contact 1 -1/e exp;
estimate 'PR positives Vs Negatives' Hivstatus 1 0 -1/e exp;
estimate 'PR Refusals Vs negatives' Hivstatus 0 1 -1/e exp;
estimate 'enroll_age' enroll_age 1 -1/e exp;
estimate 'PR  No BCG VS BCG' BCG 1 0 -1/e exp;
estimate 'PR Dont know VS BCG' BCG 0 1 -1/e exp;
title 'FULL MODEL, Household Status, Contact status Age and Sex and
Interaction age* HHold';run;
ods rtf file='P:\MANUSCRIPTS DATASETS\LTBI_Robert\Dissertation
datasets\analysis2\Tables\Analysis.rtf';
*Model three with Household Status, Contact status Age Sex  and HIVSTATUS
with interaction terms  Age*HHold and HHold*ContactStatus;
proc genmod data=contactsdataset;
where outcome in ('1','2');
class census_cat (ref='2') Index_contact (ref='2')Enroll_sex (ref='2')
Hivstatus (ref='2')BCG (ref='1');
model outcome (event='1')= Index_contact census_cat Enroll_sex enroll_age
Enroll_age*census_cat Index_contact*census_cat Hivstatus BCG/dist=bin
link=log; *variables that in your previous analysis were statistically
significnat, or any variable that you think should be part of your final
model;
estimate 'PR hh versus nhh' census_cat 1 -1/e exp;
estimate 'PR MALE VERSUS FEMALE' Enroll_sex 1 -1/e exp;
estimate 'PR Cases versus controls' Index_contact 1 -1/e exp;
estimate 'PR positives Vs Negatives' Hivstatus 1 0 -1/e exp;
estimate 'PR Refusals Vs negatives' Hivstatus 0 1 -1/e exp;
estimate 'enroll_age' enroll_age 1 -1/e exp;
estimate 'PR  No BCG VS BCG' BCG 1 0 -1/e exp;
estimate 'PR Dont know VS BCG' BCG 0 1 -1/e exp;
title 'FULL MODEL, Household Status, Contact status Age and Sex AND BOTH
INTERACTION TERMS';
run;ods rtf close;

/*Model One with Household Status, Contact status Age Sex  and HIV(Binary)*/
```

```
ods rtf file='P:\MANUSCRIPTS DATASETS\LTBI_Robert\Dissertation
datasets\analysis2\Tables\HIVBinaryFinal.rtf';
proc genmod data=contactsdataset;
where outcome in ('1','2');
class census_cat (ref='2') Index_contact (ref='2')Enroll_sex
(ref='2')Hiv(ref='2');
model outcome (event='1')= Index_contact census_cat Enroll_sex enroll_age Hiv
/dist=bin link=log; *variables that in your previous analysis were
statistically significnat*;
estimate 'PR hh versus nhh' census_cat 1 -1/e exp;
estimate 'PR MALE VERSUS FEMALE' Enroll_sex 1 -1/e exp;
estimate 'PR Cases versus controls' Index_contact 1 -1/e exp;
estimate 'PR positives Vs Negatives' Hiv 1 -1/e exp;
estimate 'enroll_age' enroll_age 1 -1/e exp;
title 'FULL MODEL, Household Status, Contact status Age and Sex and
HIV(Binary) with Intercation terms';run;
*Model two with Household Status, Contact status Age Sex  and HIVSTATUS with
interaction term HHold*ContactStatus;
proc genmod data=contactsdataset;
where outcome in ('1','2');
class census_cat (ref='2') Index_contact (ref='2')Enroll_sex (ref='2') Hiv
(ref='2');
model outcome (event='1')= Index_contact census_cat Enroll_sex enroll_age
Index_contact*census_cat Hiv/dist=bin link=log; *variables that in your
previous analysis were statistically significnat, or any variable that you
think should be part of your final model;
estimate 'PR hh versus nhh' census_cat 1 -1/e exp;
estimate 'PR MALE VERSUS FEMALE' Enroll_sex 1 -1/e exp;
estimate 'PR Cases versus controls' Index_contact 1 -1/e exp;
estimate 'PR positives Vs Negatives' Hiv 1 -1/e exp;
estimate 'enroll_age' enroll_age 1 -1/e exp;
estimate 'Index_contact*census_cat' Index_contact*census_cat 1 -1/e exp;
title 'FULL MODEL, Household Status, Contact status Age and Sex and
Interaction Contacts Status* HHold status ';run;
*Model two with Household Status, Contact status Age Sex  and HIVSTATUS with
interaction term HHold*age;
proc genmod data=contactsdataset;
where outcome in ('1','2');
class census_cat (ref='2') Index_contact (ref='2')Enroll_sex (ref='2') Hiv
(ref='2');
model outcome (event='1')= Index_contact census_cat Enroll_sex enroll_age
Enroll_age*census_cat Hiv/dist=bin link=log; *variables that in your previous
analysis were statistically significnat, or any variable that you think
should be part of your final model;
estimate 'PR hh versus nhh' census_cat 1 -1/e exp;
estimate 'PR MALE VERSUS FEMALE' Enroll_sex 1 -1/e exp;
estimate 'PR Cases versus controls' Index_contact 1 -1/e exp;
estimate 'PR positives Vs Negatives' Hiv 1 -1/e exp;
estimate 'enroll_age' enroll_age 1 -1/e exp;
title 'FULL MODEL, Household Status, Contact status Age and Sex and
Interaction Contacts Status* HHold status ';run;
*Model two with Household Status, Contact status Age Sex  and HIVSTATUS with
interaction term HHold*ContactStatus AND HHold*age;
proc genmod data=contactsdataset;
where outcome in ('1','2');
```

```
class census_cat (ref='2') Index_contact (ref='2')Enroll_sex (ref='2') Hiv
(ref='2');
model outcome (event='1')= Index_contact census_cat Enroll_sex enroll_age
Enroll_age*census_cat  Index_contact*census_cat Hiv/dist=bin link=log;
*variables that in your previous analysis were statistically significnat, or
any variable that you think should be part of your final model;
estimate 'PR hh versus nhh' census_cat 1 -1/e exp;
estimate 'PR MALE VERSUS FEMALE' Enroll_sex 1 -1/e exp;
estimate 'PR Cases versus controls' Index_contact 1 -1/e exp;
estimate 'PR positives Vs Negatives' Hiv 1 -1/e exp;
estimate 'enroll_age' enroll_age 1 -1/e exp;
title 'FULL MODEL, Household Status, Contact status Age and Sex and
Interaction Contacts Status* HHold status';run;run;ods rtf close;

*Model two with Household Status, Contact status Agecategory Sex  and
HIVSTATUS with interaction term HHold*ContactStatus AND HHold*agecategory;
proc genmod data=contactsdataset;
where outcome in ('1','2');
class census_cat (ref='2') Index_contact (ref='2')Enroll_sex (ref='2') Hiv
(ref='2')agecategory(ref='1');
model outcome (event='1')= Index_contact census_cat Enroll_sex agecategory
agecategory*census_cat Index_contact*census_cat Hiv/dist=bin link=log; *Age
is now categorica*;
estimate 'PR hh versus nhh' census_cat 1 -1/e exp;
estimate 'PR MALE VERSUS FEMALE' Enroll_sex 1 -1/e exp;
estimate 'PR Cases versus controls' Index_contact 1 -1/e exp;
estimate 'PR positives Vs Negatives' Hiv 1 -1/e exp;
estimate 'PR Age 6-15 VS 0-5' AgeCategory 1 0 0 0 0 -1/e exp;
estimate 'PR Age 16-25 VS 0-5' AgeCategory 0 1 0 0 0 -1/e exp;
estimate 'PR Age 26-35 VS 0-5' AgeCategory 0 0 1 0 0 -1/e exp;
estimate 'PR Age 36-45 VS 0-5' AgeCategory 0 0 0 1 0 -1/e exp;
estimate 'PR Age 45 above VS 0-5' AgeCategory 0 0 0 0 1 -1/e exp;
title 'FULL MODEL, Household Status, Contact status Agecategory and Sex and
Interaction Contacts Status* HHold status';run;

/*Age specific Prevalence by household status among case contacts (Table
4a)*/
proc freq data=Casecontacts;tables census_cat*outcome*agecategory/cmh;run;
/*Age specific Prevalence by household status among control contacts (table
4b*/
proc freq data=Controlcontacts;tables census_cat*outcome*agecategory/cmh;run;

/*Final model One with Household Status, Contact status Age Sex  and
HIV(Binary)* FOR CASE COTACT ONLY*/
ods rtf file='P:\MANUSCRIPTS DATASETS\LTBI_Robert\Dissertation
datasets\analysis2\Tables\HIVBinaryFinal.rtf';
proc genmod data=casecontacts;
where outcome in ('1','2');
class census_cat (ref='2') Enroll_sex (ref='2')Hiv(ref='2');
model outcome (event='1')= census_cat Enroll_sex enroll_age Hiv /dist=bin
link=log; *variables that in your previous analysis were statistically
significnat*;
estimate 'PR hh versus nhh' census_cat 1 -1/e exp;
estimate 'PR MALE VERSUS FEMALE' Enroll_sex 1 -1/e exp;
estimate 'PR positives Vs Negatives' Hiv 1 -1/e exp;
estimate 'enroll_age' enroll_age 1 -1/e exp;
```

```
title 'FULL MODEL, Household Status, Contact status Age and Sex and
HIV(Binary) with Intercation terms';run;
*Final model two with Household Status, Contact status Age Sex  and HIVSTATUS
with interaction term HHold*ContactStatus;
proc genmod data=contactsdataset;
where outcome in ('1','2');
class census_cat (ref='2') Index_contact (ref='2')Enroll_sex (ref='2') Hiv
(ref='2');
model outcome (event='1')= Index_contact census_cat Enroll_sex enroll_age
Index_contact*census_cat Hiv/dist=bin link=log; *variables that in your
previous analysis were statistically significnat, or any variable that you
think should be part of your final model;
estimate 'PR hh versus nhh' census_cat 1 -1/e exp;
estimate 'PR MALE VERSUS FEMALE' Enroll_sex 1 -1/e exp;
estimate 'PR Cases versus controls' Index_contact 1 -1/e exp;
estimate 'PR positives Vs Negatives' Hiv 1 -1/e exp;
estimate 'enroll_age' enroll_age 1 -1/e exp;
estimate 'Index_contact*census_cat' Index_contact*census_cat 1 -1/e exp;
title 'FULL MODEL, Household Status, Contact status Age and Sex and
Interaction Contacts Status* HHold status ';run;
*Final model two with Household Status, Contact status Age Sex  and HIVSTATUS
with interaction term HHold*age;
proc genmod data=contactsdataset;
where outcome in ('1','2');
class census_cat (ref='2') Index_contact (ref='2')Enroll_sex (ref='2') Hiv
(ref='2');
model outcome (event='1')= Index_contact census_cat Enroll_sex enroll_age
Enroll_age*census_cat Hiv/dist=bin link=log; *variables that in your previous
analysis were statistically significnat, or any variable that you think
should be part of your final model;
estimate 'PR hh versus nhh' census_cat 1 -1/e exp;
estimate 'PR MALE VERSUS FEMALE' Enroll_sex 1 -1/e exp;
estimate 'PR Cases versus controls' Index_contact 1 -1/e exp;
estimate 'PR positives Vs Negatives' Hiv 1 -1/e exp;
estimate 'enroll_age' enroll_age 1 -1/e exp;
title 'FULL MODEL, Household Status, Contact status Age and Sex and
Interaction Contacts Status* HHold status ';run;
*Final model two with Household Status, Contact status Age Sex  and HIVSTATUS
with interaction term HHold*ContactStatus AND HHold*age;
proc genmod data=contactsdataset;
where outcome in ('1','2');
class census_cat (ref='2') Index_contact (ref='2')Enroll_sex (ref='2') Hiv
(ref='2');
model outcome (event='1')= Index_contact census_cat Enroll_sex enroll_age
Enroll_age*census_cat  Index_contact*census_cat Hiv/dist=bin link=log;
*variables that in your previous analysis were statistically significnat, or
any variable that you think should be part of your final model;
estimate 'PR hh versus nhh' census_cat 1 -1/e exp;
estimate 'PR MALE VERSUS FEMALE' Enroll_sex 1 -1/e exp;
estimate 'PR Cases versus controls' Index_contact 1 -1/e exp;
estimate 'PR positives Vs Negatives' Hiv 1 -1/e exp;
estimate 'enroll_age' enroll_age 1 -1/e exp;
title 'FULL MODEL, Household Status, Contact status Age and Sex and
Interaction Contacts Status* HHold status';run;run;ods rtf close;
*Final model two with Household Status, Contact status Agecategory Sex  and
HIVSTATUS with interaction term HHold*agecategory;
```

```
proc genmod data=casecontacts;
where outcome in ('1','2');
class census_cat (ref='2') Enroll_sex (ref='2') Hiv
(ref='2')agecategory(ref='1');
model outcome (event='1')= census_cat Enroll_sex agecategory
agecategory*census_cat Hiv/dist=bin link=log; *Age is now categorica*;
estimate 'PR hh versus nhh' census_cat 1 -1/e exp;
estimate 'PR MALE VERSUS FEMALE' Enroll_sex 1 -1/e exp;
estimate 'PR positives Vs Negatives' Hiv 1 -1/e exp;
estimate 'PR Age 6-15 VS 0-5' AgeCategory 1 0 0 0 0 -1/e exp;
estimate 'PR Age 16-25 VS 0-5' AgeCategory 0 1 0 0 0 -1/e exp;
estimate 'PR Age 26-35 VS 0-5' AgeCategory 0 0 1 0 0 -1/e exp;
estimate 'PR Age 36-45 VS 0-5' AgeCategory 0 0 0 1 0 -1/e exp;
estimate 'PR Age 45 above VS 0-5' AgeCategory 0 0 0 0 1 -1/e exp;
title 'FULL MODEL, Household Status, Agecategory and Sex and Interaction
agecategory*HHold status';run;

/*Final model One with Household Status, Contact status Age Sex  and
HIV(Binary)* FOR CONTROL CONTACTS ONLY*/

ods rtf file='P:\MANUSCRIPTS DATASETS\LTBI_Robert\Dissertation
datasets\analysis2\Tables\HIVBinaryFinal.rtf';
proc genmod data=casecontacts;
where outcome in ('1','2');
class census_cat (ref='2') Enroll_sex (ref='2')Hiv(ref='2');
model outcome (event='1')= census_cat Enroll_sex enroll_age Hiv /dist=bin
link=log; *variables that in your previous analysis were statistically
significnat*;
estimate 'PR hh versus nhh' census_cat 1 -1/e exp;
estimate 'PR MALE VERSUS FEMALE' Enroll_sex 1 -1/e exp;
estimate 'PR positives Vs Negatives' Hiv 1 -1/e exp;
estimate 'enroll_age' enroll_age 1 -1/e exp;
title 'FULL MODEL, Household Status, Contact status Age and Sex and
HIV(Binary) with Intercation terms';run;
*Final model two with Household Status, Contact status Age Sex  and HIVSTATUS
with interaction term HHold*ContactStatus;
proc genmod data=contactsdataset;
where outcome in ('1','2');
class census_cat (ref='2') Index_contact (ref='2')Enroll_sex (ref='2') Hiv
(ref='2');
model outcome (event='1')= Index_contact census_cat Enroll_sex enroll_age
Index_contact*census_cat Hiv/dist=bin link=log; *variables that in your
previous analysis were statistically significnat, or any variable that you
think should be part of your final model;
estimate 'PR hh versus nhh' census_cat 1 -1/e exp;
estimate 'PR MALE VERSUS FEMALE' Enroll_sex 1 -1/e exp;
estimate 'PR Cases versus controls' Index_contact 1 -1/e exp;
estimate 'PR positives Vs Negatives' Hiv 1 -1/e exp;
estimate 'enroll_age' enroll_age 1 -1/e exp;
estimate 'Index_contact*census_cat' Index_contact*census_cat 1 -1/e exp;
title 'FULL MODEL, Household Status, Contact status Age and Sex and
Interaction Contacts Status* HHold status ';run;
*Final model two with Household Status, Contact status Age Sex  and HIVSTATUS
with interaction term HHold*age;
proc genmod data=contactsdataset;
where outcome in ('1','2');
```

```
class census_cat (ref='2') Index_contact (ref='2')Enroll_sex (ref='2') Hiv
(ref='2');
model outcome (event='1')= Index_contact census_cat Enroll_sex enroll_age
Enroll_age*census_cat Hiv/dist=bin link=log; *variables that in your previous
analysis were statisticalLy significnat, or any variable that you think
should be part of your final model;
estimate 'PR hh versus nhh' census_cat 1 -1/e exp;
estimate 'PR MALE VERSUS FEMALE' Enroll_sex 1 -1/e exp;
estimate 'PR Cases versus controls' Index_contact 1 -1/e exp;
estimate 'PR positives Vs Negatives' Hiv 1 -1/e exp;
estimate 'enroll_age' enroll_age 1 -1/e exp;
title 'FULL MODEL, Household Status, Contact status Age and Sex and
Interaction Contacts Status* HHold status ';run;
*Final model two with Household Status, Contact status Age Sex  and HIVSTATUS
with interaction term HHold*ContactStatus AND HHold*age;
proc genmod data=contactsdataset;
where outcome in ('1','2');
class census_cat (ref='2') Index_contact (ref='2')Enroll_sex (ref='2') Hiv
(ref='2');
model outcome (event='1')= Index_contact census_cat Enroll_sex enroll_age
Enroll_age*census_cat  Index_contact*census_cat Hiv/dist=bin link=log;
*variables that in your previous analysis were statisticalLy significnat, or
any variable that you think should be part of your final model;
estimate 'PR hh versus nhh' census_cat 1 -1/e exp;
estimate 'PR MALE VERSUS FEMALE' Enroll_sex 1 -1/e exp;
estimate 'PR Cases versus controls' Index_contact 1 -1/e exp;
estimate 'PR positives Vs Negatives' Hiv 1 -1/e exp;
estimate 'enroll_age' enroll_age 1 -1/e exp;
title 'FULL MODEL, Household Status, Contact status Age and Sex and
Interaction Contacts Status* HHold status';run;run;ods rtf close;

*Final model two with Household Status, Contact status Agecategory Sex  and
HIVSTATUS with interaction term HHold*agecategory;
proc genmod data=controlcontacts;
where outcome in ('1','2');
class census_cat (ref='2') Enroll_sex (ref='2') Hiv
(ref='2')agecategory(ref='1');
model outcome (event='1')= census_cat Enroll_sex agecategory
agecategory*census cat Hiv/dist=bin link=log; *Age is now categorica*;
estimate 'PR hh versus nhh' census_cat 1 -1/e exp;
estimate 'PR MALE VERSUS FEMALE' Enroll_sex 1 -1/e exp;
estimate 'PR positives Vs Negatives' Hiv 1 -1/e exp;
estimate 'PR Age 6-15 VS 0-5' AgeCategory 1 0 0 0 0 -1/e exp;
estimate 'PR Age 16-25 VS 0-5' AgeCategory 0 1 0 0 0 -1/e exp;
estimate 'PR Age 26-35 VS 0-5' AgeCategory 0 0 1 0 0 -1/e exp;
estimate 'PR Age 36-45 VS 0-5' AgeCategory 0 0 0 1 0 -1/e exp;
estimate 'PR Age 45 above VS 0-5' AgeCategory 0 0 0 0 1 -1/e exp;
title 'FULL MODEL, Household Status, Agecategory and Sex and Interaction
agecategory*HHold status';run;
```

APPENDIX B

SAS PROC GENMOD'S ANALYSIS CODE FOR AIM TWO

```
Libname robert "W:\MANUSCRIPTS DATASETS\LTBI_Robert\Dissertation
datasets\Analysis2\AIM 2 COHORT DATA\Datasets";

/*Droping unwanted variables*/
data Aim2 (drop= junior primary senior educ othoccup2 tstmean_base tst_meanf3
tst_meanf4 tst_meanf5   TST_base TST_FU   diff tstmean_base tst_meanf1
tst_meanf2  tst_meanf3);set Aim2;run;

/*Creating new variables" */
data Aim2;set Aim2;
MaritalStatus='.';
If married = 2 then MaritalStatus=1;Else if married = 3 then MaritalStatus=1;
Else if married = 1 then MaritalStatus=2;Else if married = 4 then
MaritalStatus=2; Else if married = 5 then MaritalStatus=2;run;
proc freq data=Aim2;table maritalstatus;run;Title 'Marital status';run;

data Aim2;set Aim2;
outcome2=.;
If outcome =1 then outcome2 = 1;
else if outcome > 1 then outcome2 = 2;
else if outcome = . then outcome2 = 2;
run;

proc freq data = Aim2;
tables BLrisk_HOMETB BLrisk_BEDROOMTB BLrisk_TIMESPENT BLrisk_NURSETB ;
run;

data Aim2;set Aim2;
Smokers='.';
If BLrisk_Habit-x=1 then smokers=1;Else if BLrisk_Habit_x=2 then
smokers=1;Else if BLrisk_Habit_x=3 then smokers=2;run;
data Aim2;set Aim2;
Smokers2='.';
If smokers=1 then smokers2=1;Else if smokers=2 then smokers2=2;Else if
smokers='.' then smokers2=2;run;

data Aim2;set Aim2;
TaxiRide='.';
If BLrisk_TAXI_x=0 then taxiride=1;Else if BLrisk_TAXI_x=1 then
taxiride=1;Else if BLrisk_TAXI_x=2 then taxiride=2;Else if BLrisk_TAXI_x=3
then taxiride=2;Else if BLrisk_TAXI_x=4 then taxiride=2;Else if
BLrisk_TAXI_x=5 then taxiride=2;Else if BLrisk_TAXI_x=99 then taxiride=3;run;

   Data Aim2;set Aim2;
EducationLevel='.';
```

```
if education='1' then EducationLevel='1';else if education='2' then
EducationLevel='2'; else if education='3' then EducationLevel='3';else if
education='4' then EducationLevel='3';Else if education ='5' then
EducationLevel='3';run;
proc freq data=Aim2;table EducationLevel;run;
data Aim2;set Aim2;
AgeCategory='.';
If age GE 15 and age LT 25 then AgeCategory=1;Else if age GE 25 and age LT 35
then AgeCategory=2;Else if age GE 35 and age LT 45 then AgeCategory=3;Else if
age GE 45 then AgeCategory=4;run;
proc freq data=Aim2;table AgeCategory;run;Title 'AgeCategory created';run;
data Aim2;set Aim2;
Occupation='.';
if occup1 in (3,4,6,8,14,16,18,19) then Occupation=1; *If you have more than
one value (3,4,6,8...) should have to use in instead of '=';
Else if occup1 in  (5,9,10,11,13,12) then Occupation=2;
else if occup1 in  (1,2,7,17,20,21,22 23 66 88 15) then Occupation=3;run;
proc freq data=Aim2;table Occupation;run;
Data Aim2;set Aim2;
SES='.';
if income='1' then SES='1';else if income='2' then SES='1'; else if
income='3' then SES='2';else if income='4' then SES='3';Else if Income ='5'
then SES='4';run;
proc freq data=Aim2;table SES;run;

data Aim2;set Aim2;
Parish2='.';
If Parish = 1 then parish2=1;Else if parish = 2 then parish2 = 2; else if
parish =3 then parish2 =1;run;

Data Aim2;set Aim2;
Religious='.';
if religion='1' then Religious='1';else if religion='2' then Religious='2';
else if religion='3' then religious='3';else if religion='4' then
Religious='4';Else if religion ='5' then Religious='4';Else if religion ='88'
then Religious='4';run;
proc freq data=Aim2;table Religious;run;
Data Aim2;set Aim2;
HomeTB='.';
if BLrisk_HOMETB='1' then HomeTB='1';else if BLrisk_HOMETB='2' then
HomeTB='2'; else if BLrisk_HOMETB='99' then HomeTB='2';run;
proc freq data=Aim2;table hometb;run;
data Aim2;set Aim2;
Bedroom='.';
If BLrisk_BEDROOMTB = 1 then Bedroom=1;Else if BLrisk_BEDROOMTB GE 2 then
Bedroom=2; if BLrisk_BEDROOMTB = '.' then bedroom=2;run;
data Aim2;set Aim2;
TimeSpent='.';
If BLrisk_TIMESPENT = 1 then timespent=1;Else if BLrisk_TIMESPENT GE 2 then
timespent=2;run;
data Aim2;set Aim2;
NurseTB='.';
If BLrisk_NURSETB = 1 then nursetb=1;Else if BLrisk_NURSETB GE 2 then
nursetb=2;run;
data Aim2;set Aim2;
Parish='.';
```

```
if par in ('2','3','5','7','9','10') then parish=1;
if par in ('1','4','6','8','11','12','13') then parish=2;
if par = '88' then parish=3;run;

data Aim2;set Aim2;
TBHome='.';
If BLrisk_TBhome = 1 then TBhome=1;Else if BLrisk_TBhome GE 2 then
tbhome=2;run;
data Aim2;set Aim2;
StayTB='.';
If BLrisk_StayTB = 1 then StayTB=1;Else if BLrisk_StayTb GE 2 then
StayTB=2;run;

data men;set men;
TBHome='.';
If BLrisk_TBhome = 1 then TBhome=1;Else if BLrisk_TBhome GE 2 then
tbhome=2;run;
data men;set men;
StayTB='.';
If BLrisk_StayTB = 1 then StayTB=1;Else if BLrisk_StayTb GE 2 then
StayTB=2;run;

data women;set women;
TBHome='.';
If BLrisk_TBhome = 1 then TBhome=1;Else if BLrisk_TBhome GE 2 then
tbhome=2;run;
data women;set women;
StayTB='.';
If BLrisk_StayTB = 1 then StayTB=1;Else if BLrisk_StayTb GE 2 then
StayTB=2;run;

data Aim2;set Aim2;
Educationlevelbinary='.';
If Educationlevel = 1 then EducationlevelBinary=1;Else if Educationlevel =2
then EducationlevelBinary=1;Else if Educationlevel=3 then
educationlevelbinary =2;run;
data men;set men;
Educationlevelbinary='.';
If Educationlevel = 1 then EducationlevelBinary=1;Else if Educationlevel =2
then EducationlevelBinary=1;Else if Educationlevel=3 then
educationlevelbinary =2;run;
data women;set women;
Educationlevelbinary='.';
If Educationlevel = 1 then EducationlevelBinary=1;Else if Educationlevel =2
then EducationlevelBinary=1;Else if Educationlevel=3 then
educationlevelbinary =2;run;

data Aim2;set Aim2;
SESrecoded='.';
If SES = 1 then SESrecoded=3;Else if SES= 2 then SESrecoded=3;Else if SES= 3
then SESrecoded=2;else if SES= 4 then SESrecoded=1;run;
data men;set men;
SESrecoded='.';
If SES = 1 then SESrecoded=3;Else if SES= 2 then SESrecoded=3;Else if SES= 3
then SESrecoded=2;else if SES= 4 then SESrecoded=1;run;
data women;set women;
```

```
SESrecoded='.';
If SES = 1 then SESrecoded=3;Else if SES= 2 then SESrecoded=3;Else if SES= 3
then SESrecoded=2;else if SES= 4 then SESrecoded=1;run;

MEN
data men;set men;
MaritalStatus='.';
If married = 2 then MaritalStatus=1;Else if married = 3 then MaritalStatus=1;
Else if married = 1 then MaritalStatus=2;Else if married = 4 then
MaritalStatus=2; Else if married = 5 then MaritalStatus=2;run;
Data men;set men;
EducationLevel='.';
if education='1' then EducationLevel='1';else if education='2' then
EducationLevel='2'; else if education='3' then EducationLevel='3';else if
education='4' then EducationLevel='3';Else if education ='5' then
EducationLevel='3';run;
data men;set men;
AgeCategory='.';
If age GE 15 and age LT 25 then AgeCategory=1;Else if age GE 25 and age LT 35
then AgeCategory=2;Else if age GE 35 and age LT 45 then AgeCategory=3;Else if
age GE 45 then AgeCategory=4;run;
data men;set men;
Occupation='.';
if occup1 in (3,4,6,8,14,16,18,19) then Occupation=1; *If you have more than
one value (3,4,6,8...) should have to use in instead of '=';
Else if occup1 in  (5,9,10,11,13,12) then Occupation=2;
else if occup1 in  (1,2,7,17,20,21,22 23 66 88 15) then Occupation=3;run;
Data men;set men;
SES='.';
if income='1' then SES='1';else if income='2' then SES='1'; else if
income='3' then SES='2';else if income='4' then SES='3';Else if Income ='5'
then SES='4';run;
Data men;set men;
Religious='.';
if religion='1' then Religious='1';else if religion='2' then Religious='2';
else if religion='3' then religious='3';else if religion='4' then
Religious='4';Else if religion ='5' then Religious='4';Else if religion ='88'
then Religious='4';run;
Data men;set men;
HomeTB='.';
if BLrisk_HOMETB='1' then HomeTB='1';else if BLrisk_HOMETB='2' then
HomeTB='2'; else if BLrisk_HOMETB='99' then HomeTB='2';run;
data men;set men;
Bedroom='.';
If BLrisk_BEDROOMTB = 1 then Bedroom=1;Else if BLrisk_BEDROOMTB GE 2 then
Bedroom=2; if BLrisk_BEDROOMTB = '.' then bedroom=2;run;
data Aim2;set Aim2;
TimeSpent='.';
If BLrisk_TIMESPENT = 1 then timespent=1;Else if BLrisk_TIMESPENT GE 2 then
timespent=2;run;
data men;set men;
NurseTB='.';
If BLrisk_NURSETB = 1 then nursetb=1;Else if BLrisk_NURSETB GE 2 then
nursetb=2;run;
data men;set men;
Parish='.';
```

```
if par in ('2','3','5','7','9','10') then parish=1;
if par in ('1','4','6','8','11','12','13') then parish=2;
if par = '88' then parish=3;run;

WOMEN

data women;set women;
MaritalStatus='.';
If married = 2 then MaritalStatus=1;Else if married = 3 then MaritalStatus=1;
Else if married = 1 then MaritalStatus=2;Else if married = 4 then
MaritalStatus=2; Else if married = 5 then MaritalStatus=2;run;
Data women;set women;
EducationLevel='.';
if education='1' then EducationLevel='1';else if education='2' then
EducationLevel='2'; else if education='3' then EducationLevel='3';else if
education='4' then EducationLevel='3';Else if education ='5' then
EducationLevel='3';run;
data women;set women;
AgeCategory='.';
If age GE 15 and age LT 25 then AgeCategory=1;Else if age GE 25 and age LT 35
then AgeCategory=2;Else if age GE 35 and age LT 45 then AgeCategory=3;Else if
age GE 45 then AgeCategory=4;run;
data women;set women;
Occupation='.';
if occup1 in (3,4,6,8,14,16,18,19) then Occupation=1; *If you have more than
one value (3,4,6,8...) should have to use in instead of '=';
Else if occup1 in  (5,9,10,11,13,12) then Occupation=2;
else if occup1 in  (1,2,7,17,20,21,22 23 66 88 15) then Occupation=3;run;
Data women;set women;
SES='.';
if income='1' then SES='1';else if income='2' then SES='1'; else if
income='3' then SES='2';else if income='4' then SES='3';Else if Income ='5'
then SES='4';run;
Data women;set women;
Religious='.';
if religion='1' then Religious='1';else if religion='2' then Religious='2';
else if religion='3' then religious='3';else if religion='4' then
Religious='4';Else if religion ='5' then Religious='4';Else if religion ='88'
then Religious='4';run;
Data women;set women;
HomeTB='.';
if BLrisk_HOMETB='1' then HomeTB='1';else if BLrisk_HOMETB='2' then
HomeTB='2'; else if BLrisk_HOMETB='99' then HomeTB='2';run;
data women;set women;
Bedroom='.';
If BLrisk_BEDROOMTB = 1 then Bedroom=1;Else if BLrisk_BEDROOMTB GE 2 then
Bedroom=2; if BLrisk_BEDROOMTB = '.' then bedroom=2;run;
data women;set women;
TimeSpent='.';
If BLrisk_TIMESPENT = 1 then timespent=1;Else if BLrisk_TIMESPENT GE 2 then
timespent=2;run;
data women;set women;
NurseTB='.';
If BLrisk_NURSETB = 1 then nursetb=1;Else if BLrisk_NURSETB GE 2 then
nursetb=2;run;
data women;set women;
```

```
Parish='.';
if par in ('2','3','5','7','9','10') then parish=1;
if par in ('1','4','6','8','11','12','13') then parish=2;
if par = '88' then parish=3;run;

/*Checking distribution of variables*/
ods rtf file='P:\MANUSCRIPTS DATASETS\LTBI_Robert\Dissertation
datasets\Analysis2\Aim2 cohortdata\datasets\OutputTables\Completeness
Assessment.rtf';
proc freq data=Aim2;table AGE write ageyr sex married occupation agecategory
educationlevel maritalstatus BLrisk_KNEWTB_x    BLrisk_HABIT_x
      BLrisk_DRINK_x    BLrisk_PRISON_x   BLrisk_WORKHOME_x BLrisk_MINER_x
      BLrisk_TAXI_x     BLrisk_TRANS_x    BLrisk_HOSP_x outcome BLrisk_ADMIT
      BLrisk_TBHOME     BLrisk_STAYTB     BLrisk_HOMETB     BLrisk_BEDROOMTB
      BLrisk_TIMESPENT  BLrisk_NURSETB;run;

/*Completing Table 1(a) while checking the relationship with exposure*/
proc freq data=Aim2;tables Sex*agecategory/or chisq cmh;run;
proc freq data=Aim2;tables Sex*Hivresult/or chisq cmh;run;
proc freq data=Aim2;tables Sex*maritalstatus/or chisq cmh;run;
proc freq data=Aim2;tables Sex*Religious/or chisq cmh;run;
proc freq data=Aim2;tables Sex*educationlevel/or chisq cmh;run;
proc freq data=Aim2;tables Sex*SES/or chisq cmh;run;
proc freq data=Aim2;tables Sex*Occupation/or chisq cmh;run;ods rtf close;run;
proc freq data=Aim2;tables Sex*parish/or chisq cmh;run;
/*Completing Table 1(b) while checking the relationship with exposure*/
ods rtf file='P:\MANUSCRIPTS DATASETS\LTBI_Robert\Dissertation
datasets\analysis2\Tables_AIM2\Table 1(A).rtf';
proc freq data=Aim2;tables Sex*BLrisk_KNEWTB_x/or chisq cmh;run;
proc freq data=Aim2;tables Sex*BLrisk_HABIT_x/or chisq cmh;run;
proc freq data=Aim2;tables Sex*smokers2/or chisq cmh;run;
proc freq data=Aim2;tables Sex*BLrisk_DRINK_x/or chisq cmh;run;
proc freq data=Aim2;tables Sex*BLrisk_WORKHOME_x/or chisq cmh;run;
proc freq data=Aim2;tables Sex*BLrisk_MINER_x/or chisq cmh;run;
proc freq data=Aim2;tables Sex*BLrisk_TAXI_x/or chisq cmh;run;
proc freq data=Aim2;tables Sex*BLrisk_TRANS_x/or chisq cmh;run;
proc freq data=Aim2;tables Sex*BLrisk_HOSP_x/or chisq cmh;run;
proc freq data=Aim2;tables Sex*BLrisk_ADMIT/or chisq cmh;run;
proc freq data=Aim2;tables Sex*outcome2/or chisq cmh;run;
proc freq data=Aim2;tables Sex*BLrisk_STAYTB/or chisq cmh;run;
proc freq data=Aim2;tables Sex*BEDROOM/or chisq cmh;run;
proc freq data=Aim2;tables Sex*TIMESPENT/or chisq cmh;run;
proc freq data=Aim2;tables Sex*NURSETB/or chisq cmh;run;
proc freq data=Aim2;tables Sex*parish/or chisq cmh;run;
proc freq data=Aim2;tables Sex*HOMETB/or chisq cmh;run;
proc freq data=Aim2;tables Sex*hivresult/or chisq cmh;run;

/* Using the variable F1_hivno, all the failed to tress represent TRUE failed
to tress*/

proc freq data=Loss2fup;tables Sex*F1_ othhivno;run;

/*Completing Table 1(a) while checking the relationship with Outcome*/
proc freq data=Aim2;tables Outcome*agecategory/or chisq cmh;run;
proc freq data=Aim2;tables Outcome*Hivresult/or chisq cmh;run;
proc freq data=Aim2;tables Outcome*maritalstatusor chisq cmh;run;
```

```
proc freq data=Aim2;tables Outcome*Religious/or chisq cmh;run;
proc freq data=Aim2;tables Outcome*educationlevel/or chisq cmh;run;
proc freq data=Aim2;tables Outcome*SES/or chisq cmh;run;
proc freq data=Aim2;tables Outcome*Occupation/or chisq cmh;run;ods rtf
close;run;
proc freq data=Aim2;tables Outcome*parish/or chisq cmh;run;

/*Completing Table 1(b) while checking the relationship with Outcome*/
ods rtf file='P:\MANUSCRIPTS DATASETS\LTBI_Robert\Dissertation
datasets\analysis2\Tables AIM2\Table 1(A).rtf';
proc freq data=Aim2;tables Outcome*BLrisk_KNEWTB_x/or chisq cmh;run;
proc freq data=Aim2;tables Outcome*BLrisk_HABIT_x/or chisq cmh;run;
proc freq data=Aim2;tables Outcome*BLrisk_DRINK_x/or chisq cmh;run;
proc freq data=Aim2;tables Outcome*BLrisk_WORKHOME_x/or chisq cmh;run;
proc freq data=Aim2;tables Outcome*BLrisk_MINER_x/or chisq cmh;run;
proc freq data=Aim2;tables Outcome*BLrisk_TAXI_x/or chisq cmh;run;
proc freq data=Aim2;tables Outcome*BLrisk_TRANS_x/or chisq cmh;run;
proc freq data=Aim2;tables Outcome*BLrisk_HOSP_x/or chisq cmh;run;
proc freq data=Aim2;tables Outcome*BLrisk_ADMIT/or chisq cmh;run;
proc freq data=Aim2;tables Outcome*outcome/or chisq cmh;run;
proc freq data=Aim2;tables Outcome*BLrisk_STAYTB/or chisq cmh;run;
proc freq data=Aim2;tables Outcome*BEDROOM/or chisq cmh;run;
proc freq data=Aim2;tables Outcome*TIMESPENT/or chisq cmh;run;
proc freq data=Aim2;tables Outcome*NURSETB/or chisq cmh;run;
proc freq data=Aim2;tables Outcome*parish/or chisq cmh;run;
proc freq data=Aim2;tables Outcome*HOMETB/or chisq cmh;run;


/*Completing Table 1(a) while checking the relationship with Outcome*/
proc freq data=Aim2;tables agecategory*Outcome*Sex/or chisq cmh;run;
proc freq data=Aim2;tables Hivstatus*Outcome*Sex/or chisq cmh;run;
proc freq data=Aim2;tables maritalstatus*Outcome*Sex/or chisq cmh;run;
proc freq data=Aim2;tables Sex*Outcome*Religious/or chisq cmh;run;
proc freq data=Aim2;tables sex*Outcome*educationlevel/or chisq cmh;run;
proc freq data=Aim2;tables sex*Outcome*SES/or chisq cmh;run;
proc freq data=Aim2;tables sex*Outcome*Occupation/or chisq cmh;run;
proc freq data=Aim2;tables sex*Outcome*parish/or chisq cmh;run;

proc freq data=men;tables agecategory*outcome;run;
proc freq data=men;tables maritalstatus*outcome;run;
proc freq data=men;tables religious*outcome;run;
proc freq data=men;tables educationlevel*outcome;run;
proc freq data=men;tables SES*outcome;run;
proc freq data=men;tables occupation*outcome;run;
proc freq data=men;tables parish*outcome;run;
proc freq data=women;tables agecategory*outcome;run;
proc freq data=women;tables maritalstatus*outcome;run;
proc freq data=women;tables religious*outcome;run;
proc freq data=women;tables educationlevel*outcome;run;
proc freq data=women;tables SES*outcome;run;
proc freq data=women;tables occupation*outcome;run;
proc freq data=women;tables parish*outcome;run;

*/Above analysis this time using AIM2 dataset/*

Data=Aim2;
```

```
proc freq data=Aim2;tables BLrisk_KNEWTB_x*outcome;run;
proc means data= Aim2 sum mean min max ;
class outcome2 BLrisk_KNEWTB_x;
var person_time;
run;
proc means data= Aim2 sum mean min max ;
class outcome2 Taxiride;
var person_time;
run;


proc freq data=Aim2;tables BLrisk_DRINK_x*outcome;run;
proc freq data=Aim2;tables BLrisk_WORKHOME_x*outcome;run;
proc freq data=Aim2;tables Taxiride*outcome;run;
proc freq data=Aim2;tables StayTB*outcome;run;
proc freq data=Aim2;tables TBhome*outcome;run;
proc freq data=Aim2;tables homeTB*outcome;run;
proc freq data=Aim2;tables bedroom*outcome;run;
proc freq data=Aim2;tables timespent*outcome;run;

/*Creating New datasets*/
Data Men;set Aim2; If sex=1 then output Men;run;
Data Women;set Aim2; If sex=2 then output Women;run;
Data Loss2FUP;set Aim2;If outcome='.' then output Loss2FUP;run;

proc univariate data=Contactsdataset plot;var Enroll_age;histogram
Enroll_age;run; *Examining the Distribution of continous variables;
proc univariate data=Contactsdataset plot;var DM_Income;histogram
DM_Income;run;

PROC FREQ data= Aim2;
tables BLrisk_TAXI_x;
run;

/* Person time by year */

proc means data= Aim2 sum mean min max ;
class agecategory sex;
var person_time;
run;

* chi2 sex vs age by outcome;
proc sort data = Aim2;
by sex;
run;
PROC FREQ data= Aim2;
tables maritalstatus*outcome2 / chisq relrisk or;
by sex;
run;

/* Calculating person years for men and then women for both tables*/
proc means data= men sum mean min max ;
class agecategory sex;
var person_time;
run;
proc means data= women sum mean min max ;
```

```
class agecategory sex;
var person_time;
run;
proc means data= men sum mean min max ;
class maritalstatus sex;
var person_time;
run;
proc means data= women sum mean min max ;
class maritalstatus sex;
var person_time;
run;
proc means data= men sum mean min max ;
class Religious sex;
var person_time;
run;
proc means data= women sum mean min max ;
class religious sex;
var person_time;
run;
proc means data= men sum mean min max ;
class educationlevel;
var person_time;
run;
proc means data= women sum mean min max ;
class educationlevel;
var person_time;
run;
proc means data= men sum mean min max ;
class SES sex;
var person_time;
run;
proc means data= women sum mean min max ;
class SES sex;
var person_time;
run;
proc means data= men sum mean min max ;
class occupation sex;
var person_time;
run;
proc means data= women sum mean min max ;
class occupation sex;
var person_time;
run;
proc means data= men sum mean min max ;
class parish sex;
var person_time;
run;
proc means data= women sum mean min max ;
class parish sex;
var person_time;
run;
proc means data= men sum mean min max ;
class BLrisk_KNEWTB_x sex;
var person_time;
run;
proc means data= women sum mean min max ;
```

```
class BLrisk_KNEWTB_x sex;
var person_time;
run;
proc means data= men sum mean min max ;
class smokers sex;
var person_time;
run;
proc means data= women sum mean min max ;
class smokers sex;
var person_time;
run;

proc means data= men sum mean min max ;
class smokers2 sex;
var person_time;
run;
proc means data= women sum mean min max ;
class smokers2 sex;
var person_time;
run;

proc means data= men sum mean min max ;
class BLrisk_drink_x sex;
var person_time;
run;
proc means data= women sum mean min max ;
class BLrisk_drink_x sex;
var person_time;
run;

proc means data= men sum mean min max ;
class BLrisk_WORKHOME_x sex;
var person_time;
run;
proc means data= women sum mean min max ;
class BLrisk_WORKHOME_x sex;
var person_time;
run;
proc means data= men sum mean min max ;
class tbhome sex;
var person_time;
run;
 proc means data= women sum mean min max ;
class tbhome sex;
var person_time;
run;
proc means data= men sum mean min max ;
class staytb sex;
var person_time;
run;
 proc means data= women sum mean min max ;
class staytb sex;
var person_time;
run;
proc means data= men sum mean min max ;
class hometb sex;
```

```
var person_time;
run;
 proc means data= women sum mean min max ;
class hometb sex;
var person_time;
run;
proc means data= men sum mean min max ;
class bedroom sex;
var person_time;
run;
 proc means data= women sum mean min max ;
class bedroom sex;
var person_time;
run;
proc means data= men sum mean min max ;
class timespent sex;
var person_time;
run;
 proc means data= women sum mean min max ;
class timespent sex;
var person_time;
run;

proc means data= men sum mean min max ;
class hivresult sex;
var person_time;
run;
 proc means data= women sum mean min max ;
class hivresult sex;
var person_time;
run;

PROC FREQ data= Aim2;
tables BLrisk_TBhome BLrisk_StayTB HomeTB Bedroom TimeSpent;
run;

/*creating outcome variable2*/
data Aim2;
set Aim2;
if outcome = 1 then outcome2 = 1;
if outcome in (.,2,3) then outcome2 = 2;
run;

/*Univariate analysis computing relative risk between men and women
DEMOGRAPHICS TABLE*/

title  "Age sex vs outcome";
proc sort data = Aim2;
by agecategory;
run;
proc freq data = Aim2;
tables sex*outcome2/ chisq or relrisk;
by agecategory;
run;
```

```
title  "Age sex vs married";
proc sort data = Aim2;
by maritalstatus;
run;
proc freq data = Aim2;
tables sex*outcome2/ chisq or relrisk;
by maritalstatus;
run;

title  "Educationlevel vs sex";
proc sort data = Aim2;
by educationlevelbinary;
run;
proc freq data = Aim2;
tables sex*outcome2/ chisq or relrisk;
by educationlevelbinary;
run;


title  "SES vs sex";
proc sort data = Aim2;
by SESrecoded;
run;
proc freq data = Aim2;
tables sex*outcome2/ chisq or relrisk;
by SESrecoded;
run;

title  "Occupation vs sex";
proc sort data = Aim2;
by occupation;
run;
proc freq data = Aim2;
tables sex*outcome2/ chisq or relrisk;
by occupation;
run;

title  "Parish vs sex";
proc sort data = Aim2;
by parish2;
run;
proc freq data = Aim2;
tables sex*outcome2/ chisq or relrisk;
by parish2;
run;

proc means data= aim2 sum mean min max ;
class hivresult sex;
var person_time;
run;
 proc means data= women sum mean min max ;
class hivresult sex;
var person_time;
run;

proc means data= men sum mean min max ;
```

```
class timespent sex;
var person_time;
run;
 proc means data= women sum mean min max ;
class timespent sex;
var person_time;
run;

proc means data= men sum mean min max ;
class hivresult sex;
var person_time;
run;
 proc means data= women sum mean min max ;
class hivresult sex;
var person_time;
run;

proc means data= Loss2fup sum mean min max ;
class  outcome2 sex;
var person_time;
run;
 proc means data= women sum mean min max ;
class outcome sex;
var person_time;
run;

/*getting person time  for men and then for women for smokers2 who
converted*/
proc means data= men sum mean min max ;
class  outcome2 smokers2;
var person_time;
run;
 proc means data= women sum mean min max ;
class outcome2 smokers2;
var person_time;
run;

*/Univariate analysis computing relative risk between men and women RISK
FACTORS TABLE

title  "Knew someone with TB vs sex";
proc sort data = Aim2;
by BLrisk_KNEWTB_x;
run;
proc freq data = Aim2;
tables sex*outcome2/ chisq or relrisk;
by BLrisk_KNEWTB_x;
run;

title  "Smokers vs sex";
proc sort data = Aim2;
by smokers2;
run;
proc freq data = Aim2;
tables sex*outcome2/ chisq or relrisk;
by smokers2;
```

```
run;

title  "Alcohol use vs sex";
proc sort data = Aim2;
by BLrisk_DRINK_x;
run;
proc freq data = Aim2;
tables sex*outcome2/ chisq or relrisk;
by BLrisk_DRINK_x;
run;

title  "Works form home vs sex";
proc sort data = Aim2;
by BLrisk_WORKHOME_x;
run;
proc freq data = Aim2;
tables sex*outcome2/ chisq or relrisk;
by BLrisk_WORKHOME_x;
run;

title  "Taxi riders vs sex";
proc sort data = Aim2;
by taxiride;
run;
proc freq data = Aim2;
tables sex*outcome2/ chisq or relrisk;
by Taxiride;
run;

proc sort data = Aim2;
by tbhome;
run;
proc freq data = Aim2;
tables sex*outcome2/ chisq or relrisk;
by tbhome;
run;

proc sort data = Aim2;
by stayTb;
run;
proc freq data = Aim2;
tables sex*outcome2/ chisq or relrisk;
by stayTB;
run;

proc sort data = Aim2;
by HomeTB;
run;
proc freq data = Aim2;
tables sex*outcome2/ chisq or relrisk;
by HomeTB;
run;


proc sort data = Aim2;
by Bedroom;
```

```
run;
proc freq data = Aim2;
tables sex*outcome2/ chisq or relrisk;
by Bedroom;
run;

proc sort data = Aim2;
by Timespent;
run;
proc freq data = Aim2;
tables sex*outcome2/ chisq or relrisk;
by Timespent;
run;

proc sort data = Aim2;
by hivresult;
run;
proc freq data = Aim2;
tables sex*outcome2/ chisq or relrisk;
by hivresult;
run;

/* Model building using GENMOD */
proc freq data=Aim2;
proc genmod data=Aim2;
where outcome2 in (1,2);
class sex (ref='2');
model outcome2 (event='1')=sex/dist=bin link=log;
estimate 'RR men versus women' sex 1 -1/e exp;
title 'Crude Prevalence ratio by sex';
run;

proc freq data=Aim2;
proc genmod data=Aim2;
where outcome2 in (1,2);
class agecategory(ref='1');
model outcome2 (event='1')= agecategory/dist=bin link=log;
estimate 'PR Age 25-34 VS 0-5' AgeCategory 1 0 0 -1/e exp;
estimate 'PR Age 35-44 VS 0-5' AgeCategory 0 1 0 -1/e exp;
estimate 'PR Age 45+' AgeCategory 0 0 1 -1/e exp;run;

proc freq data=Aim2;
proc genmod data=Aim2;
where outcome2 in (1,2);
class BLrisk_KNEWTB_x(ref='2');
model outcome2 (event='1')= BLrisk_KNEWTB_x/dist=bin link=log;
estimate 'RR No versus Yes' BLrisk_KNEWTB_x 1 -1/e exp;run;

proc freq data=Aim2;
proc genmod data=Aim2;
where outcome2 in (1,2);
class smokers2(ref='2');
model outcome2 (event='1')= smokers2/dist=bin link=log;
estimate 'RR Smokers versus NonSmokers' smokers2 1 -1/e exp;run;

proc freq data=Aim2;
```

```
proc genmod data=Aim2;
where outcome2 in (1,2);
class BLrisk_WORKHOME_x(ref='2');
model outcome2 (event='1')= BLrisk_WORKHOME_x/dist=bin link=log;
estimate 'RR works from home versus away form home' BLrisk_WORKHOME_x 1 -1/e
exp;run;

proc freq data=Aim2;
proc genmod data=Aim2;
where outcome2 in (1,2);
class tbhome(ref='2');
model outcome2 (event='1')= tbhome/dist=bin link=log;
estimate 'RR Never lived versus Lived' tbhome 1 -1/e exp;run;

proc freq data=Aim2;
proc genmod data=Aim2;
where outcome2 in (1,2);
class stayTB(ref='2');
model outcome2 (event='1')= stayTB/dist=bin link=log;
estimate 'RR stayed with TB versus never' stayTB 1 -1/e exp;run;

proc freq data=Aim2;
proc genmod data=Aim2;
where outcome2 in (1,2);
class HomeTB(ref='2');
model outcome2 (event='1')= HomeTB/dist=bin link=log;
estimate 'RR stayed with TB versus never' HomeTB 1 -1/e exp;run;

proc freq data=Aim2;
proc genmod data=Aim2;
where outcome2 in (1,2);
class Bedroom(ref='2');
model outcome2 (event='1')= Bedroom/dist=bin link=log;
estimate 'RR stayed with TB versus never' Bedroom 1 -1/e exp;run;

proc freq data=Aim2;
proc genmod data=Aim2;
where outcome2 in (1,2);
class Timespent(ref='2');
model outcome2 (event='1')= Timespent/dist=bin link=log;
estimate 'RR stayed with TB versus never' Timespent 1 -1/e exp;run;

proc freq data=Aim2;
proc genmod data=Aim2;
where outcome2 in (1,2);
class hivresult(ref='2');
model outcome2 (event='1')= hivresult/dist=bin link=log;
estimate 'RR hiv +ve versus hiv -ve' hivresult 1 0 -1/e exp;
estimate 'Other versus hiv -ve' hivresult 0 1 -1/e exp;run;

/*All significant variables at Univariate level (Age and sex and bedroom)
plus HIV*/

/*Final association btn sex and LTBI from this model is RR=1.69;P<0.01*/
proc freq data=Aim2;
proc genmod data=Aim2;
```

```
where outcome2 in (1,2);
class sex (ref='2')  Bedroom(ref='2') hivresult(ref='2')
AgeCategory(ref='1');
model outcome2 (event='1')= sex Bedroom hivresult AgeCategory/dist=bin
link=log;
estimate 'Men versus women' sex 1 -1/e exp;
estimate 'shared bedroom versus never shared' bedroom 1 -1/e exp;
estimate 'PR Age 25-34 VS 0-5' AgeCategory 1 0 0 -1/e exp;
estimate 'PR Age 35-44 VS 0-5' AgeCategory 0 1 0 -1/e exp;
estimate 'PR Age 45+' AgeCategory 0 0 1 -1/e exp;
estimate 'HIV positive versus HIVNegative' hivresult 1 0 -1/e exp;
estimate 'HIV Other versus HIVNegative' hivresult 0 1 -1/e exp;run;
```