

https://dspace.org/

Prevalence of latent tuberculosis using the tuberculin skin test in two human immunodeficiency virus endemic municipalities in South Africa (2016 – 2018)

Murphy, Williams Victor

2019

https://hdl.handle.net/10539/29881

Downloaded from a DSpace repository. https://dspace.org/

PREVALENCE OF LATENT TUBERCULOSIS USING THE TUBERCULIN SKIN TEST IN TWO HUMAN IMMUNODEFICIENCY VIRUS ENDEMIC MUNICIPALITIES IN SOUTH AFRICA (2016 – 2018)



UNIVERSITY OF THE
WITWATERSRAND,
JOHANNESBURG

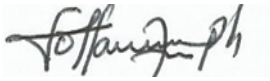
Dr. Williams, Victor Murphy

A research report submitted to the Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, in partial fulfillment of the requirements for the degree of Master of Science in Epidemiology in the field of Epidemiology and Biostatistics

Johannesburg, August 2019

DECLARATION

I, Dr. Williams, Victor Murphy hereby declare that this research report is my own work. It is being submitted in partial fulfillment for the degree of MSc in the field of Epidemiology and Biostatistics, School of Public Health, Faculty of Health Sciences, University of the Witwatersrand. It has not been submitted before for any degree or examination at this or any other university. Where I have used the thoughts or ideas of others, the required referencing conventions have been adhered to.

Signed: 

Date: 12th August 2019

DEDICATION

I dedicate this research report to my wife Idorenyin for all her support and my mother Dom who made me realize the value of education at a young age.

ACKNOWLEDGMENTS

My supervisors, Dr. Neil Martinson and Dr. Kennedy Otwombe of the Perinatal HIV Research Unit (PHRU) provided enormous support during the protocol development and writing of this research report. I thank them for their supervision and mentoring. My gratitude goes to the academic and non-academic staff in the Division of Epidemiology and Biostatistics for their regular support during lectures, consultations and mentoring which provided a solid foundation for this research report. I am also grateful to Mrs. Zodwa Ndlovu and Innocentia Vukeya for their guidance which has enabled me to complete this report.

Many thanks to the PHIRST Study principal investigators Prof. Cheryl Cohen and Dr. Jocelyn Moyes of the National Institute for Communicable Diseases (NICD) for granting me permission to analyze the PHIRST study data and also providing guidance on the research report whenever it was needed. I acknowledge the contributions of the entire PHIRST Study team including the field teams at Agincourt and Klerksdorp. Without their effort, this research report would not have been possible.

While studying and writing this research report, I received immense support from my family especially my wife Idorenyin for her regular inspiration, my mother Dom and Prof. Victor Akpan. I appreciate Dr. Marianne Calnan, Dr. Jude Unuafe and Rev. Fr. Enimabasi Akpan for their constant encouragement and motivation through this period.

ABSTRACT

Background: Tuberculosis is a global health threat and a leading cause of mortality in low and middle-income countries of Africa, Asia and Western Pacific responsible for about 1.6 million deaths in 2017. In South Africa, it is the highest cause of mortality. This is compounded by the high HIV prevalence which has been noted to be the leading cause of immunosuppression leading to reactivation of latent tuberculosis to active tuberculosis disease.

Methods: The research adopted a cross-sectional study design approach. This was in the form of secondary analysis of baseline data from a completed cohort study (PHIRST Study: 2016 - 2018) to determine transmission dynamics of infectious respiratory conditions including tuberculosis. The Analysis included determination of prevalence of latent tuberculosis infection at two sites - Agincourt and Klerksdorp using TST, and annual risk of tuberculosis infection. A Logistic and multi-level analysis was used to predict risk factors for latent tuberculosis infection both at the individual and household level.

Results: The overall prevalence of TST-positive was 20.5%. Prevalence was higher at Klerksdorp (24.3 %) compared to Agincourt (16.3%) and it increased with increase in age. A prevalence of 1.9 - 45% was seen in Agincourt and 5.3 - 51.9% in Klerksdorp. The lowest prevalence was seen in <5 age group and the highest in the 41-45 age group at both sites. A significant association was seen between HIV status and TST-positive prevalence (OR=1.63; 95% CI: 1.13, 2.35, p=0.009) but this became insignificant after controlling for age and sex (OR=1.12; 95% CI: 0.76, 1.65; p=0.567). Overall household prevalence was 54.2% with Klerksdorp having a higher household prevalence of 70.5% and Agincourt 36%.

The annual risk of TB infection (ARTI) was 3.8% in participants aged ≤15 years and under-five age group had the highest value. ARTI reduced with increase in age, higher in males and in those with negative and unknown HIV status. Individual risk factors for latent TB infection were being a resident of Klerksdorp and increasing age. HIV status, past TB history, sex, and BMI were not significant risk factors for TST-positivity. At the household level, the number of windows, smoking within households and wealth index were significant predictors of latent TB infection.

Conclusion: Prevalence of TST-positivity is higher in Klerksdorp both at the individual and household level compared to Agincourt. Younger age was associated with a higher risk of TB infection and the risk of infection decreased with increase in age. Increasing age, living in Klerksdorp, number of windows in a house, smoking within households and wealth index significantly predicted TST-positivity.

Keywords: *Latent TB infection, prevalence of latent TB infection, TST positive, risk factors for latent TB infection, Annual risk of tuberculosis infection.*

TABLE OF CONTENTS

DECLARATION ii

DEDICATION iii

ACKNOWLEDGMENTS iv

ABSTRACT v

TABLE OF CONTENTS vii

LIST OF FIGURES x

LIST OF TABLES xi

LIST OF APPENDICES xii

ABBREVIATIONS xiii

DEFINITION OF TERMS xiv

1 CHAPTER ONE: INTRODUCTION 1

1.1 BACKGROUND 1

1.2 LITERATURE REVIEW 3

1.2.1 Latent Tuberculosis Infection (LTBI) 3

1.2.2 Latent Tuberculosis Infection Screening 4

1.2.3 Limitations of TST 5

1.2.4 TST in high HIV and Tuberculosis prevalence 6

1.2.5 Conceptual Framework for determination of TST-positivity 7

1.3 PROBLEM STATEMENT 8

1.4 JUSTIFICATION 8

1.5 RESEARCH QUESTION, HYPOTHESIS, AIM, AND OBJECTIVES 8

1.5.1 Research Question 8

1.5.2 Hypothesis 8

1.5.3 Aim 9

1.5.4 Objectives 9

2 CHAPTER TWO: METHODOLOGY 10

2.1 THE PRIMARY STUDY 10

2.1.1 Background and study setting 10

2.1.2 Study design 10

2.1.3 Study Site 10

2.1.4 Study population 12

2.1.5 Sampling 13

2.1.6 Inclusion Criteria 13

2.1.7 Data Collection 13

2.2 SECONDARY DATA ANALYSIS 14

2.2.1 Study design 14

2.2.2 Data source 14

2.2.3 Study population 15

2.2.4 Power and Sample Size 15

2.2.5 Description of Variables 15

2.2.6 Generation of new variables 17

2.3 DATA MANAGEMENT 18

2.4 STATISTICAL ANALYSIS 19

2.4.1 Descriptive Analysis 19

2.4.2 Prevalence of Latent Tuberculosis Infection (TST-Positive) 19

2.4.3 Annual risk of tuberculosis infection (ARTI) 21

2.4.4 TST-Positive prevalence and HIV status 21

2.4.5 Predictors of TST-Positive 21

2.5 ETHICAL CONSIDERATIONS 23

3 CHAPTER THREE: RESULTS 25

3.1 BASELINE CHARACTERISTICS 25

3.2 PREVALENCE OF TST-POSITIVE 28

3.2.1 Distribution of TST measurements 28

3.2.2 TST-Positive prevalence overall and by site 29

3.2.3 TST-Positive prevalence using different TST cut off 32

3.2.4 TST-Positive Prevalence and nutritional status 34
3.2.5 Household prevalence of TST-Positive 35
3.3 ANNUAL RISK OF TB INFECTION (ARTI) 36
3.3.1 ARTI by age and site 37
3.3.2 ARTI by year, sex and site 39
3.3.3 ARTI by HIV status, site, and sex 39
3.4 TST-POSITIVITY BASED ON HIV STATUS 41
3.4.1 Descriptive Epidemiology 41
3.4.2 Controlling for age and sex 41
3.5 RISK FACTORS OF TST-POSITIVITY 41
3.5.1 Individual risk factors 42
3.5.2 Household-level risk factors 45
4 CHAPTER FOUR: DISCUSSION 48
4.1 SUMMARY OF KEY RESULTS 48
4.2 COMPARISON WITH OTHER STUDIES 49
4.2.1 Prevalence of latent tuberculosis 49
4.2.2 Annual risk of tuberculosis infection 53
4.2.3 TST-Positivity based on HIV status 54
4.2.4 Risk factors for a Positive TST 54
4.3 STRENGTH AND LIMITATIONS OF THE STUDY 57
4.3.1 Strength of the study 57
4.3.2 Limitations 57
5 CHAPTER FIVE: RECOMMENDATIONS AND CONCLUSION 59
5.1 RECOMMENDATIONS 59
5.2 CONCLUSION 60
5.3 RECOMMENDATIONS FOR FUTURE RESEARCH 60
REFERENCES 61

LIST OF FIGURES

Figure 1.1 Conceptual framework for determination of TST-positivity..... 7
Figure 2.1 Map showing Agincourt research site in Mpumalanga Province, South Africa (obtained from Anstey et al, 2018) 11
Figure 2.2 Map of South Africa showing the location of Klerksdorp in the Northwest Province 12
Figure 2.3 Schema for selection of study participants from the data set for analysis..... 15
Figure 3.1 Distribution of TST measurements greater than zero by site 28
Figure 3.2 TST measurements greater than zero by age category and site..... 29
Figure 3.3 TST-positive prevalence by site and age category 31
Figure 3.4 A bar graph of TST-positive prevalence by HIV status, site, and sex 32
Figure 3.5 TST-positive Prevalence at 5mm, 10mm and 15mm TST cut-off by age..... 34
Figure 3.6 Median ARTI by site and age category for all participants..... 38
Figure 3.7 Median ARTI by site in participants <15 years of age..... 39
Figure 3.8 Bar graph of median ARTI by year, sex and site 40
Figure 3.9 Bar graph median ARTI by HIV status, sex and site 40

LIST OF TABLES

Table 3.1 Baseline characteristics of participants and households in the study 26
Table 3.2 Comparison of TST-Positive prevalence between sites and by some variables using TST measurement of ≥10 mm (≥5 mm if HIV +ve) as a positive test 30
Table 3.3 TST-Positive Prevalence using different cut-offs of TST Measurement 33
Table 3.4 TST-Positive prevalence based on upper arm circumference and BMI 35
Table 3.5 Comparison of household TST-positive prevalence by site 36
Table 3.6 Description of median ARTI by site 37
Table 3.7 Logistic model of individual-level factors which predict TST-Positive ignoring clustering within households (Participants ≤15 years of age) 42
Table 3.8 Logistic model of individual-level factors which predict TST-Positive ignoring clustering within households (Participants ≥16 years of age) 43
Table 3.9 Multi-level modeling (two-level) of household risk factors of TST-Positive taking into account clustering of households within the site 46
Table 3.10 Logistic model of household risk factors of TST-Positive at Klerksdorp 47

LIST OF APPENDICES

Appendix 1 Ethical clearance certificate 69

Appendix 2 Plagiarism declaration form 70

Appendix 3 Approval from PHIRST study principal investigator 71

Appendix 4 Graphical description of wealth quintile by site..... 72

Appendix 5 Graphical description of 0.0 mm TST measurement by age..... 73

Appendix 6 Smoothed graphs describing TST-Positive prevalence..... 74

Appendix 7 Logistic model of individual factors which predict TST-positive within household in Klerksdorp..……..75

Appendix 8 Logistic models of individual predictive factors based on HIV status 76

Appendix 9 Multi-level modelling (three levels) of individual risk factors of TST-Positive considering clustering within households 78

ABBREVIATIONS

ARTI	Annual Risk of Tuberculosis Infection
BCG	Bacille Calmette-Guérin
BMI	Body Mass Index
CD4+	Cell Differentiation 4
CI	Confidence Interval
HIV	Human Immunodeficiency Virus
IGRA	Interferon Gamma Release Assay
LTBI	Latent Tuberculosis Infection
NICD	National Institute for Communicable Diseases
OR	Odds ratio
PLHIV	People Living with Human Immunodeficiency Virus
PHIRST	A Prospective Household observational cohort study of Influenza, Respiratory Syncytial virus and other respiratory pathogens community burden and Transmission dynamics in South Africa
PHRU	Perinatal HIV Research Unit
PPD	Purified Protein Derivative
STATSA	Statistics South Africa
TB	Tuberculosis
TST	Tuberculin Skin Test
WHO	World Health Organisation

DEFINITION OF TERMS

Latent TB Infection	A state of a persistent immune response to stimulation by *Mycobacterium tuberculosis* antigens without evidence of clinically manifested active TB.
Prevalence	The proportion of a population that is affected by a certain disease at a specified time.
Sensitivity	The ability of a test to correctly identify those with a particular disease.
Specificity	The ability of a test to correctly identify those without the disease.
Crowding Index	The number of people living in a dwelling divided by the number of rooms.

1 CHAPTER ONE: INTRODUCTION

This chapter describes the tuberculosis epidemic in South Africa and globally with its public health impact in the advent of HIV. It reviews relevant literature on the significance of latent tuberculosis infection in the TB epidemic and the role of TST in control efforts. The chapter in the concluding part discusses the problem statement, justification, and objectives of the study.

1.1 Background

Tuberculosis is caused by *Mycobacterium tuberculosis* in susceptible individuals and is transmitted primarily by inhalation of TB bacilli in aerosol into the lungs. Tuberculosis causes high mortality globally with approximately 10 million new cases and 1.6 million deaths in 2017 (1). Of all the new cases, 58% were in males and 90% were in adults aged ≥15 years (1). Nine percent (0.9 million) of the recorded TB cases in 2017 occurred amongst people living with HIV and 72% of these reside in Africa (2). Tuberculosis has been identified as one of the highest causes of mortality globally and in 2017, it accounted for about 300, 000 deaths in HIV infected individuals (1, 2). Globally, the incidence of tuberculosis is noted to be highest in South-East Asia, the Western Pacific Region, and Africa which also has the highest number of infections associated with HIV and mortality (1). Thirty high burden countries account for 87% of the global TB incidence and eight are responsible for two-thirds of the global tuberculosis incidence with India having the highest number of new infections annually, closely followed by China (1, 2). Whereas South Africa, Lesotho, and the Philippines have the highest annual incidence (1). The world's tuberculosis burden is complicated by the emergence and gradual increase in cases of the resistant form of the tuberculosis bacteria to effective first-line drugs. Drug-resistant TB is more difficult to treat and according to the World Health Organisation (WHO), only 55% of patients achieve cure after completing treatment globally (1, 3). A recent publication of treatment outcome of MDR-TB patients in Myanmar showed a treatment success rate of 80%. HIV positive status without treatment and age >55 years were positive predictors of the poor treatment outcome (1, 4).

In Africa, South Africa has one of the highest-burden of tuberculosis after Nigeria and is one of the eight countries which contributes to 66% of the total global annual TB incidence (1, 2, 5). The prevalence of TB in South Africa was 696/100, 000 population in 2014 (6) with an

annual incidence of 567/100, 000 population in 2017 (7). Tuberculosis is currently the leading cause of mortality in South Africa, causing an estimated 78, 000 deaths in 2017 with a case fatality rate of 20% (1, 8). These high indices are noted to be due to the HIV epidemic which has plagued the country further leading to a very high mortality rate (2, 3, 5). Though a lot of effort has gone into HIV care and TB control, the burden of tuberculosis is still high and a large percentage of the South African population are infected with the latent form of tuberculosis (5) which can progress to active disease, thus limiting control efforts as most control activities are targeted towards active disease (9).

Prevalence studies of latent tuberculosis have previously focused on high-risk groups and in few locations, without considering the general population (10-12). These studies have indicated the prevalence of latent tuberculosis infection varies from 19 - 89% of the South African population depending on the age group, location and population group (5, 10, 13). However, there are no contemporaneous estimates of the population prevalence of LTBI. Thus, it is necessary to develop an effective means of identifying the percentage of the entire population infected with latent tuberculosis infection for prophylaxis to minimize progression to active disease (3). This forms part of the first pillar of End-TB Strategy which is integrated patient-centered care and prevention. WHO has adopted this to ensure people in the high-risk group e.g. PLHIV, miners, children below 5 years of age coming in contact with infectious cases are protected from developing active tuberculosis infection (14).

Besides identifying those eligible for prophylaxis against TB, accurate determination of latent TB prevalence provides a basis for cross-sectional measures to estimate incidence rate: the annual risk of tuberculosis infection (ARTI) in a population (15, 16). ARTI indicates the proportion of a particular population who will become infected in a community within a particular year and is based on Styblo rule that 50 incident cases of tuberculosis /100, 000 population per year corresponds to 1% ARTI (17, 18). Although this rule has been invalidated because it was postulated when there were no functional TB control programs, effective TB treatments, and no HIV epidemic, it nevertheless, provides trend estimates useful in planning TB control activities and measuring the impact of HIV on tuberculosis transmission (16, 18).

1.2 Literature review

Early diagnosis and treatment of tuberculosis are vital for its control and the subsequent elimination in high burden countries and globally (1, 19). Achieving this will require adopting approaches that will enhance the detection of undiagnosed LTBI and providing adequate therapy to limit progression to active disease in people who are at risk (9, 14, 19). A major challenge to management of the latent form of tuberculosis is the lack of ideal diagnostic technic which presently is limited to the tuberculin skin test (TST) and interferon-gamma release assays (IGRA) (14). A large number of new tuberculosis cases are diagnosed yearly despite control efforts and this is believed to be due to reactivation of previous latent infections in high-risk groups and new infections in children and PLHIV who are in contact with people who have active tuberculosis disease (3).

1.2.1 Latent Tuberculosis Infection (LTBI)

This is the presence of an immune response to stimulation by *Mycobacterium tuberculosis* antigen without manifestation of clinical signs of tuberculosis disease (9, 14). This immune suggests there is the potential for progression to active disease if the individual becomes immunosuppressed for any reason including HIV infection, diabetes or malnutrition (9). The global burden of LTBI is not known but current estimation is that about 1.7 billion people (i.e. 23% of the world's population) are infected with latent TB and 5 – 15% of these infected people will go on to develop active TB disease in their lifetime (1, 2, 9, 14). A 2018 systematic review and meta-analysis estimated the global prevalence of LTBI to be 24.8% using IGRA and 24.1% using a TST cut off of 10.0 mm (20). In some tuberculosis endemic countries, more than 80% of the population has latent tuberculosis infection (5), especially amongst high-risk population groups including PLHIV, miners and healthcare workers (10, 21, 22).

Advanced age, past treatment for tuberculosis and living in households with residents who have active tuberculosis have been identified as some of the risk factors for latent tuberculosis infection (5, 23, 24). The prevalence of LTBI is noted to be different amongst different risk groups in different countries. This variability depends largely on the availability of access to screening services, prophylactic treatment, and prevalence of HIV which has been noted to contribute significantly to the progression of latent tuberculosis to active disease (1).

Environmental and household factors such as regular exposure to tobacco smoke and household indoor air pollution from cooking and heating have been noted to contribute to TB infection. According to WHO, smoking accounts for 7.9% of TB cases worldwide (2). This is further collaborated by two systematic reviews conducted in 2007 and 2015 respectively pointing to the fact that exposure to tobacco smoke and smoke from different types of domestic fuel used in households increases the risks of developing TB particularly in children and women (25, 26).

1.2.2 Latent Tuberculosis Infection Screening

Although no standardized test for LTBI is available, WHO recommends screening of at-risk populations with either TST or in full at first appearance IGRA, with a view to offering those identified to be positive, prophylactic treatment to minimize progression to active tuberculosis disease (14). Reactivation of latent tuberculosis in the populations at risk e.g. people infected with HIV, miners, and children <5 years of age can lead to active disease and ongoing infection of those they come in contact with (3).

TST is widely used for diagnosis of previous *M. tuberculosis* infection with a view to providing preventive treatment (14, 27), and as an ancillary diagnostic for children. In research, it is useful in latent tuberculosis prevalence studies in different population groups (10, 12, 27, 28). The test was first used by Robert Koch and demonstrates delayed-type cutaneous reactivity to tuberculin in persons who have been previously sensitized (29). It involves the injection of 0.1ml (5 tuberculin unit) of the Purified Protein Derivative (PPD) into the anterior surface of the forearm, and the induration formed (palpable swelling) is read 48-72 hours later (29, 30).

IGRA is a more recent screening method which measures the blood concentration of interferon released by leucocytes in response to *M. tuberculosis* antigen stimulation (14, 19). The measured value of interferon is read off against standardized values to determine if it is positive or not. Two types of IGRA are commercially available for use - QuantiFERON®-TB Gold In-Tube and T-SPOT®.*TB* (14, 19).

TST and IGRA are comparable and recommended with a sensitivity of 75-90% and 80-95% respectively, and a specificity of 70-95% and 95-100% respectively (19, 31). IGRA is noted to

have a higher specificity but more expensive and technical to operationalize compared to TST, thus limiting its use in resource-limited settings (14, 28, 32)**.** Neither test can predict with absolute certainty, the progression to active TB disease (33). However, findings in a recently published (2018) prospective cohort study conducted in the UK to determine the prognostic value of IGRA and TST in predicting the development of active TB disease showed values of incidence by IGRA and TST are comparable, depending on the type of IGRA used and the TST cut off (34). Incidence was higher amongst participants who had a positive screening test at baseline with an incidence of 13.2 per 1000 person-years using T-SPOT.*TB*, 10.1 per 1000 person-years using QuantiFERON-TB Gold In-Tube and 11.1 per 1000 person-years using TST with cut off of 15.0 mm (34).

1.2.3 Limitations of TST

Although useful for determining latent TB infection, false-positive test, false-negative test and inability to differentiate between latent and active TB are some of the limitations of TST. These limitations are responsible for the variability observed in sensitivity and specificity of TST making it less efficient as a screening test for LTBI especially in the presence of factors which affect the immune response to tuberculin.

Different studies have documented the possible association between TST and BCG or some clinical conditions (35-37) which may result in either a false positive or false negative TST result. TST response was found to be unreliable with a false positive response in young BCG vaccinated children and those with short intervals between BCG vaccination and receiving TST (29, 38). Where BCG vaccination does not affect TST outcome include being an adult (≥15 years of age), having TST more than 15 years after BCG vaccination and a TST measurement greater than 15.0 mm (38, 39). Infection with non-tuberculous mycobacteria, poor tuberculin injection technic and errors during measurement and interpretation of induration could also produce a false positive TST result (29).

Severe malnutrition, being on treatment for rheumatoid arthritis and a positive HIV status can produce a diminished TST response (29, 37, 38, 40) leading to a false negative test. This is due to a dysfunctional immune system associated with these conditions. Insufficient time (between

infection and TST testing) for the body to mount an immune response, old tuberculous infection or overwhelming TB disease are also predictors of a false negative TST result (41).

Considering the fact some people receive preventive treatment for tuberculosis based on the TST result (9), it is important to interpret TST with caution given that a large proportion of the population in most tuberculosis endemic countries have received BCG vaccination. Although TST provides more reliable results in tuberculosis endemic countries compared to less endemic countries (38), care must, however, be taken to differentiate positive TST as a result of other environmental *mycobacterial* infection and *M. tuberculosis* (29, 31).

1.2.4 TST in high HIV and Tuberculosis prevalence

In HIV and TB endemic settings, TST has been successfully utilized in screening for latent TB in different high-risk populations (e.g. children, miners, healthcare workers) and the general population (5, 10, 13, 21, 22, 42), but it showed low predictive value amongst HIV positive participants in a previous latent tuberculosis prevalence study conducted in Johannesburg, South Africa (13). HIV status also did not have any effect on TST in a study to determine the prevalence of TST-positivity amongst South African miners and PLHIV in India (10, 40).

In the Johannesburg - South Africa study conducted by Ncayiyana et al in 2016, the overall prevalence of TST-positive was 34% varying from 19% in the 0 – 14 years age group to 45% amongst those aged 45 years and above. No association was noted between HIV status and prevalence in this study. Rather, age, sex, and socioeconomic status were noted to be associated with latent tuberculosis prevalence (13). In another study completed in 2009 to determine the effect of increasing age on prevalence in HIV negative individuals in Cape Town, TST had good predictive value with TST-positive prevalence ranging from 28% in those aged 5 – 10 years to 88.2% in the 31 – 35 years age bracket. No statistical difference in TST-positive prevalence was noted between males and females (5).

A prevalence of 60.4% was noted when TST was used to determine the prevalence of latent TB amongst 328 healthcare workers in Iran (22). Being a nurse, increased years of experience and contact with TB patients were the main risks linked with a positive test. There was no association between age, sex, and BCG scar with a positive skin test (22). However, age of

contact, presence of BCG scar, relationship with an index TB case and proximity of sleeping site between contact and index case determined positivity of TST in a prevalence study amongst tuberculosis contacts in Pakistan (43, 44).

1.2.5 Conceptual Framework for determination of TST-positivity

The conceptual framework for the study is centered on the Social-Ecological Model which examines the multilevel interactions of personal, social and environmental factors which determines a health outcome (45). The different exposure variables in the study are grouped into five- clinical, demographic, household, social and nutritional /anthropometric variables. The conceptual framework explores the various possible interactions which exist between the exposure variables and could result in or have an effect on the outcome of interest – TST-positivity. The framework is described in figure 1.1.

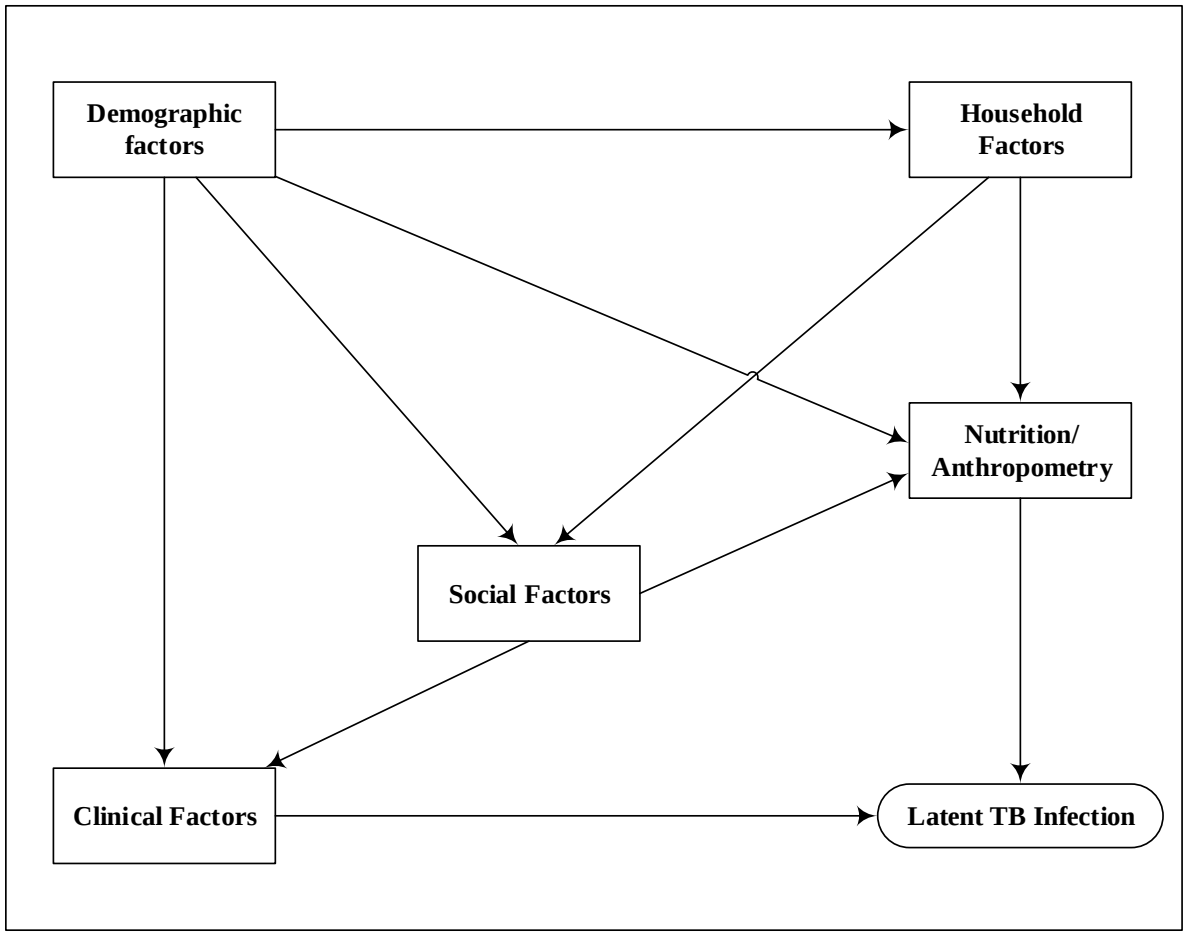


Figure 1.1 Conceptual framework for determination of TST-positivity

1.3 Problem statement

The true prevalence of latent TB is not known in the general population (5, 13) as more studies have considered specific high-risk populations and not the general population (10-12). This lack of knowledge of prevalence in the general population, has limited treatment of latent tuberculosis to the high-risk groups, neglecting the general population which may become a source of tuberculosis infection whenever those with untreated latent tuberculosis develop the active disease (24). The high prevalence of HIV in South Africa complicates the challenge of latent tuberculosis as it has been documented to be the commonest cause of progression to active tuberculosis disease and mortality amongst tuberculosis patients (7).

1.4 Justification

This study will help determine the prevalence of latent tuberculosis infection in the sampled population. This will represent the general population of Agincourt and Klerksdorp, two very different populations but both representative of a large part of the population of South Africa thus providing an estimate of latent tuberculosis infection (14) that could be extrapolated to similar areas in South and Southern Africa. Limitations identified in previous studies such as restricting participants to a certain age group and HIV status, the inadequate sample size to account for children and impact of exposure to indoor air pollution within households on TST outcome will be addressed (5, 13). The possible association between different exposures and tuberculin skin test positivity will also be tested.

1.5 Research question, Hypothesis, Aim, and Objectives

1.5.1 Research Question

What is the prevalence of latent tuberculosis infection measured with the TST in Klerksdorp and Agincourt Research Site from 2016 - 2018?

1.5.2 Hypothesis

There is no difference in the prevalence of latent tuberculosis infection at Klerksdorp and Agincourt Research Site.

1.5.3 Aim

To determine the prevalence and predictive factors of latent tuberculosis infection in Klerksdorp (Northwest Province) and Agincourt Research site (Mpumalanga Province) from 2016 - 2018, using the Mantoux method of TST administration and interpretation.

1.5.4 Objectives

i. To determine the age-specific prevalence of latent tuberculosis infection using the tuberculin skin test at both sites.
ii. To determine the annual risk of tuberculosis infection (ARTI) amongst participants by age and gender.
iii. To compare the prevalence of TST-positivity amongst HIV positive and HIV negative participants, controlling for age and sex.
iv. To determine risk factors for TST-positivity overall and by age, gender and HIV status.

2 CHAPTER TWO: METHODOLOGY

This chapter describes the setting in which the research was conducted, the study design, study sites, and population, and variables used in the analysis. Also, the analysis technic, data management processes and ethical considerations are reviewed.

2.1 The Primary Study

2.1.1 Background and study setting

The primary study was a household level prospective cohort study conducted at two sites by the National Institute for Communicable Diseases (NICD), Agincourt Health and Socio-Demographic Surveillance Site (HDSS) and the Perinatal HIV Research Unit (PHRU), a research unit of Wits University. It was titled A Prospective Household observational cohort study of Influenza, Respiratory Syncytial virus, and other respiratory pathogens community burden and Transmission dynamics in South Africa (The PHIRST Study).

The PHIRST Study was conducted in one rural community (Agincourt – Mpumalanga Province) and an urban community (Klerksdorp (Matlosana) – Northwest Province). Residents of these two communities were followed up annually for three consecutive years (2016 – 2018) and different specimens were collected at intervals specified in the study protocol after baseline information was obtained. This was used to determine the burden and transmission dynamics associated with respiratory pathogens at the two study sites including tuberculosis.

2.1.2 Study design

A cohort study design was adopted for the PHIRST study. New cohorts were recruited annually and followed up for 12 months from 2016 – 2018.

2.1.3 Study Site

The study was conducted at Klerksdorp in the Northwest Province and Agincourt Health and socio-Demographic Surveillance Site (HDSS), Bushbuckridge, Mpumalanga Province South Africa. Klerksdorp was selected because it is peri-urban and has a higher TB prevalence compared with Agincourt which is rural (6). These sites would give an indication of progress

towards past TB control effort and the possible need to scale up or change TB control and prevention strategies.

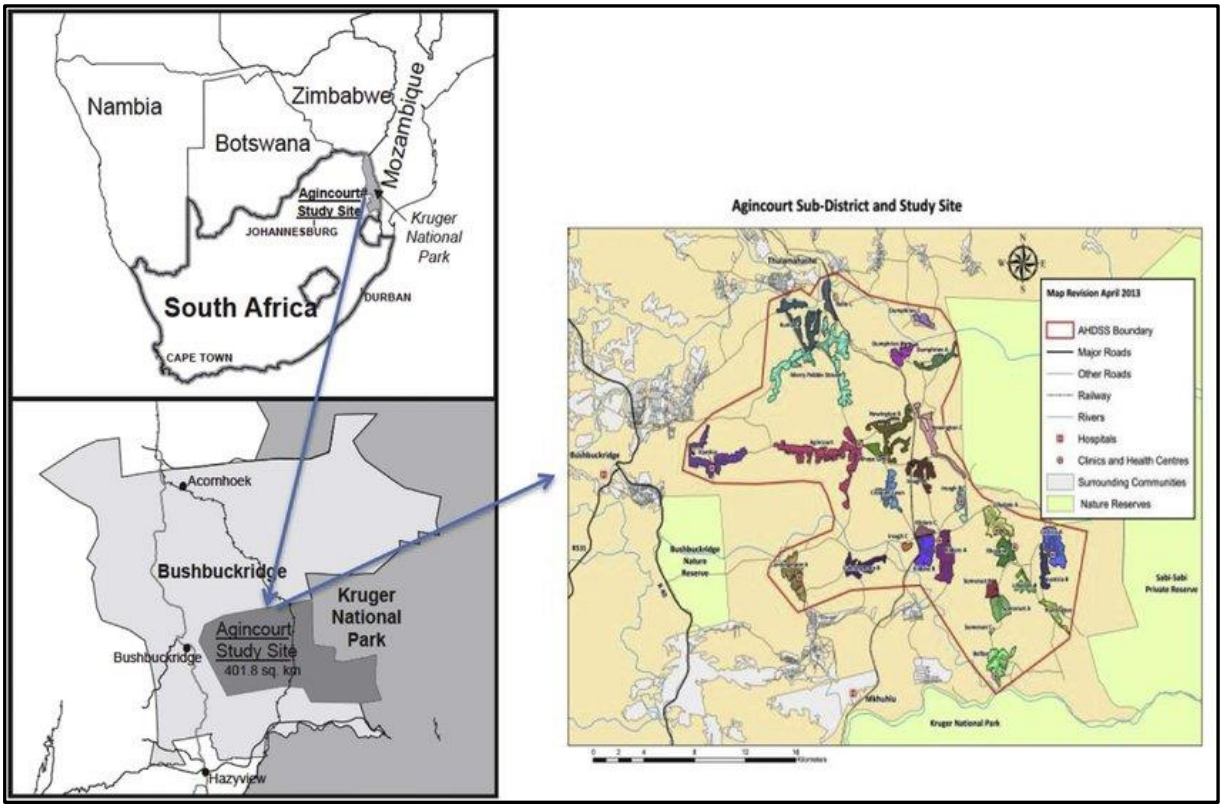


Figure 2.1 Map showing Agincourt research site in Mpumalanga Province, South Africa (obtained from Anstey et al, 2018)

Agincourt is the Wits University and South African Medical Research Councils Rural Research Study Site (MRC/Wits-Agincourt Unit). It is located in Bushbuckridge sub-district, Mpumalanga province and has an area of 420 Km² (46) (Figure 2.1). It was established in 1992 with a research focus on adult health, aging, migration, and health (46). The site is made up of 31 research villages with a population of 116, 549 people who live in 22, 721 households (47, 48). Residents here are largely made of immigrants who were refugees displaced during the Mozambican civil war. Estimated HIV incidence for people aged 15 years and above is 19.4% and HIV/TB is the highest cause of mortality. The research site has two health centers and six clinics (46, 49).

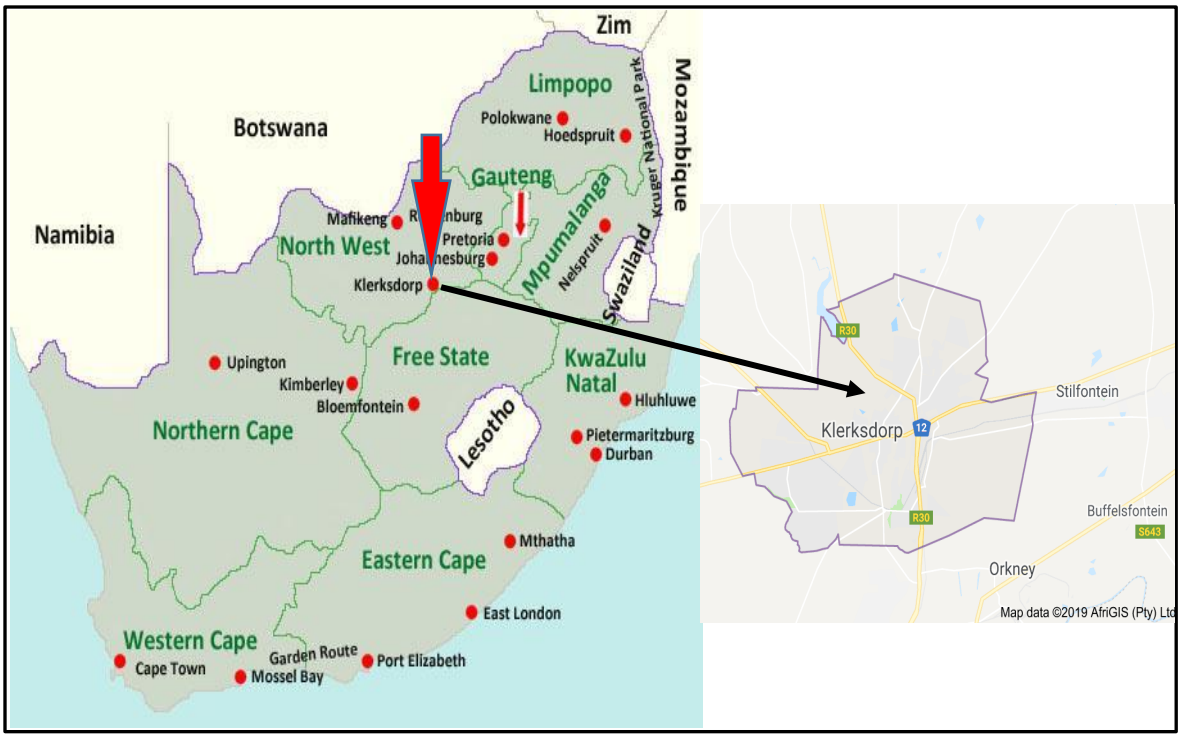


Figure 2.2 Map of South Africa showing the location of Klerksdorp in the Northwest Province (Obtained from https://www.google.co.za/maps/place/Klerksdorp/ & https://showme.co.za/facts-about-south-africa/the-maps-of-south-africa/)

Klerksdorp is the largest city in the Northwest Province of South Africa. It is approximately 164 km Southwest of Johannesburg and covers an area of about 3 625km² (50). It has an estimated population of 417, 282 inhabitants with 92% of these, resident in urban areas (mining villages incorporated in the urban areas) (50) (Figure 2.2). Twenty-nine percent of its population is less than 15 years, 15 – 64 age group account for 66% of the population and the elderly (≥65 years) account for 5% of the population (50). The main economic activities revolve around mining, agriculture, transportation, manufacturing and construction (50). There is one major public hospital – The Klerksdorp Tshepong Hospital Complex which receives referrals from the entire Northwest Province with 16 health clinics. Tuberculosis incidence in Klerksdorp was estimated at 560-764/100, 000 population in 2014 (6).

2.1.4 Study population

Residents of Klerksdorp - Northwest Province, and Agincourt Research Site, Bushbuckridge Municipality, Mpumalanga Province South Africa irrespective of age, sex or HIV status.

2.1.5 Sampling

A sample size of 1500 (750 at each location) was projected for the study. This allowed estimation of 10% risk of tuberculosis infection and 5% risk of illness with a 95% confidence interval (CI) and 5% precision in the sampled population. A Stratified two-stage cluster sampling technic was used to select 300 households, overall, 150 in each study location but new 50 households were enrolled each year in January and February. The plan was to enroll five residents per household for the study. The first sampling step identified 2 – 3 township extensions in Klerksdorp and 2 -3 villages in Agincourt. The households were then randomly selected using a sampling frame provided by Statistics South Africa (STATSA).

Sampled households were visited to ascertain if they met the inclusion criteria and to also obtain informed consent before commencing the study. None consenting households and those not meeting the inclusion criteria were excluded. Sampling and household verification continued till the required households were enrolled (50 households per site per year). Five additional households were enrolled at each site to accommodate withdrawal or loss to follow up of a household from the study. A household was withdrawn from the study if more than 20% of its participants suspended participation in the study.

2.1.6 Inclusion Criteria

Only randomly selected households with at least three residents were included in the study. Household residents were to give informed consent to participate, must have lived in Klerksdorp or Agincourt for one year before the commencement of the study, and had no intentions of relocating from their residence. Fifty percent of the households selected were to have children with age below five years as a resident.

2.1.7 Data Collection

Data collection was done by a team comprising a trained data collector and a study nurse. Each household selected for the study was given a unique household identity number and residents of the household who consented were given a unique individual identification number linked to the household. Baseline demographic information, medical history of participants and

household information was collected using an interviewer-administered questionnaire in Redcap.

At the baseline visit, anthropometric measures of all participants were taken- height (cm), weight (kg) and upper arm circumference (cm). HIV negative participants were offered HIV counseling and testing. The test was conducted and the results issued to those who consented. Those who were newly identified as HIV positive were referred for treatment. A Nasal swab was taken from all residents in a household and this was labeled using the individual identity number. Baseline sputum was collected from residents who could produce it and blood sample for CD4 count and viral load from HIV positive participants. These were labeled using the identification number and preserved in a cooler box with ice for onward transportation to the laboratory.

Tuberculin skin testing was conducted at baseline and at a follow-up visit 6 months later by the trained study nurse. Separate consent was obtained before the tuberculin was placed. 0.1ml (5 tuberculin unit) of the Purified Protein Derivative (PPD) was injected into the anterior surface of the forearm, and the induration formed (palpable swelling) was read 48-72 hours later by the nurse using a transparent ruler. Unit of measurement was millimeter (mm). Repeat visits were made to collect follow up information and specimen for a nasal swab, CD4 and viral load.

2.2 Secondary Data Analysis

2.2.1 Study design

A Cross-Sectional study design in the form of a secondary analysis of baseline data was adopted to enable us to determine the prevalence of TST-positive at the two sites and predictors of TST-positivity at the individual and household level.

2.2.2 Data source

Data used for this research was from the PHIRST Study. This was baseline data of participants and households sequentially collected from three sequentially recruited households in 2016, 2017 and 2018.

2.2.3 Study population

Residents of Klerksdorp and Agincourt without consideration of age, sex or HIV status.

2.2.4 Power and Sample Size

Two previous studies in South Africa determined the prevalence of latent tuberculosis infection to be 45% and 34% respectively (5, 13). At an alpha level of 5%, power of 90% and an effect size of 0.11, at least 826 participants (413 at each location) was required for the secondary analysis of the PHIRST cohort study data to test the stated hypothesis.

2.2.5 Description of Variables

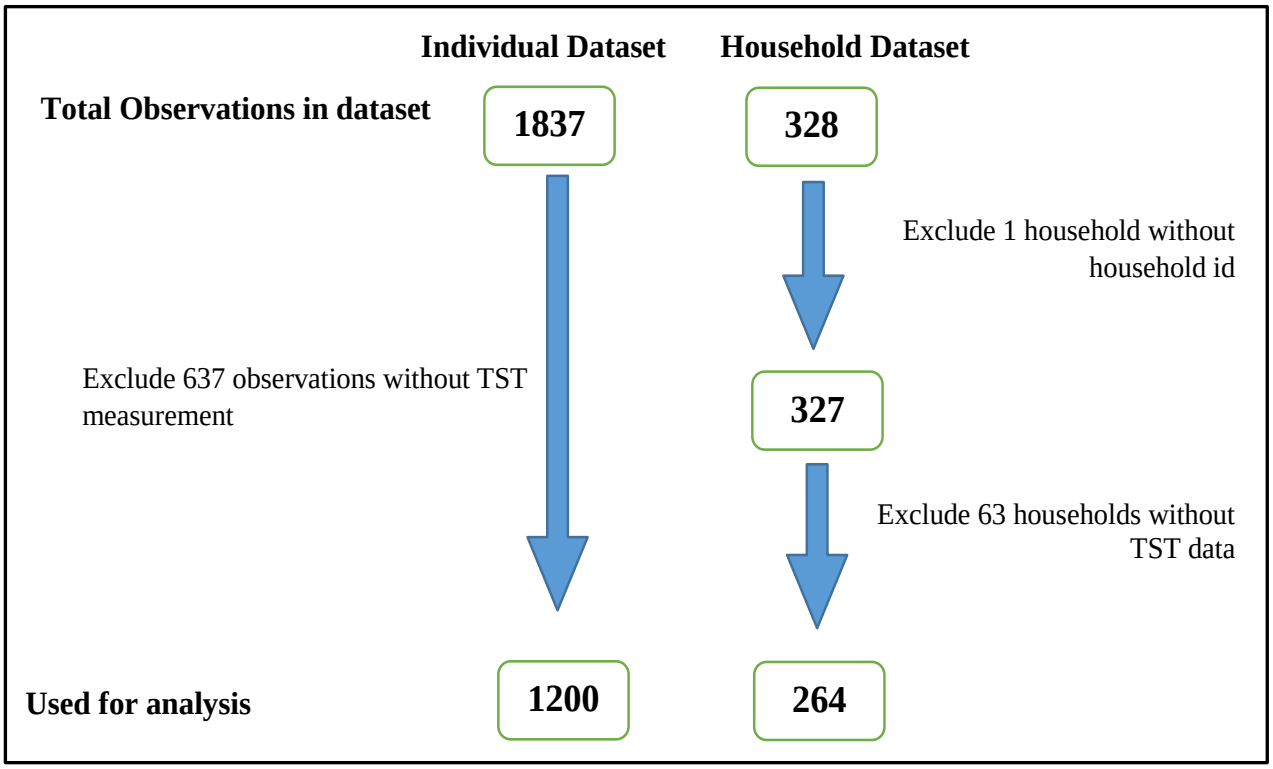


Figure 2.3 Schema for selection of study participants from the data set for analysis

Variables used for analysis were obtained from the baseline data of 1837 study participants from 328 households selected for the study (figure 2.3). They were categorized into 10 individual-level variables and 13 household-level variables. Of the 1837 individuals in the data set, 1200 observations (570 from Agincourt and 630 from Klerksdorp) from 264 households

which had documented tuberculin skin test measurements were used for analysis. The observations and households excluded from the analysis did not have information on TST measurement which is the primary outcome of the study. These individuals were either not eligible for TST, did not provide informed consent or were not available for reading of TST measurements after two visits. A household was dropped from analysis if no resident had a TST result. The individual-level variables were grouped into demographic, anthropometric and clinical variables. The primary outcome variable was TST result derived from the TST measurement. A TST measurement ≥10mm (≥5mm in HIV positive participants) was regarded as TST-positive result and a TST measurement of <10mm was regarded as TST-negative (41, 51).

Demographic variables are study site (Agincourt /Klerksdorp), year of study (2016, 2017 and 2018), sex (male/female), age at consent in years measured on a continuous and a categorical scale (<5, 5-10, 11-15, 16-20, 21-25, 26-30, 31-35, 36-40, 41-45, >45). Anthropometric variable included body mass index (BMI) and upper arm circumference. BMI was obtained by finding the ratio of weight in kilogram to height square in meter (Wt. (kg)/ Ht^2 (m)). It was grouped into four (Underweight <18.5, Normal 18.5- <25, Overweight 25- <30, Obese ≥30) and measured in kilogram per meter squared (kgM^{-2}). Upper arm circumference was measured in millimeters on a continuous scale for participants aged ≤15 years.

Clinical variables are TST measurement, TST result (positive/negative), HIV status (negative/positive/unknown), present cough (Yes/No), CD4 (measured on a continuous scale), and Past TB Treatment (Yes/No). The Social variable is smoking in a household (binary-Yes/No).

Thirteen variables from 264 households (Agincourt 125, Klerksdorp 139) were analyzed. The variables analyzed was presence of TST positive in the households, type of house (formal house built by professional, formal house built by self, informal, other - traditional or flat), age of house, number of people resident in household, number of windows, number of bedrooms, number of toilets inside the house, number of people resident in household measured on continuous scale and categorised into <5, 5, >5, smoking within household (yes/no), number in

household who smoke, number in household who smoke inside, fuel used for cooking (electricity, paraffin, gas, wood, other) and fuel used for heating the house (electricity, paraffin, gas, wood, coal/ imbawula, none).

Two variables, crowding index and wealth index, were generated to describe household parameters which were absent in the data set.

2.2.6 Generation of new variables

2.2.6.1 Crowding Index

Crowding index gives an estimate of overcrowding within a household and was determined as the ratio of the number of residents in a household to the number of bedrooms in the household. Households with 0 as the number of rooms were allocated 1 to indicate where the residents sleep. The formula is shown below:

$$\textbf{Crowding Index} = \frac{Number\ of\ residents\ in\ a\ household}{Number\ of\ bedrooms\ in\ the\ household} \quad (52, 53)$$

2.2.6.2 Wealth Index

Wealth index was generated as a proxy for socio-economic status which was absent from the two data sets. It was calculated from six household variables using the principal component analysis method and presented as quintiles (54, 55). The household variables were the type of house, number of residents in a household, number of bedrooms in a household (Crowding Index), presence of toilet inside the house, cooking fuel and heating fuel.

Steps taken to generate wealth index quintiles for the households are outlined below:

Step 1: Categorical variables type of house, cooking fuel, and heating fuel were recoded to 0/1.

a. Formal house built by professional or self was coded 1. Informal dwelling/ mud/thatch was coded 0.
b. Cooking fuel was coded 1 if it was electricity, gas or paraffin. Wood or other types of cooking fuel was coded 0.
c. Heating fuel was coded 1 if it was electricity, gas, coal or paraffin. Wood, other kinds of heating or no heating was coded 0.

Step 2: Run the principal component analysis for the identified variables – the type of house, cooking fuel, heating fuel, crowding index (derived from the number of rooms and number of residents in a household) and presence of toilet inside the house with weighting to account for the total number of households at each site.

Step 3: Prediction of the wealth index score which is assigned to each household in the dataset.

Step 4: Generation of quintile for the wealth index score with weighting to account for the total number of households at each site.

A bar chart describing household level quintile by site is available as Appendix 4.

2.2.6.3 Smoking within households

Three variables which described smoking within the household (smoking within household (yes/no), number in a household who smoke and number in a household who smoke inside) was merged into one. The new variable was binary. It was coded 1 for a household if any of the variables above is "yes" or >0 and coded 0 if variables above are "no" or <1.

2.2.6.4 The Outcome variable - TST Result

The outcome variable for individual and household level analysis was binary (TST positive or negative). TST result was generated using values of TST measurement to enable determination of TST-positive prevalence at the individual level. It was either positive or negative. A TST positive result was taken as TST measurement ≥10mm or ≥5mm in HIV positive participants and TST measurement <10mm was regarded as TST negative result (31, 41, 51, 56). A cut-off of 10.0 mm was adopted as this is mostly used in TB prevalence studies and will make comparison easier with other studies. Prevalence is also determined using other cut-off points (5.0 mm and 15.0 mm) for comparison. A lower TST cut off for HIV positive participants was to account for anergy which occurs in immunosuppression associated with HIV.

2.3 Data Management

Data for the analysis was received de-identified in STATA format. It consisted of individual-level data and household-level data. Data processing and cleaning were done using STATA 15.0 software (StataCorp LLC College Station, TX). Duplicates, outliers, and inconsistencies

were checked for using the unique household and individual number for each entry. Missing entries were cross-checked with the original data file and inputted if available. An observation or variable was excluded from the analysis if >10% of its entries were missing. Generation of new variables and recoding was done as required for analysis.

2.4 Statistical analysis

2.4.1 Descriptive Analysis

A cross-tabulation of baseline characteristics of variables in the two datasets is presented in a table (Table 1) stratified by site. Continuous variables were described using mean and standard deviation for normally distributed variables; median and interquartile range for non-normally distributed variables. Frequency tables (n) and percentage (%) was used to describe categorical variables. For comparison of differences between two categorical variables, the Pearson Chi-Square test ($\chi 2$) or Fischer's exact test was used. A t-test was used to compare differences in continuous variables that are normally distributed while the Wilcoxon Rank Sum Test was used to compare differences in non-normally distributed continuous variables.

TST measurements are described for values = 0 mm by age and >0 mm by age category and site using a box plot.

2.4.2 Prevalence of Latent Tuberculosis Infection (TST-Positive)

2.4.2.1 Individual-level prevalence of TST-Positive

The first objective was to determine the age-specific prevalence of TST-positive using the tuberculin skin test result at both sites. Prevalence was calculated (with 95% confidence interval) per site and by age group for TST result using the formula below:

$$\text{Prevalence} = \frac{\text{Number of TST+ve}}{\text{Total Number of TST}} * 100$$

Where TST+ve = total participants with a positive TST result,

And Total TST = total number of participants who had TST done.

Three tables were used to describe TST-positive prevalence based on some sociodemographic variables and using different TST cut off for children and adults. A Pearson's Chi-Square test

was used for the comparison of the difference in TST-positive prevalence between the two sites and by some sociodemographic variables. The generated p-value is tabulated and indicates if there is a difference in prevalence between the two sites based on the variables. Three graphs were used to describe the prevalence of TST-positive: by site and age category; HIV status, site and sex and by different TST measurements.

Analysis of the repeat TST testing was excluded as data was only available for 30% of the participants who had a baseline TST and would not have provided a good comparison with the baseline analysis. This lack of data was due to challenges encountered in the administration and reading of TST measurements towards the end of each study year.

2.4.2.2 Household-level prevalence of TST-positive

For the household analysis, households were considered positive if they had at least one household member with a TST-positive result (using 10.0 mm TST cut-off), and negative if no member had a positive TST result. Household-level prevalence was determined with 95% CI based on the number of households considered to be positive or negative at each site. The formula is shown below:

$$\text{Household Prevalence} = \frac{\text{Number of households with +ve TST result}}{\text{Total Number of households with TST result}} * 100$$

A description of TST-positive prevalence at the household level is also presented based on-site, crowding index and smoking within the household. Household prevalence based on wealth quintile (a measure of socioeconomic status) is estimated and presented as a percentage with a positive result (and 95% confidence intervals) by site. Findings and comparison of household-level prevalence of TST-positive are presented in a table. In 2017, there were challenges with administration and collection of TST data which resulted in a fewer number of study participants in the 2017 cohort compared to other years. This limited yearly comparison of derived prevalence.

2.4.3 Annual risk of tuberculosis infection (ARTI)

The second objective was to determine the percentage Annual risk of tuberculosis infection (ARTI (%)). This was estimated (with 95% CI) to give an indication of tuberculosis incidence rate at both sites for the different age groups using the formula below:

$$\text{ARTI} = 1 - (1 - P)^{1/a}$$

Where P = Overall LTBI prevalence from TST and 'a' = mean age in years (57).

Overall TST-positive prevalence was used to determine the overall ARTI with its 95% CI. ARTI and their 95% CI for the sites were calculated using TST-positive prevalence from the respective site. A descriptive summary of ARTI based on site is presented in a table. Line and bar graphs are used to describe median ARTI. These are by age category and site; site and ≤15 years of age; year, sex and site; and by HIV status site and sex.

2.4.4 TST-Positive prevalence and HIV status

The third objective was to compare the prevalence of TST-positive amongst HIV positive and HIV negative participants, controlling for age and sex. Pearson's Chi-Square test was used to determine the difference in proportion between TST-positive and HIV status. Bivariate logistic regression analysis between TST-positive and HIV status was used to determine the unadjusted OR with 95% CI. The adjusted odds ratio of the association between TST-positive and HIV status controlling for age and sex was obtained by conducting a logistic regression analysis of TST-positive and HIV, controlling for age and sex. Marginal estimates of the difference in the prevalence of TST-positive between HIV positive and negative participants were predicted for the two sites.

2.4.5 Predictors of TST-Positive

The fourth objective aimed at determining risk factors for TST-positivity in the studied population. To achieve this, separate logistic models were developed: one for individual-level variables and the other for the household level variables.

2.4.5.1 Logistic analysis for Individual predictors of TST-Positivity

To predict individual-level risk factors for TST-positivity, a logistic model ignoring clustering and a nested multilevel logistic model with random effects at three levels (site, household, and individual) to account for clustering was used. The analysis was done separately for children (age ≤15 years) and adults (age ≥16 years).

A Separate logistic analysis ignoring clustering was done for children and adults based on HIV status to ascertain if there was a difference in predicting factors at the individual level based on HIV status. A different logistic analysis was also done to ascertain if risk factors will differ at Klerksdorp only.

A univariate logistic analysis was conducted to determine the independent association of each predictor variable with the outcome variable - TST result. Variables with $p<0.2$ were included in the multivariate analysis. This was confirmed by including all the variables in the model and dropping insignificant variables serially, starting with variables with the largest p-value till a final model is obtained. After variable selection for the multivariate analysis, some variables (age, past TB history, and HIV status) which are known confounders and have been identified to predict TST-positivity in previous studies were included if they were initially excluded.

Test of interaction between all variables in the final model was done to identify any significant interaction for inclusion in the model. Statistical significance was at $p<0.05$ and the Hosmer-Lemeshow test was used to assess the fitness of the final logistic model.

For the nested multilevel logistic model with random effects at three levels (site, household, and individual), similar steps as above were taken and statistical significance was at $p<0.05$ with their accompanying odds ratio (OR) and 95% CI. This was done separately for adults and children.

2.4.5.2 Logistic Analysis for household predictors of TST-Positivity

To identify household predictors of TST-positivity, a nested multilevel logistic model with random effects at two levels (site and household) was used. This was done for both sites combined and for Klerksdorp only.

Univariate analysis identified the independent association of each household predictor variable with the outcome variable – household TST result. Variables with $p<0.2$ were included in the multivariate analysis. This was confirmed by including all the variables in the model and dropping insignificant variables serially starting with variables with the largest p-value till a final model is obtained. All variables used for generating wealth index was excluded from the analysis as wealth index was included in the analysis. The three variables which describe smoking in a household were merged into one variable and included in the analysis.

Test of interaction between all variables in the final multivariate model was done to identify any significant interaction for inclusion in the model. Statistical significance was at $p<0.05$ with its accompanying odds ratio (OR) and 95% CI. A quadrature check was done to ascertain if the standard 12 integration points were adequate for the model.

2.5 Ethical considerations

Ethics approval was received from the University of the Witwatersrand Human Research Ethics Committee (Medical) before analysis of this data set (Number: M181005). The certificate of ethics clearance is attached (Appendix 1). Ethical approval for the primary study was obtained from the University of the Witwatersrand Human Research Ethics Committee (Medical) (Number: M150808). Permission to analyze the baseline data from the cohort study was obtained from the primary study principal investigator (Appendix 3). The data file was stored in a password protected hard drive to restrict access to only the principal investigator and the supervisors.

This research did not receive funding from any organization and was supported by personal funds provided by the principal investigator. The principal investigator was responsible for the data management, analysis, interpretation, and writing of the first draft of the research report.

The supervisors provided guidance on the writing of the protocol and the research. They further reviewed different drafts of the research report before submission of the final draft.

3 CHAPTER THREE: RESULTS

The results from data analysis are presented in this chapter. It includes a description of baseline characteristics of the variables in the research, the determined prevalence of latent tuberculosis infection by sociodemographic variables, the annual risk of tuberculosis infection described by site and age, and comparison of TST-positive between HIV positive and negative participants controlling for some demographic factors. The risk factors which predict TST-positivity at the individual and household level is identified through logistic and multilevel modeling.

3.1 Baseline Characteristics

Table 3.1 shows the baseline characteristics of the study participants. The total number of individual participants included in this analysis was 1200 who had TST results representing 65.3% of the total cohort recruited to the PHIRST study. Six hundred and thirty (52.5%) were from Klerksdorp. Females accounted for 60.5% (n=726) of the participants at both sites and more participants were seen in 2018 (41.3%) compared to other years of the study. The median age was 14 (IQR: 6, 31) at Agincourt and 17.0 (IQR: 9, 38) at Klerksdorp (P<0.0002). The 6 – 10 age category accounted for the largest proportion of participants at both sites – 23.3% at Agincourt and 19.8% in Klerksdorp. Participants were older in Klerksdorp (18.1%) compared to Agincourt (13.3%) (p=0.0240). The mean upper arm circumference was 18.9 cm (SD: 4.1) at Agincourt and 18.6cm (SD: 4.2) at Klerksdorp. Mean BMI was higher in females (28.3 KgM^{-2}; SD: 7.6) compared to males (23.3 KgM^{-2}; SD: 4.8) and it was similar at both sites - 26.7 KgM^{-2} (SD: 6.6) at Agincourt and 27.0 KgM^{-2} (SD: 7.6) at Klerksdorp. The median TST measurement was 0 mm (IQR: 0, 0) at Agincourt and 0 mm (IQR: 0, 8) at Klerksdorp (p<0.0001). Majority of participants (80.4%) had a negative HIV status while 14.4% had a positive status with no difference between sites. Sixty-two participants (5.2%), however, had unknown HIV status (73% of these were aged ≤20 years). Most of the participants had neither been treated for tuberculosis in the past (99.2% and 95.7%) nor had a cough at the time of visit (98.3% and 94.2%) at Agincourt and Klerksdorp respectively.

At the household level, 264 households were included in the analysis – 125 (47.4%) from Agincourt and 139 (52.6%) from Klerksdorp. Mean age of houses at both sites was similar 16.3

years (SD: 9.9) at Agincourt and 16.6 years (SD: 14.0) at Klerksdorp. The median number of windows in households were, however, different at both sites (p=0.0002). Smoking within households was more in Klerksdorp (48.9%) compared to 14.4% at Agincourt. Overall mean crowding index was 2.6 (SD: 1.6). Comparison of crowding index between sites showed Agincourt had 2.3 (SD: 1.3) and Klerksdorp 2.8 (SD: 1.8) (p= 0.0123). Mean wealth index was -1.16 (SD: 1.21) at Agincourt and 0.52 (SD: 0.82) at Klerksdorp (p<0.0001). More than half of the population at Agincourt (55.5%) belonged the lowest wealth quintile (very poor) while Klerksdorp had a more evenly distributed population across the wealth quintile with only 6.1% in the lowest quintile and a higher wealth index. A graphical description of wealth quintile by site is attached as Appendix 4.

Table 3.1 Baseline characteristics of participants and households in the study

Variable	Agincourt N=570 (47.5%)	Klerksdorp N=630 (52.5%)	P-value
TST Measurement (mm)			
Median (IQR)	0 (IQR: 0, 0)	0 (IQR: 0, 8)	<0.0001 [a]
Sex			
Male (%)	218 (38.3)	256 (40.6)	0.3978
Female (%)	352 (61.7)	374 (59.4)	0.3978
Year			
2016	267 (46.8)	162 (25.7)	<0.0001
2017	65 (11.4)	210 (33.3)	<0.0001
2018	238 (41.8)	258 (41.0)	0.7781
Age (years)			
0 – 5 (%)	104 (18.5)	75 (11.9)	0.0021
6 - 10 (%)	133 (23.3)	125 (19.8)	0.1414
11 – 15 (%)	74 (13.0)	87 (13.8)	0.6746
16 – 20 (%)	57 (10.0)	67 (10.6)	0.7182
21 – 25 (%)	22 (3.9)	39 (6.2)	0.0664
26 – 30 (%)	37 (6.5)	30 (4.8)	0.1926
31 – 35 (%)	22 (3.9)	36 (5.7)	0.1347
36 – 40 (%)	25 (4.4)	30 (4.7)	0.7558
41 – 45 (%)	20 (3.5)	27 (4.3)	0.4884
>45 (%)	76 (13.3)	114 (18.1)	0.0240
Median (IQR)	14 (IQR: 6, 31)	17 (IQR: 9, 38)	0.0002 [a]
Upper Arm Circumference (cm)			
Mean (SD)	18.9 (4.1)	18.6 (4.2)	0.3905 [b]
Body Mass Index (KgM^{-2})*			

Underweight (%)	11 (4.6)	25 (7.8)	0.1268
Normal (%)	99 (41.6)	122 (38.2)	0.4237
Overweight (%)	59 (24.8)	67 (21.0)	0.2906
Obese (%)	69 (29.0)	105 (32.9)	0.3229
Mean (SD)	26.7 (6.6)	27.0 (7.6)	0.5587[b]
HIV Status			
Negative (%)	467 (81.9)	497 (79.0)	0.2041
Positive (%)	82 (14.4)	91 (14.5)	0.9680
Unknown (%)	21 (3.7)	41 (6.5)	0.0269
CD4+ (cells/ml)			
Mean (SD)	660.4 (357.3)	658.3 (320.2)	0.9877[b]
Median (IQR)	702.5 (IQR: 423, 899)	646 (IQR:410, 903)	0.8208[a]
Min/Max	30, 1463	214, 1126	-
Past TB History			
No (%)	531 (99.2)	574 (95.7)	0.0002
Yes (%)	4 (0.8)	26 (4.3)	
Present Cough			
No (%)	527 (98.3)	570 (94.2)	0.0003
Yes (%)	9 (1.7)	35 (5.7)	
HOUSEHOLD VARIABLES	**N=125 (47.4%)**	**N=139 (52.6%)**	**P-Value**
Age of House			
Mean (SD)	16.3 (9.9)	16.6 (14.0)	0.8137[b]
Number of Windows			
Median (IQR)	7 (IQR: 4, 10)	5 (IQR: 4, 5)	0.0002[a]
Smoking in Household:			
No (%)	107 (85.6)	71 (51.1)	<0.0001
Yes (%)	18 (14.4)	68 (48.9)	
Crowding Index			
Mean (SD)	2.3 (1.3)	2.8 (1.8)	0.0123[b]
Wealth Quintile			
Very poor	66 (55.5)	8 (6.1)	<0.0001
Poor	25 (21.0)	23 (17.5)	0.8484
Rich	16 (13.5)	35 (26.7)	0.0093
Very Rich	9 (7.6)	35 (26.7)	0.0001
Wealthy	3 (2.5)	30 (22.9)	0.0210[d]
Mean (SD)	-1.16 (1.21)	0.52 (0.82)	<0.0001[b]

[a]Ranksum test, [b]T-Test**,** [d]Fisher's exact [c]See appendix 4 for a graphical description of wealth index by site. Median is reported for non-normally distributed continuous variables and mean for normally distributed variables. *****Underweight (<18.5), Normal (18.5- <25), Overweight (25- <30), Obese (≥30) BMI reported for those aged ≥16 years. Upper arm circumference reported for those aged ≤15 years.

CD4 is reported for only HIV positive participants. IQR=Interquartile range

3.2 Prevalence of TST-Positive

3.2.1 Distribution of TST measurements

Prevalence of TST-positive was determined using TST measurements in millimeter (mm). Eight hundred and nineteen participants (68.3%) had TST measurement of 0 mm. Distribution of TST measurement of 0 mm by age shows most of these participants were ≤15 years of age (Appendix 5). Median TST measurement in Agincourt and Klerksdorp was 0 mm (IQR: 0, 0) and 0 mm (IQR: 0, 8) respectively. Considering TST measurements >0 mm, median TST measurement at Agincourt was 10 mm (IQR: 5, 18) and Klerksdorp 12 mm (IQR: 5, 16) (p=0.4860). Median TST measurement >0 mm was similar between the sexes. Females had – 12 mm (IQR: 5, 17) and males had 10 mm (IQR: 5, 16) (p=0.3526). Distribution of median TST measurements >0 mm by age category and site is shown in figure 3.1 and 3.2. Figure 3.1 suggests clear digit preference at Agincourt. And 8mm appeared to be less likely selected in Klerksdorp.

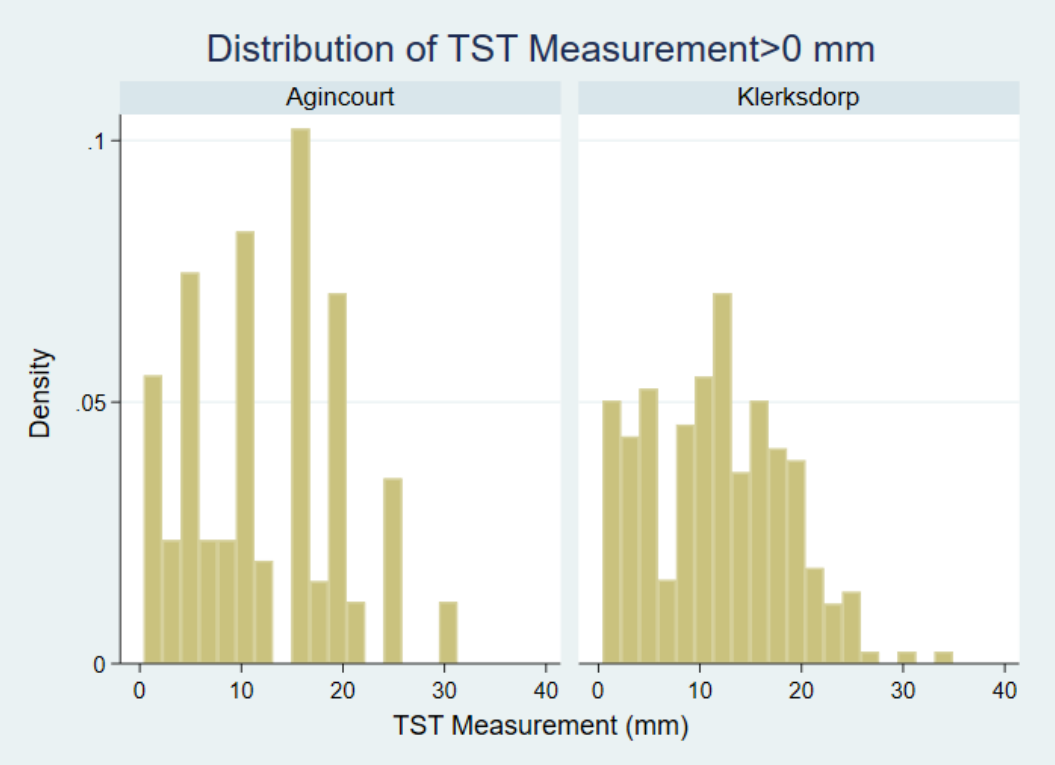


Figure 3.1 Distribution of TST measurements greater than zero by site

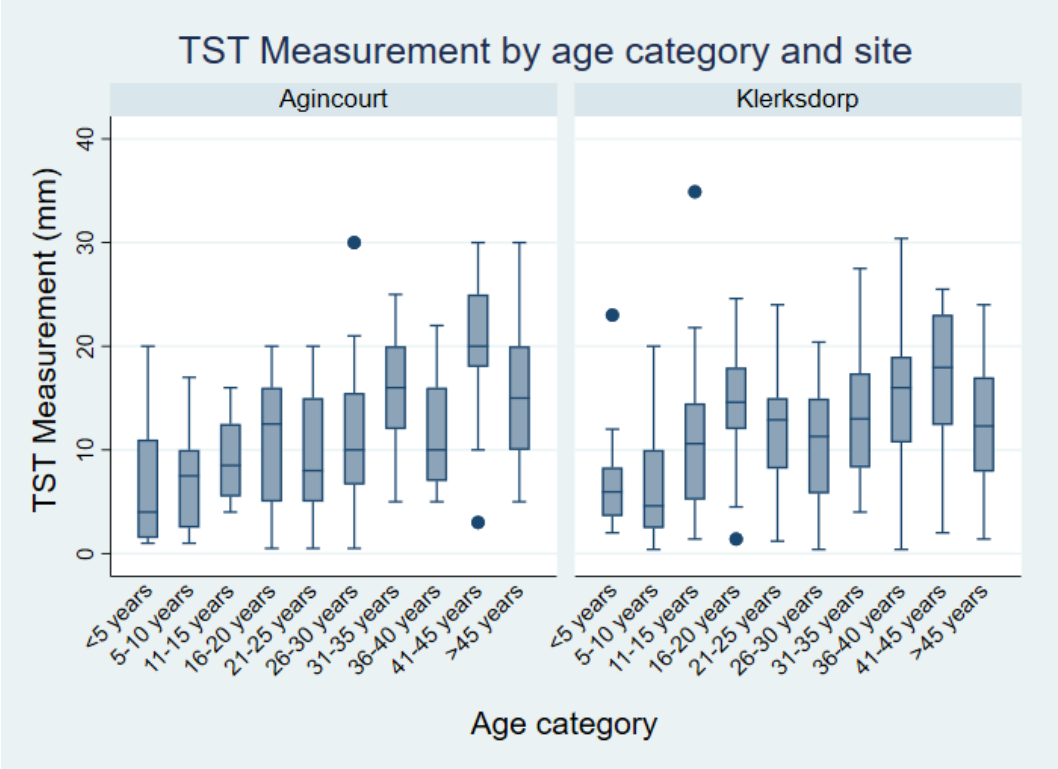


Figure 3.2 TST measurements greater than zero by age category and site

TST measurement ≥2 mm in TST negative participants

One hundred and twelve participants (9.3%) with a negative baseline TST result (using TST cut off of ≥10 mm, ≥5 mm for HIV positive participants) had a TST measurement ≥2 mm with a median TST measurement of 5 mm (IQR: 4, 7). Seventy-four (66%) of these participants were from Klerksdorp and 73.2% from the 2018 participants. Sixty-four (51%) were females, participants aged ≤15 years accounted for 47.3% and those aged >45 years accounted for 20.5%. HIV positive participants with negative baseline TST ≥2 mm accounted for 10.7% (range of TST measurement 2 – 5 mm, median 4 (IQR: 3, 4).

3.2.2 TST-Positive prevalence overall and by site

The overall prevalence of TST-positive was 20.5% (95% CI: 18.3, 22.9) using ≥10 mm TST cut off. Considering the two sites, overall prevalence of latent TB infection was 24.3% (95% CI: 21.1, 27.8) at Klerksdorp and 16.3% (95% CI: 13.5, 19.6) at Agincourt p=0.001 (Table 3.2). There is a significant difference in prevalence in 2016 and 2017 between the two sites (p<0.001). In 2018, there is no difference in prevalence between sites. Overall prevalence was

19.4% in males and 21.2% in females (P= 0.450). Males at Agincourt had a prevalence of 14.2% in comparison with those at Klerksdorp which was 23.8% (P=0.008). Prevalence in females was 17.6% at Agincourt and 24.6% at Klerksdorp (P=0.021).

Table 3.2 Comparison of TST-Positive prevalence between sites and by some variables using TST measurement of ≥10 mm (≥5 mm if HIV +ve) as a positive test

Variable	Agincourt Prevalence (95% CI)	Klerksdorp Prevalence (95% CI)	P-value
Site	16.3 (13.5, 19.6)	24.3 (21.1, 27.8)	0.001
Year			
2016	0.7 (0.2, 3.0)	30.2 (23.6, 37.8)	<0.001
2017	76.9 (65.1, 85.6)	23.3 (18.1, 29.6)	<0.001
2018	17.2 (12.9, 22.6)	21.3 (16.7, 26.8)	0.249
Sex**:**			
Male	14.2 (10.2, 19.5)	23.8 (19.0, 29.4)	0.008
Female	17.6 (14.0, 22.0)	24.6 (20.5, 29.2)	0.021
Age Cat (yrs.)			
<5	1.9 (0.5, 7.4)	5.3 (2.0, 13.5)	0.211
5 – 10	7.5 (4.1, 13.4)	7.2 (3.8, 13.3)	0.922
11 – 15	9.5 (4.6, 18.6)	20.7 (13.4, 30.5)	0.050
16 – 20	14.0 (7.1, 25.8)	29.9 (20.1, 41.9)	0.036
21 – 25	22.7 (9.6, 45.0)	38.5 (24.5, 54.6)	0.209
26 – 30	21.6 (11.1, 37.9)	30.0 (16.2, 48.7)	0.433
31 – 35	40.9 (22.5, 62.3)	33.3 (19.8, 50.3)	0.560
36 – 40	28.0 (13.7, 48.7)	46.7 (29.7, 64.5)	0.156
41 – 45	45.0 (24.9, 66.9)	51.9 (33.3, 69.9)	0.642
>45	36.8 (26.7, 48.3)	33.3 (25.3, 42.5)	0.619
Body Mass Index (KgM^{-2})#			
Underweight	45.5 (19.3, 74.4)	20.0 (8.4, 40.5)	0.116
Normal	27.3 (19.4, 36.9)	37.7 (29.5, 46.7)	0.101
Overweight	30.5 (20.0, 43.5)	35.8 (25.2, 48.0)	0.528
Obese	34.8 (24.4, 46.8)	31.4 (23.2, 41.0)	0.645
HIV Status			
Negative	14.6 (11.6, 18.1)	24.1 (20.6, 28.1)	<0.001
Positive	28.0 (19.4, 38.8)	28.6 (20.2, 38.7)	0.939
Unknown	9.5 (2.3, 31.9)	17.1 (8.3, 31.9)	0.425
Past TB History			
No (%)	17.3 (14.3, 20.8)	22.8 (19.6, 26.4)	0.023
Yes (%)	25.0 (2.4, 82.0)	34.6 (18.8, 54.7)	0.704
Present Cough			
No (%)	17.5 (14.4, 20.9)	23.0 (19.7, 26.6)	0.023
Yes (%)	11.1 (1.4, 53.2)	28.6 (16.0, 45.7)	0.281

#Underweight (<18.5), Normal (18.5- <25), Overweight (25- <30), Obese (≥30)

Prevalence was generally noted to increase with an increase in age category but with differences in age groups. This was seen at both sites with a much steeper increase at Klerksdorp with dips in two age categories at Agincourt and one in Klerksdorp. It increased up to a maximum seen at the 41-45 age category before declining at the >45 age category (Table 3.2, Figure 3.3). A smoothed version of Figure 3.3 which clearly shows a difference in prevalence between the two sites is attached as Appendix 6. Prevalence across the different age categories at both sites was variable with the range in Agincourt being 1.9% in the <5 age group to 45% in the 41-45 age group. The range in Klerksdorp was 5.3% in the <5 age group and 51.9% in the 41-45 age group.

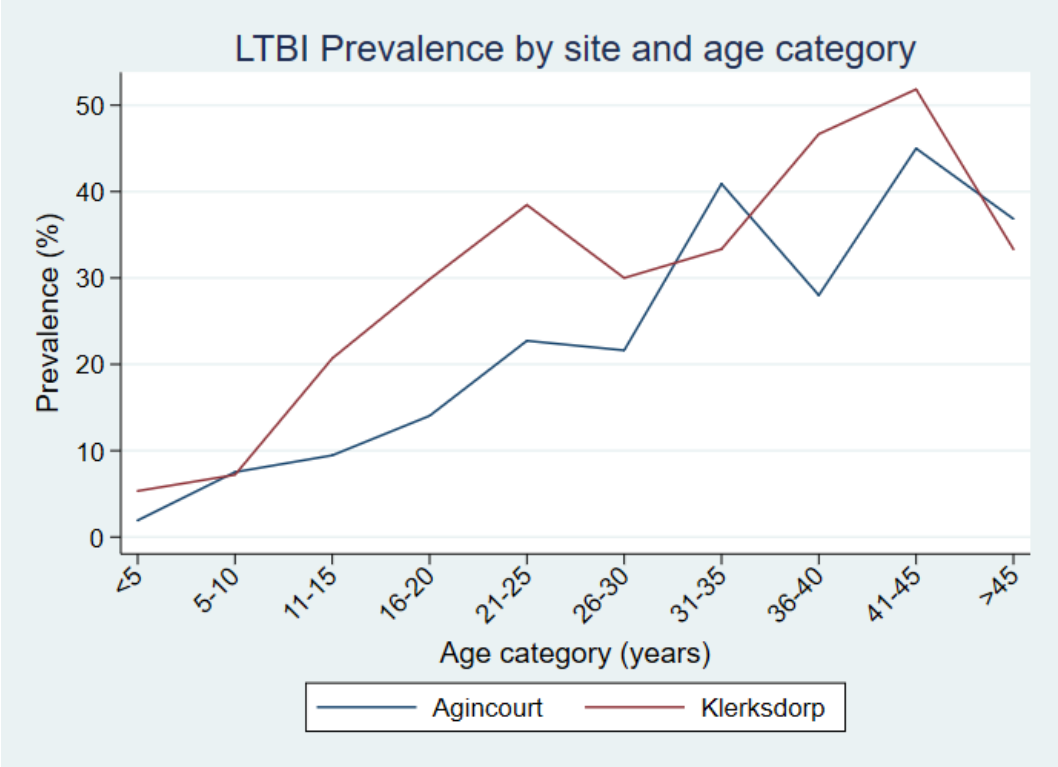


Figure 3.3 TST-positive prevalence by site and age category

Prevalence of TST-positive overall was 19.5% amongst HIV negative participants, 28.3% amongst HIV positive participants and 14.5% in those with unknown status. Between sites, TST-positive prevalence amongst HIV negative was 24.1% at Klerksdorp compared with Agincourt which was 14.6% (P<0.001). The prevalence amongst HIV positive participants at both sites was similar- 28.0% at Agincourt and 28.6% at Klerksdorp (P=0.939) (Table 3.2).

Within site, the prevalence was statistically different between HIV positive and negative participants at Agincourt (p=0.0025) but not in Klerksdorp (p=0.3689). TST-positive prevalence by HIV status, site and sex are described in Figure 3.4. It shows HIV positive males at Agincourt have a higher prevalence compared to females. A small number of participants in this study had a past history of tuberculosis for which they were treated (30/1135) and 10 (33.3%) of these had a positive tuberculin skin test. At the site level, TST-positive prevalence was higher amongst those with past TB history than in those with no TB history and higher at Klerksdorp (Table 3.2). TST-positive prevalence between the two sites by past TB treatment and history of cough is further described in Table 3.2.

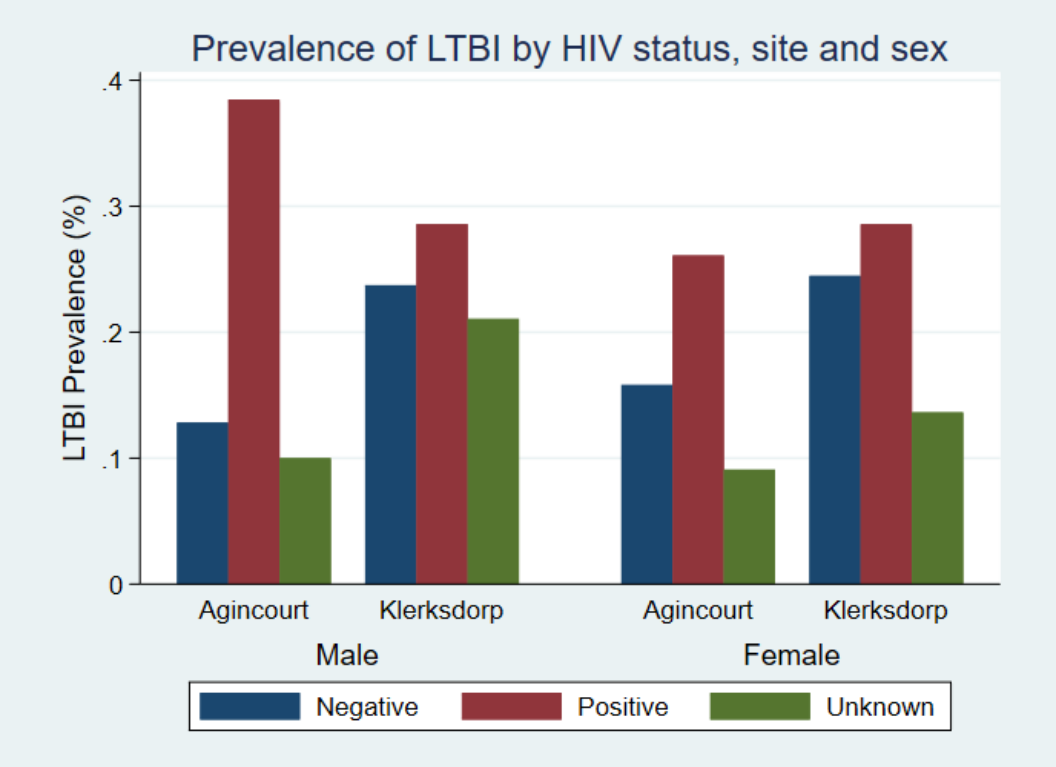


Figure 3.4 A bar graph of TST-positive prevalence by HIV status, site, and sex

3.2.3 TST-Positive prevalence using different TST cut off

The overall TST-positive prevalence is described using a TST cut off of 5 mm, 10 mm and 15 mm for all the participants (Table 3.3). The prevalence is presented at a yearly interval up to ten years of age and thereafter in categories. Prevalence is highest at the 5 mm cut off followed by 10 mm. It is observed to increase and decrease alternately at the yearly age interval up to 10

years of age with the highest prevalence seen at 6 years of age. The prevalence thereafter increases sharply with alternate peaks and dips up to a maximum seen at the 41-45 years age category. A decline is seen at the >45 years age category. This trend in prevalence is similar across the three TST cut-offs (Figure 3.5). A smoothed version of Figure 3.5 is attached as Appendix 6.

Table 3.3 TST-Positive Prevalence using different cut-offs of TST Measurement

Age (years)	N (%) N= 1200	Positive TST at 5mm	Positive TST at 10mm	Positive TST at 15mm
		Prevalence (95% CI)	Prevalence (95% CI)	Prevalence (95% CI)
<1	16 (1.3)	12.5 (3.0, 39.8)	6.3 (0.8, 35.1)	-
1	37 (3.1)	-	-	-
2	41 (3.4)	12.2 (5.1, 26.4)	4.9 (1.2, 17.8)	2.4 (0.3, 15.8)
3	48 (4.0)	4.2 (1.0, 15.4)	2.1 (0.3, 13.7)	2.1 (0.2, 13.7)
4	37 (3.1)	13.5 (5.7, 28.9)	5.4 (1.3, 19.6)	-
5	46 (3.8)	6.5 (2.1, 18.6)	6.5 (2.1, 18.6)	-
6	35 (2.9)	22.9 (11.7, 39.8)	17.1 (7.8 (33.6)	8.6 (2.7, 23.8)
7	45 (3.8)	6.7 (2.1, 19.0)	6.7 (2.1, 19.0)	2.2 (0.3, 14.5)
8	37 (3.1)	13.5 (5.7, 28.9)	8.1 (2.6, 22.7)	-
9	49 (4.1)	8.2 (3.1, 20.0)	4.1 (1.0, 15.1)	-
10	46 (3.8)	6.5 (2.1, 18.6)	4.3 (1.1, 16.1)	2.2 (0.3, 14.2)
11 – 15	161 (13.4)	21.1 (15.5, 28.1)	15.5 (10.7, 22.0)	7.5 (4.3, 12.7)
16 – 20	124 (10.3)	25.0 (18.1, 33.4)	22.6 (16.0, 30.8)	15.3 (10.0, 22.8)
21 – 25	61 (5.1)	39.3 (27.9, 52.1)	32.8 (22.1, 45.6)	18.0 (10.2, 29.8)
26 – 30	67 (5.6)	32.8 (22.6, 45.0)	25.4 (16.3, 37.2)	17.9 (10.4, 29.1)
31 – 35	58 (4.8)	44.8 (32.5, 57.8)	36.2 (24.8, 49.4)	27.6 (17.6, 40.5)
36 – 40	55 (4.6)	41.8 (29.5 , 55.3)	38.2 (26.3, 51.7)	29.1 (18.6, 42.5)
41 – 45	47 (3.9)	48.9 (35.0, 63.1)	48.9 (34.9, 63.1)	40.4 (27.3, 55.0)
>45	190 (15.8)	42.6 (35.8, 49.8)	34.7 (28.3, 41.8)	26.8 (21.0, 33.6)

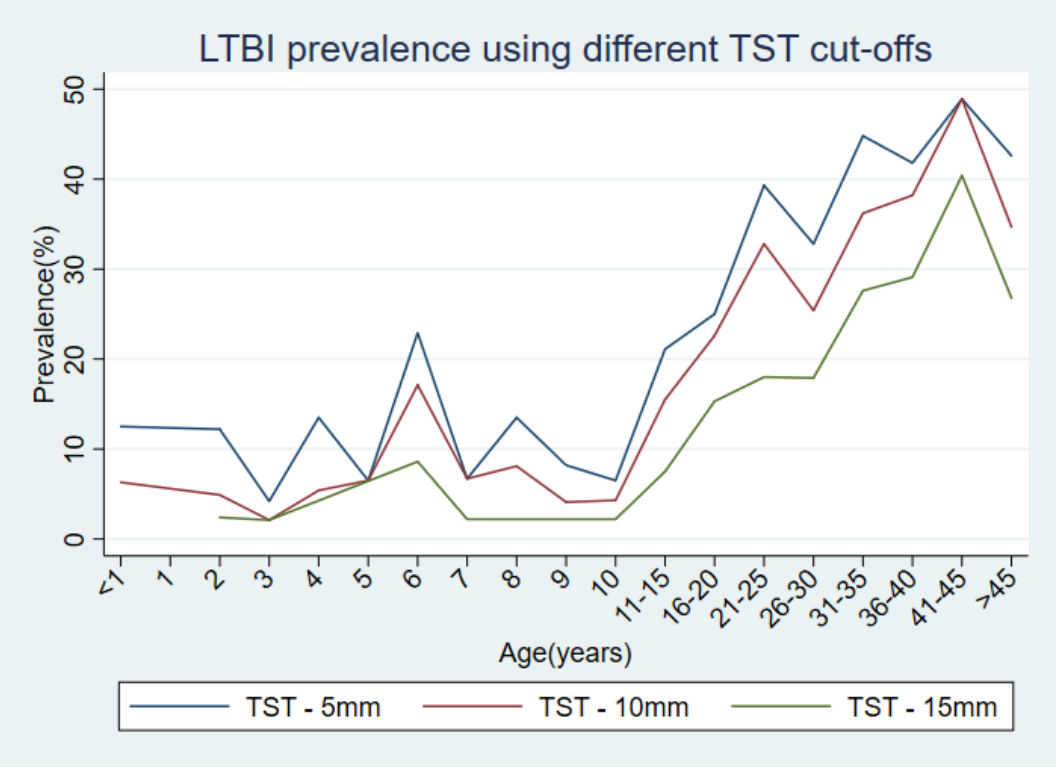


Figure 3.5 TST-positive Prevalence at 5mm, 10mm and 15mm TST cut-off by age

3.2.4 TST-Positive Prevalence and nutritional status

TST-positive prevalence was assessed using two variables commonly used for the description of nutritional status - BMI and upper arm circumference (Table 3.4). Upper arm circumference is used for assessment of children and BMI for adults. Mean upper arm circumference was 18.6 (SD: 4.2) in participants with a negative TST and 20.0 (SD: 4.0) in TST positive participants. No participant with upper arm circumference <15.0 cm had a positive TST. Prevalence was noted to initially increase slightly with an increase in upper arm circumference but decreased again. No significant difference in upper arm circumference was seen between the two sites.

Underweight participants in Agincourt had a higher TST-positive prevalence at 45.5% compared to the other BMI categories in Agincourt and the underweight in Klerksdorp which had the lowest prevalence of 20.0%. Participants with a normal BMI had the highest prevalence of 37.7% in Klerksdorp in comparison to Agincourt where they had the lowest prevalence at 27.3%. No statistically significant difference in BMI category is seen within and between the sites.

Table 3.4 TST-Positive prevalence based on upper arm circumference and BMI

Upper arm Circumference (cm)	Agincourt	Klerksdorp	P-value
	Prevalence (95% CI)	Prevalence (95% CI)	
<11	-	-	-
<12	-	-	-
<13	-	-	-
<14	-	-	-
<15	-	-	-
<16	-	6.12 (1.94, 17.73)	-
<17	4.82 (1.79, 12.29)	7.69 (3.68, 15.39)	0.540
<18	4.92 (2.21, 10.58)	8.96 (5.13, 15.17)	0.2070
<19	5.85 (3.16, 10.57)	7.78 (4.56, 12.98)	0.4797
<20	5.47 (3.05, 9.64)	7.65 (4.65, 12.33)	0.3800
Body mass index (KgM^{-2})			
Underweight	45.5 (19.3, 74.4)	20.0 (8.4, 40.5)	0.116
Normal	27.3 (19.4, 36.9)	37.7 (29.5, 46.7)	0.101
Overweight	30.5 (20.0, 43.5)	35.8 (25.2, 48.0)	0.528
Obese	34.8 (24.4, 46.8)	31.4 (23.2, 41.0)	0.645

Upper arm circumference (age ≤15 years), BMI (age ≥16 years)

3.2.5 Household prevalence of TST-Positive

The overall household prevalence of TST-positive was 54.2%. Household prevalence at Agincourt was 36.0% while Klerksdorp was 70.5% (P<0.0001). The average number of TST positives per household was 0.6 (SD = 1.10) at Agincourt and 1.2 (SD = 1.2) at Klerksdorp (p=0.0001) whereas mean crowding index in households with a positive TST was 2.9 (SD: 2.0) and 2.3 (SD: 0.9) for households without a positive TST (p= 0.0012). Households with positive TST had fewer windows (Median=5; IQR: 4, 6) compared to those with negative TST (Median=6; IQR: 4, 10) (p=0.0086).

TST-positive prevalence was higher in households where residents smoke compared to those where they do not smoke but this difference was marginally significant at Klerksdorp (p=0.060) and not significant at Agincourt (p=0.181). Household prevalence was significantly higher in households where members smoke in Klerksdorp (77.9%) compared to Agincourt at 50.0% (p= 0.019) (Table 3.5)

The overall mean wealth index was significantly higher in TST positive households (Mean=0.01; SD 0.1) compared to the negative households (Mean= -0.6; SD 0.1) (p=0.0002). When stratified by site there was no difference in prevalence at both sites (Agincourt – p=0.2297; Klerksdorp – p=0.6160). Based on wealth quintiles, our analysis suggests there is an overall increase in household TST-positive prevalence with an increase in wealth quintile but this increase was marginally significant (p=0.072). At the site level, TST-positive prevalence increased and decreased alternately with an increase in wealth quintile at Agincourt households. At Klerksdorp, TST-positive prevalence was similar across the quintiles and households in the lowest and 3rd/4th wealth quintile had the largest and lowest prevalence of 75.0% and 65.7% respectively (Table 3.5). Statistically, there is no difference in prevalence within each site based on wealth quintile (Agincourt: p= 0.855, Klerksdorp: p= 0.946) but a significant difference is seen between both sites in the first three wealth quintiles.

Table 3.5 Comparison of household TST-positive prevalence by site

Household Variable	Agincourt		Klerksdorp		P-value
	N (%)	Prevalence* (95% CI)	N (%)	Prevalence (95% CI)	
Site	125 (47.4%)	36.0 (28.0, 44.8)	139 (52.6%)	70.5 (62.3, 77.5)	<0.001
Smoking in Household:					
No	107 (85.6)	33.6 (25.3, 43.2)	71 (51.1)	63.4 (51.5, 73.8)	<0.001
Yes	18 (14.4)	50.0 (27.8, 72.2)	68 (48.9)	77.9 (66.4, 86.3)	0.019
Wealth quintile					
Very poor	66 (55.5)	37.9 (26.9, 50.2)	8 (6.1)	75.0 (35.0, 94.4)	0.044
Poor	25 (21.0)	40.0 (22.7, 60.2)	23 (17.5)	69.6 (47.9, 85.1)	0.040
Rich	16 (13.5)	25.0 (9.3, 51.9)	35 (26.7)	65.7 (48.5, 79.6)	0.007
Very Rich	9 (7.6)	44.4 (16.5, 76.5)	35 (26.7)	65.7 (48.5, 79.6)	0.242
Wealthy	3 (2.5)	33.3 (2.5, 90.5)	30 (22.9)	73.3 (54.6, 86.3)	0.151

*****Prevalence is based on TST measurement ≥10 mm (≥5 mm if HIV +ve)

3.3 Annual risk of TB infection (ARTI)

The overall median annual risk of TB infection was calculated using the overall TST-positive prevalence of 20.5% for all participants. Our results suggest the risk of TB infection is 2.22% annually for all participants (95% CI: 2.1, 2.3) (IQR: 1.6, 3.4). ARTI at each site was calculated using the TST-positive prevalence for each site (Agincourt 16.3%, Klerksdorp 24.3%). The median annual risk of infection of 2.09% (95% CI: 1.96, 2.18) was seen at Agincourt. This was

lower compared with Klerksdorp at 2.29% (95% CI: 2.17, 2.46). Description of median ARTI by year, age category and sex between the two sites is illustrated in Table 3.6.

Incident symptomatic TB during the period of the study could not be determined. This was due to challenges encountered in administration and measurement of TST during the follow-up period towards the end of the year.

Table 3.6 Description of median ARTI by site

Variable	Agincourt	Klerksdorp
	ARTI % (95% CI)	**ARTI % (95% CI)**
Median ARTI	2.09 (1.96, 2.18)	2.29 (2.17, 2.46)
Year		
2016	2.28 (2.09, 2.39)	2.46 (2.29, 2.79)
2017	1.39 (1.33, 1.53)	2.37 (2.06, 2.55)
2018	2.28 (2.09, 2.53)	2.22 (1.95, 2.46)
Age Cat (yrs.)		
<5	6.10 (6.10, 8.65)	8.77 (8.77, 12.65)
5 - 10	2.91 (2.91, 3.19)	3.91 (3.91, 4.33)
11 – 15	2.17 (2.09, 2.18)	2.94 (2.79, 2.94)
16 – 20	1.90 (1.85, 1.90)	2.29 (2.29, 2.37)
21 – 25	1.68 (1.64, 1.70)	1.97 (1.97, 2.01)
26 – 30	1.55 (1.53, 1.57)	1.80 (1.78, 1.83)
31 – 35	1.45 (1.45, 1.48)	1.71 (1.71, 1.73)
36 – 40	1.40 (1.40, 1.41)	1.61 (1.60, 1.63)
41 – 45	1.35 (1.34, 1.36)	1.54 (1.54, 1.55)
>45	1.25 (1.24, 1.27)	1.39 (1.38, 1.41)
Sex:		
Male	2.7 (2.39, 2.91)	2.79 (2.46, 2.94)
Female	1.90 (1.70, 1.96)	2.06 (1.93, 2.29)

3.3.1 ARTI by age and site

Figure 3.6 describes the median ARTI by age category and site for all participants. ARTI is observed to decrease with increase in age category in a similar pattern at both sites and in the overall. The <5 age category has the highest ARTI at both sites and in the overall ARTI, while >45 years age category has the lowest ARTI (Fig 3.2.1, Table 3.6).

Further determination of median ARTI restricted to participants aged <15 years (n=556) at Agincourt and Klerksdorp showed overall ARTI is 3.8% (95% CI: 3.4, 3.8). At Agincourt,

ARTI for those <15 years is 3.2% (95% CI: 2.9, 3.6) and 3.9% (95% CI: 3.9, 4.3) at Klerksdorp. A steep exponential reduction in ARTI in children aged <15 years is seen overall and for the two sites with ARTI of approximately 24% in those aged 1 year at Klerksdorp and 16% at Agincourt (Figure 3.7). This shows the risk of TB infection is highest amongst young children and decreases with increase in age.

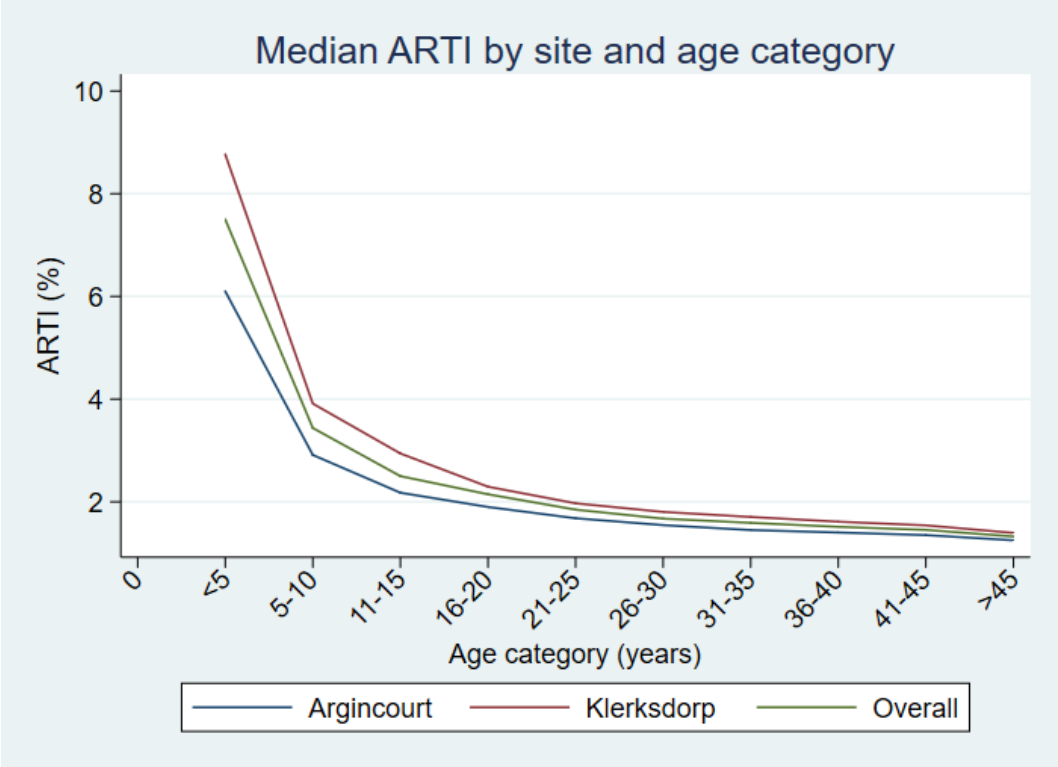


Figure 3.6 Median ARTI by site and age category for all participants
ARTI is based on Prevalence: Agincourt -16.3%, Klerksdorp -24.3%, Overall -20.5%

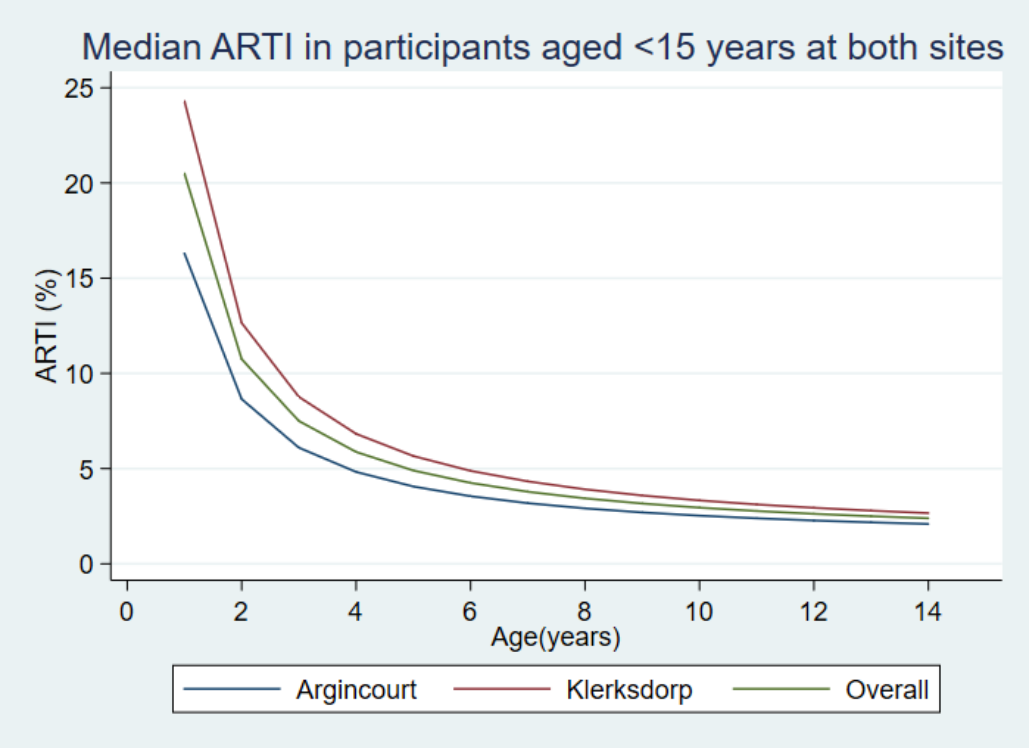


Figure 3.7 Median ARTI by site in participants <15 years of age

Correlation test shows there is a significant negative correlation between ARTI and age with a coefficient of -0.51 at both sites ($p<0.0001$). This corroborates the graphs above.

3.3.2 ARTI by year, sex and site

Figure 3.8 describes ARTI by year, sex and site for all participants. At Agincourt, ARTI is lower in 2017 for males and females compared with the other two years where males have a higher ARTI compared to females. At Klerksdorp, ARTI is slightly higher in 2016 in both males and females when compared with 2017 and 2018.

3.3.3 ARTI by HIV status, site, and sex

ARTI by HIV status, site and sex are illustrated in Figure 3.9. ARTI is noted to be higher amongst participants with a negative and unknown HIV status compared to those with a positive HIV status at both sites. Males with a negative HIV status have similar ARTI at both sites.

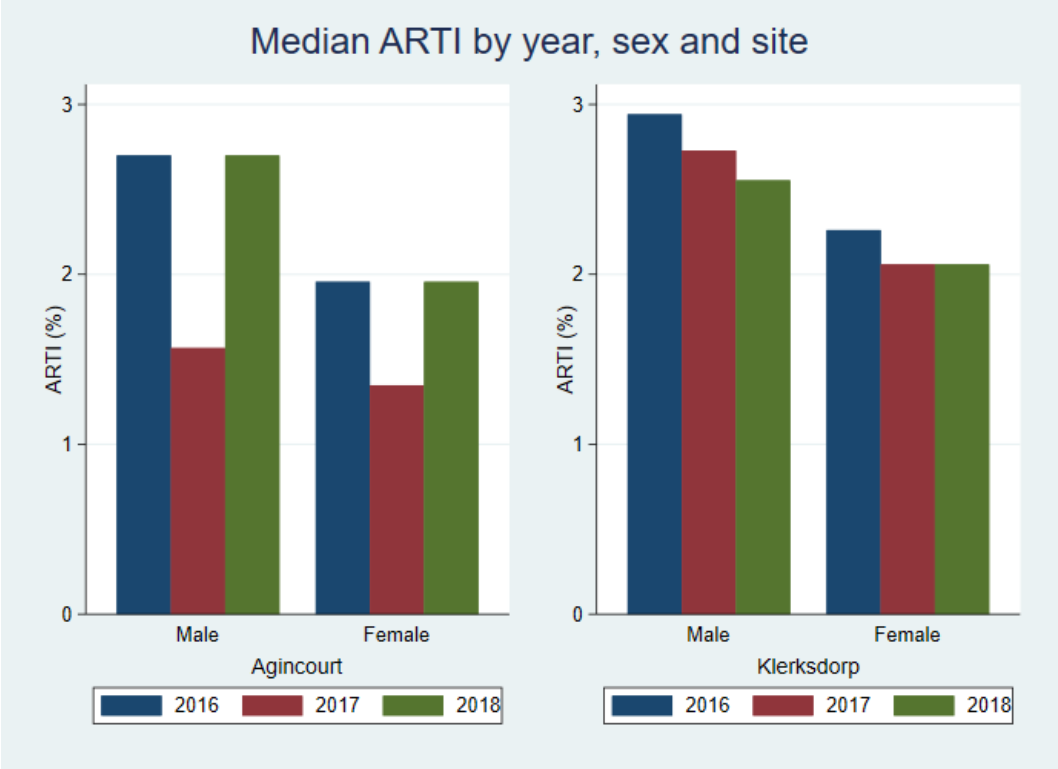


Figure 3.8 Bar graph of median ARTI by year, sex and site

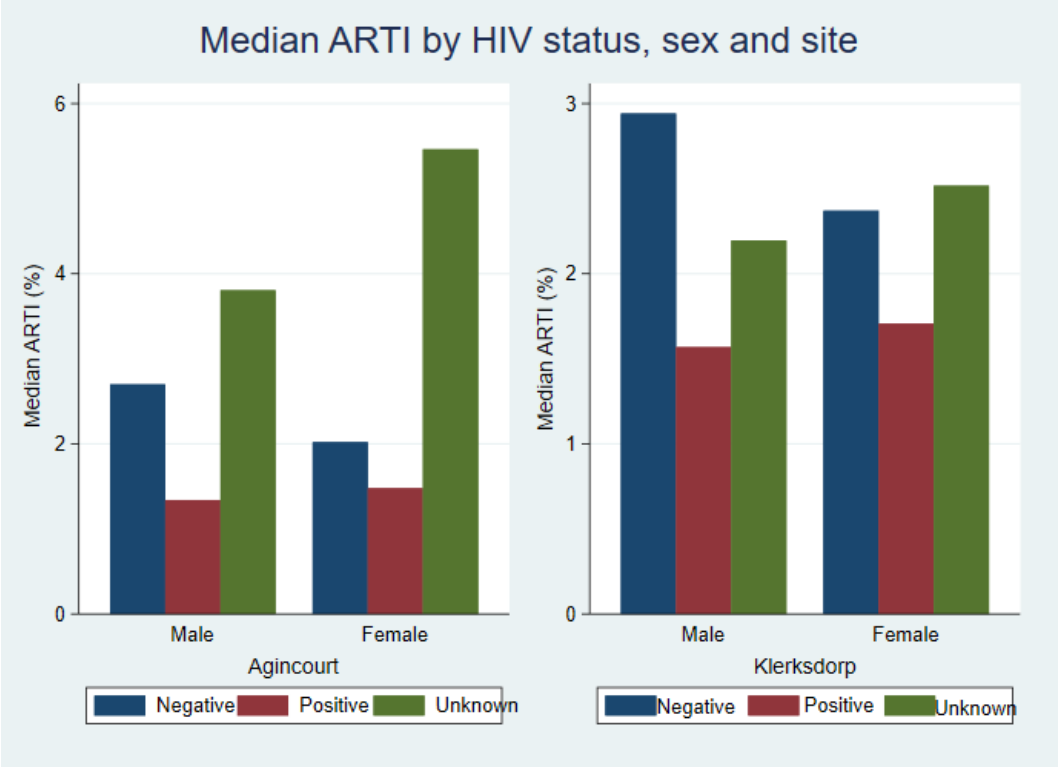


Figure 3.9 Bar graph median ARTI by HIV status, sex, and site

3.4 TST-Positivity based on HIV status

3.4.1 Descriptive Epidemiology

The TST-positive prevalence amongst HIV positive participants is 28.3%, 19.5% in those with a negative HIV status and 14.5% in those with unknown status. Test of association between TST result and HIV status using Pearson's Chi-square comparison test shows there is a significant relationship between the two variables (Chi-square with two degrees of freedom = 8.44, p= 0.015). Further assessing for odds shows the unadjusted odds of having a positive TST for an HIV +ve participant using HIV negative as the baseline is 1.63 (95% CI: 1.13, 2.35) p=0.009. This was however different at both sites as Agincourt had an odds of 2.29 (95% CI: 1.32, 3.95) p=0.003 and Klerksdorp 1.26 (95% CI: 0.76, 2.07) p=0.370.

3.4.2 Controlling for age and sex

After controlling for age and sex the odds of a positive TST in HIV positive participants at both sites compared to those with a negative or unknown status was 1.12 (95% CI: 0.76, 1.65) indicating that a positive HIV status did not significantly lead to a positive TST result (p=0.567) (Agincourt: OR=1.42; 95% CI: 0.79, 2.56; p=0.238, Klerksdorp: OR=0.94; 95% CI: 0.55, 1.58; p=0.801).

Report on Margins after controlling for age and sex

The mean marginal difference in TST positivity between those who were HIV negative and HIV positive after controlling for age and sex is 0.02. HIV negative participants had a margin value of 0.18 (p<0.001) and HIV positive participants had 0.20 (p<0.001). A test of the difference between HIV positive and negative/unknown indicates predictive margins did not differ between the two (p=0.5670).

3.5 Risk factors of TST-Positivity

Risk factors for a positive TST were assessed at two levels: individual and household risk factors.

3.5.1 Individual risk factors

In Tables 3.7 and 3.8, the odds with 95% CI and p-values for the univariate and multivariate logistic model for prediction of TST-positive risk factors are presented by age. The multivariate model identified variables that were significant independent predictors of a positive TST after controlling for other variables.

3.5.1.1 Logistic modeling for prediction of individual risk factors of TST-Positive

Individual Risk factors for participants aged ≤15 years

Table 3.7 Logistic model of individual-level factors which predict TST-Positive ignoring clustering within households (Participants ≤15 years of age)

Variable	Univariate analysis		Multivariate analysis	
	Odds (95% CI)	P-Value	Odds (95% CI)	P-Value
Site				
Agincourt	1		1	
Klerksdorp	1.86 (1.03 3.37)	0.041	1.84 (0.99, 3.38)	0.051
Sex				
Male	1		1	
Female	0.76 (0.42, 1.36)	0.351		
Age (yrs)	1.14 (1.06, 1.22)	0.001	1.12 (1.02, 1.22)	0.019
Upper-arm Circum. (cm)	1.08 (1.01, 1.15)	0.020	1.02 (0.94, 1.10)	0.697
HIV Status:				
Negative	1		1	
Positive	3.17 (1.001, 10.03)	0.050	2.53 (0.77, 8.32)	0.126
Unknown	0.32 (0.04, 2.37)	0.263	0.40 (0.05, 3.05)	0.375
Present Cough				
No	1		1	
Yes	0.61 (0.08, 4.69)	0.636		

Univariate and multivariate analysis for predictors of TST-positivity in participants aged ≤15 years is described in Table 3.7. In the univariate analysis, the odds of being TST-positive was 86% higher for participants resident in Klerksdorp (OR=1.86; 95% CI: 1.03, 3.37; p=0.041) compared to those in Agincourt and this was significant. There was a 14% increase in the odds of being TST-positive for a unit increase in age (OR=1.14; 95% CI: 1.06, 1.22; p=0.001) and 8% increase for a unit increase in upper arm circumference (OR=1.08; 95% CI: 1.01, 1.15; p=0.020). For HIV positive participants, the odds of a positive TST result increased by 3.17

when compared with those who had a negative HIV status (95% CI: 1.001, 10.03) but this was marginally significant (p=0.050). Sex and present history of cough were not significant at the 5% significance level though they showed reduced odds and no participant had a past history of tuberculosis.

In the multivariate analysis, only age was statistically significant, indicating that a unit increase in age increased the odds of being TST-positive by 12% (OR=1.12; 95% CI: 1.02, 1.22; =0.019). Although site and a positive HIV status were included in the model, they were not significant at the 5% level of significance. The model fitness was assessed with the Hosmer-Lemeshow test and had a test statistic of 9.00, p=0.3421.

Individual Risk factors for participants aged ≥16 years

Table 3.8 Logistic model of individual-level factors which predict TST-Positive ignoring clustering within households (Participants ≥16 years of age)

Variable	Univariate analysis		Multivariate analysis	
	Odds (95% CI)	P-Value	Odds (95% CI)	P-Value
Site				
Agincourt	1	-	1	
Klerksdorp	1.38 (0.97, 1.96)	0.070	1.37 (0.96, 1.95)	0.079
Sex				
Male	1	-	1	
Female	0.80 (0.55, 1.15)	0.224		
Age (yrs)	1.01 (1.001, 1.02)	0.036	1.01 (1.0001 1.0193)	0.047
BMI	1.01 (0.98, 1.03)	0.661		
BMI Category				
Underweight	1	-		
Normal	1.17 (0.55, 2.49)	0.692		
Overweight	1.18 (0.53, 2.62)	0.681		
Obese	1.15 (0.53, 2.49)	0.720		
HIV Status:				
Negative	1	-	1	
Positive	0.79 (0.53, 1.18)	0.256	0.80 (0.53, 1.19)	0.271
Unknown	0.86 (0.37, 2.03)	0.732	0.82 (0.34, 1.95)	0.646
Past TB History				
No	1	-	1	
Yes	1.23 (0.55, 2.74)	0.612		
Present Cough				
No	1	-	1	
Yes	1.31 (0.58, 2.95)	0.512		

For participants aged ≥16 years (Table 3.8), age was significant in the univariate model showing that a unit increase in age increased the odds of being TST-positive by 1% (OR=1.01; 95% CI: 1.001, 1.02; p=0.036). The odds were increased for those living in Klerksdorp compared to Agincourt but this was only marginally significant at the 5% level (OR=1.38; 955 CI: 0.97, 1.96; p=0.070). There was a decreased odds of being TST-positive for females compared to males (OR=0.80; 955 CI: 0.55, 1.15; p=0.224) and HIV positive participants compared with HIV negative participants (OR=0.79; 95% CI: 0.53, 1.18; p=0.256), but these were not significant. Participants with a history of TB had an increased odds of being TST-positive compared to those without a history of TB but this finding was not statistically significant (OR=1.23; 95% CI: 0.55, 2.74; p=0.612).

In the multivariate analysis, only age was a statistically significant predictor of a positive TST (OR=1.01; 95% CI: 1.0001, 1.0193; p=0.047). Site and HIV status though included in the model were not statistically significant. The Hosmer-Lemeshow goodness of fit test statistic for the multivariate model was 8.86, p=0.3546.

In a separate analysis conducted for residents of Klerksdorp only (see Appendix 7), no variable approached statistical significance in the univariate and multivariate analysis for participants aged ≥16 years. For participants aged ≤15 years, age, upper arm circumference and a positive HIV status were significant in the univariate analysis. In the multivariate analysis, only age (OR=1.14; 95% CI: 1.01, 1.28; p=0.037) was statistically significant. A positive HIV status was marginally significant (OR=4.08; 95% CI: 0.89, 18.77; p=0.071).

3.5.1.2 Individual predictive factors based on HIV status

The tables of the univariate and multivariate logistic model of individual predictive factors of TST-positive based on HIV status is attached as Appendix 8. No variable was statistically significant in the univariate and multivariate analysis for participants aged ≥16 years who were HIV infected at the time of TST testing. Only age approached statistical significance in the univariate (OR=1.02; 95% CI: 0.99, 1.05; p=0.085) and multivariate analysis (OR=1.02; 95% CI: 0.99, 1.05; p=0.099). For participants of the same age who were HIV negative at the time of TST testing, residing at Klerksdorp was significant in the univariate (OR=1.64; 95% CI:

1.08, 2.49; p=0.019) and multivariate analysis (OR=1.58; 95% CI: 1.04, 2.41; p=0.032). Age was a marginally significant predictor in the univariate (OR=1.01; 95% CI: 0.99, 1.02; p=0.055) and multivariate analysis (OR=1.01; 95% CI: 0.99, 1.02; p=0.058). Findings in participants aged ≤15 years were similar to that from the logistic model for all participants aged ≤15 years. Residing in Klerksdorp, age and upper arm circumference were significant in the univariate analysis while only age was a significant predictor in the multivariate analysis.

3.5.1.3 Multi-level modeling for prediction of individual risk factors of TST-Positive

Findings from the multi-level modeling analysis were similar to the logistic model and are attached as Appendix 9. In participants aged ≥16 years, age was the only significant predicting factor in the univariate (OR= 1.02; 95% CI: 1.01, 1.03; p=0.007) and multivariate analysis (OR=1.02; 95% CI: 1.004, 1.03; p=0.012). For participants aged ≤15 years, age (OR=1.16; 95% CI 1.06, 1.26; p=0.001) and upper arm circumference (OR=1.09; 95% CI: 1.01, 1.18; p=0.031) were significant predictors in the univariate analysis while in the multivariate analysis, only age was significant (OR=1.15; 95% CI: 1.03, 1.28; p=0.019). The standard deviation for household random intercept was 1.84 (95% CI: 1.33, 2.53) for those aged ≥16 years and 1.38 (95% CI: 0.73, 2.63) for ≤15 years old. Site random intercept had a negligible standard deviation for both age groups.

3.5.2 Household-level risk factors

3.5.2.1 Multi-level modeling for prediction of household risk factors of TST-Positive

Univariate multi-level logistic model

The univariate analysis for predictors of positive TST result within households is presented in Table 3.8. Where residents of households smoke, the odds was 2.11 greater (95% CI: 1.15, 3.85; p=0.015) in comparison to households where the residents do not smoke. An increase in the number of windows in the household reduced the odds of having a resident with a positive TST by 8% (OR=0.92; 95% CI: 0.85, 1.002), but this was marginally significant (p=0.055). A unit increase in wealth index increased the odds of a positive TST in the household by 15% (OR=1.15; 95% CI: 0.89, 1.50) but this increase missed statistical significance (p=0.287). Age of house and wealth quintiles did not statistically significantly impact on household TST.

Multivariate multilevel logistic model

The number of windows, smoking in households and wealth index significantly predicted positive TST within households in the multivariate analysis. An increase in the number of windows decreased the odds of a positive TST by 10% (OR=0.90; 95% CI: 0.82, 0.98; p=0.011) while a unit increase in wealth index increased the odds of a positive TST by 39% (OR=1.39; 95% CI: 1.13, 1.70; p=0.002). The odds of a positive TST increased by 2.47 (95% CI: 1.37, 4.44; p=0.003) in households where the residents smoke compared to those where they do not smoke. The intra-cluster effect was negligible (OR=7.64e-07) indicating that within site variation provided a minimal explanation for TST positivity within the households.

Table 3.9 Multi-level modeling (two-level) of household risk factors of TST-Positive taking into account clustering of households within the site

Household Variable	Univariate analysis		Multivariate analysis	
	Odds (95% CI)	P-Value	Odds (95% CI)	P-Value
Age of House	0.99 (0.98, 1.02)	0.962		
Number of Windows	0.92 (0.85,1.002)	0.055	0.90 (0.82, 0.98)	0.011
Smoking in Household:				
No	1		1	
Yes	2.11 (1.15, 3.85)	0.015	2.47 (1.37, 4.44)	0.003
Wealth Index	1.15 (0.89, 1.50)	0.287	1.38 (1.13, 1.70)	0.002
Wealth Quintile				
Very poor	1			
Poor	1.06 (0.48, 2.36)	0.886		
Rich	0.77 (0.33, 1.80)	0.539		
Very Rich	0.96 (0.38, 2.41)	0.937		
Wealthy	1.22 (0.43, 3.48)	0.712		

Findings from the univariate and multivariate analysis are different when compared with a logistic model including only households in Klerksdorp (Table 3.9). No variable statistically significantly predicted a positive household TST. Only smoking within households approached statistical significance in both the univariate and multivariate analysis.

Table 3.10 Logistic model of household risk factors of TST-Positive at Klerksdorp

Household Variable	Univariate analysis		Multivariate analysis	
	Odds (95% CI)	P-Value	Odds (95% CI)	P-Value
Age of House	0.99 (0.97, 1.02)	0.584	0.99 (0.96, 1.01)	0.320
Number of Windows	0.98 (0.83, 1.17)	0.857		
Smoking in Household:				
No	1		1	
Yes	2.04 (0.96, 4.32)	0.062	2.15 (0.97, 4.78)	0.060
Wealth Index	0.89 (0.56, 1.41)	0.613	0.80 (0.49, 1.31)	0.380
Wealth Quintile				
Very poor	1			
Poor	0.76 (0.12, 4.75)	0.771		
Rich	0.64 (0.11, 3.66)	0.615		
Very Rich	0.64 (0.11, 3.66)	0.615		
Wealthy	0.92 (0.15, 5.51)	0.924		

Hosmer-Lemeshow test statistic = 7.28, p=0.5073

4 CHAPTER FOUR: DISCUSSION

This chapter reviews the major findings from the preceding analysis and compares this with that of similar studies conducted within South Africa and other parts of the world. The strength and limitations of the study are identified and discussed to provide guidance for the obtained results and for future studies.

4.1 Summary of key results

From the analysis, the general prevalence of TST-positive was 20.5%. It was 1.5 times higher in Klerksdorp (24.3%) compared to Agincourt (16.3%), and similar in both sexes. Prevalence was noted to increase with an increase in age, only dipping at the 26-30 years age bracket and increasing again after 30 years of age. It was also higher amongst HIV positive participants compared with those who had a negative or unknown HIV status. For prevalence at the household level, the overall prevalence was 54.2% and households at Klerksdorp had a higher prevalence (70.5%) compared to Agincourt (36%).

The overall ARTI was 2.22% when all participants were considered but 3.8% for participants aged <15 years. It was higher in Klerksdorp compared to Agincourt. Males were at a higher risk compared to females and the risk was highest in the youngest age group (<5 years), gradually reducing with an increase in age to the lowest risk at the >45 years age category. The risk was higher amongst participants with negative and unknown HIV status at both sites. The odds of a positive TST was significantly higher amongst HIV positive participants compared to those with negative and unknown status (OR=1.63; 95% CI: 1.13, 2.35, p=0.009). After controlling for age and sex, a positive HIV status was not a significant determinant of a positive tuberculin skin test result (OR=1.12; 95% CI: 0.76, 1.65; p=0.567).

Significant individual risk factor which predicted TST-positive after controlling for other factors was the age in participants aged ≤15 years and ≥16 years. Positive HIV status and past history of tuberculosis were not significant and there was no significant association between sex and latent TB infection. Significant household factors which predicted a positive TST

within households were smoking within households by the residents, the number of windows in the house and the wealth index.

4.2 Comparison with other studies

The overall prevalence of TST-positive from this study showed significant variability in the age category and between the two sites which participants were drawn from. It is close to the global estimate of 24.1% given by Cohen et al in their systematic review and meta-analysis (20) but lower than a finding of 34.3% in a cross-sectional study conducted in a Johannesburg township South Africa (13) and amongst urban residents of Kampala, Uganda (58). It is also lower than a similar study conducted in Cape Town where TST-positive prevalence was between 28.0 – 88.2% depending on the age category with an overall prevalence of 45.0% (5). This difference possibly could be attributed to the fact overall prevalence from our analysis is a combined estimate from an urban and a rural area. Considering the fact our study participants were drawn from the general population with minimal restriction, our determined prevalence is observed to be lower compared to some studies above and in some high-risk population groups within South Africa and other middle and low-income countries (10, 21, 22). Our finding nevertheless, is much higher when compared with a study conducted in New York to determine prevalence and predictors of TST-positivity amongst injection drug users where they estimated prevalence to be 14% (12) and in Germany where 11.3% of participants screened for latent TB prior to commencement of treatment for rheumatology related conditions had a positive TST result (32).

4.2.1 Prevalence of latent tuberculosis

A higher prevalence was seen in the urban site of our study. A possible explanation for this could be the fact that Klerksdorp is already known to have a high TB prevalence (6, 59) with associated high transmission rate and the urban nature of Klerksdorp which has a larger population, with about 66% of its inhabitants being in the 15-64 age bracket who are involved in mining, construction and other occupations such as health which are identified risk factors for tuberculosis (50). Another reason could be the residential arrangement of Klerksdorp where mining villages are established within urban areas allowing for regular interaction between miners and those in other occupations who may not necessarily be at risk of infection at work (50). The availability of large social networks and more opportunities for mixing which exists

in an urban setting leading to its residents being more exposed to infection compared to those residents in rural areas provides another justification for the difference in prevalence between our study sites (58, 60, 61). Residents of Agincourt, on the contrary, are settled immigrants engaged in farming and other subsistence activities. There are no mines or similar occupations which could result in increased exposure to TB infections given that the risk of infection has been shown to increase with employment as documented by a study conducted in Uganda which showed that leaving home for school or work increased the risk of TB infection (58). We did not have a variable describing occupation so its effect could not be assessed.

4.2.1.1 Effect of sex and age on TST-Positive prevalence

The similarity in prevalence between males and females overall and within site was not surprising as studies conducted within South Africa (5, 62), Ethiopia (63), Tanzania (64) and Asian countries of China and Indonesia (44, 65) had a similar outcome. However, a community-based study conducted in Ethiopia (66) and Uganda (58) indicated a higher TST-positive prevalence in males. These conflicting findings show the possible differences which exist in the different populations and circumstances under which the studies were conducted. Males and females are exposed to different conditions and contexts which determines their risks of infection but it can be argued that neither males nor females who participated in our study were exposed to exceptional risks to warrant a statistically different prevalence. Also, there is a possibility of sampling bias or selection bias during placement of TST as males accounted for only 39.5% of participants with a TST result. This also indicates there are fewer males in the households selected for the study.

Our determined TST-positive prevalence based on age and site was variable. This variability and increase in prevalence with an increase in age category are comparable with other studies conducted in South Africa and other parts of the world (5, 13, 58, 65). In a previous study conducted in Matlosana (Klerksdorp) by Lebina et al to determine the prevalence of LTBI in children aged 5 and 7 years, TST-positive prevalence was 15.1% and 19.1% respectively (11), slightly higher than 7.2% (95% CI: 3.8% - 13.3%) obtained in our study for children in the 5-10 years age group. This indicates a possible decrease in TST-positive prevalence at the site due to increased HIV/TB prevention programs and treatment services but when considering

that the number of participants in our 5-10 age group is few (n=125) compared to the earlier study, a much larger sample will be required to make a confident assessment. Like it was observed in other studies, the prevalence was higher in the older age group. This could be credited to the fact there is accumulated exposure with increase in age, waning immunity and increased access to social networks amongst the older population compared to young children (5, 13). More importantly, a higher prevalence observed in Klerksdorp can be attributed to the age distribution of our participants where Klerksdorp had more residents in the >45 age category, and Agincourt had more residents in the <5 age category (Table 1).

4.2.1.2 TST-Positive prevalence and nutritional status

The effect of nutritional status of participants on TST-positive prevalence assessed using upper arm circumference for participants aged ≤15 years and BMI for participants aged ≥16 years (Table 3.4) did not reveal any definite trend. Although malnutrition has been linked with increased risk of tuberculosis in susceptible people, previous studies which assessed the effect of BMI on LTBI prevalence similarly did not show any significant difference in TST-positive prevalence based on BMI (13, 63, 67). The risk of TB infection and mortality amongst adult PLHIV was studied in South Africa and findings presented by Hanrahan et al (2010) shows PLHIV who are overweight or obese had reduced risk of mortality and reduced risk of TB infection compared to those with a normal weight (68). In contradiction, a population-based survey in rural China discovered that the risk of TB infection was higher amongst those who were obese (69). These different findings could be attributed to the difference in the two populations with China having an older population who are at an increased risk of TB infection.

4.2.1.3 TST-Positive prevalence and HIV

Previous studies conducted in South Africa and Zambia showed there was no difference in TST-positive prevalence between HIV positive and negative participants (13, 70) but an earlier study conducted in Khayelitsha South Africa (published 2007) showed that approximately 69% of the study participants who were HIV positive had a positive TST (36). The prevalence of TST-positive based on HIV status varied between sites in this research (Figure 3.4). Due to the low sensitivity of TST amongst HIV positive participants (40), the TST cut off for HIV positive participants was set at ≥5 mm but there was still no significant difference in prevalence between

HIV positive and negative participants at the site with a high HIV prevalence. A significant difference between HIV positive and negatives are seen at Agincourt could be attributed to the few numbers of participants at the site with a positive TST with a greater proportion being HIV positive. The guidance by WHO to provide TB prophylactic treatment to all who are HIV positive regardless of TST status after ruling out active TB disease is based on their susceptibility to TB infection and the poor ability of available tests to diagnose latent TB in those who are HIV positive (14, 71). Though there is a shift in the use of CD4 count for monitoring of PLHIV to viral load, participants in the study with a positive HIV status who had a CD4 done had a high median CD4 count as illustrated in Table 3.1. This indicates a greater proportion of the participants had an optimal immune system.

4.2.1.4 Prevalence using different TST cut-offs

A high prevalence of TST-positive as seen across the three TST cut-offs indicates a very high prevalence of TST-positives in our sampled population, even at ≥15 mm. This is similar to other studies with a high TST-positive prevalence using a TST cut off of ≥10 mm (5, 13, 67) and demonstrates the need for more effort in increasing access to TB prevention and treatment services within communities and overall improvement in healthcare. In a previous study conducted in New York, there was no difference in prevalence when TST cut-off was reduced to ≥5 mm from ≥10 mm as a majority of the participants who had a positive TST had measurements greater than 10 mm (12). In South Africa, prevalence was high at the three cut-off points in a study to determine effect of environmental tobacco smoke exposure on the risk of TB infection amongst children in a high TB endemic locality and also in another study to determine ARTI using different methods in high HIV/TB prevalent communities (16, 72). Furthermore, we noted a trend of increasing prevalence with age using TST measurements of ≥5 mm and ≥15 mm (Figure 3.5). This incontrovertibly supports earlier findings of age being a major determinant of TST-positive prevalence.

4.2.1.5 Household TST-Positive prevalence

The high prevalence of household TST-positive overall and at the sites is representative of residents in the households and indicates a higher TST-positive prevalence at Klerksdorp. Some of the household characteristics evaluated were crowding index and number of windows. Our

analysis revealed Klerksdorp had a higher crowding index with fewer windows. This, in previous studies, has been linked to a higher TST-positive prevalence (70, 71).

Compared to households where residents do not smoke, the prevalence was higher in households where residents smoke but this difference was marginally significant at Klerksdorp and not statistically significant at Agincourt. Smoking and air pollution from environmental tobacco smoke within households have been identified as a risk for TB infection in adults and young children (25, 26, 58, 71), and the higher household prevalence where residents smoke in our study confirms these previous findings.

Data used for our analysis did not include household socioeconomic status, but a wealth index derived using available household variables presented as wealth quintiles (Table 3.5, Appendix 4). Although no visible trend was seen within each site, a marginally significant finding of increasing prevalence with an increase in wealth quintile overall is contrary to a generally held belief that tuberculosis is a disease associated with poverty. This raises a concern in our study as the 264 households studied was probably few (we used only households linked to participants with a TST result). A minimum of 752 households (376 at each site) at an alpha level of 5% and power of 80% would have been required to determine a 10% difference between the two sites. A review of baseline household characteristics at our two sites (Table 1) shows that the site with a higher household TST-positive prevalence, Klerksdorp, has more risks for TST-positive compared to Agincourt. These include a fewer median number of windows, a higher crowding index and a greater proportion of residents who smoke within the households.

4.2.2 Annual risk of tuberculosis infection

Our second objective was to determine ARTI by age and sex using values of TST-positive prevalence determined for each site and overall (Table 3.6). ARTI was observed to be higher in the analysis for those aged <15 years compared to that for all participants. Values were initially different at the younger ages but became similar at both sites as age increased, with the highest risk of infection in the youngest participants and decreasing with increase in age (Figure 3.6 & 3.7). A slightly higher ARTI seen in males when viewed by sex and year at both sites (Figure 3.8) could be due to the mean age used for determination of ARTI which was lower in

males compared to females. Also, ARTI seen amongst HIV negative and unknown participants was higher compared to those with a positive status. This could be due to our sample population where 97% of the participants aged <15 (children had the highest ARTI) had a negative and unknown HIV status (Figure 3.9).

Most studies conducted to determine ARTI in South Africa has been in children and adolescents, and findings from our study are similar to these studies which we have observed used different age distribution for determination of ARTI (5, 11, 13, 16, 73). An overall ARTI of 3.8% in participants aged <15 as seen in our study indicates a very high ongoing TB infection rate and this is higher than 2.9% seen when ARTI was determined in 5 & 7-year-olds at Matlosana (Klerksdorp) in 2015 (11) which is one of our study sites. In comparison to our findings, studies conducted in Zambia and Malawi had lower estimates of ARTI, whatever the method used for its calculation (57). This could possibly be attributed to the lower prevalence of TB and HIV in these countries compared to South Africa.

4.2.3 TST-Positivity based on HIV status

Studies conducted in South Africa and in other countries indicated there is no association between HIV status and TST-positive (10, 13, 58). In this study, the initial test of association using a Pearson's Chi-square comparison test showed a significant association between the two, but this became insignificant after controlling for age and sex. Thus we can conclude our findings are similar to other studies.

4.2.4 Risk factors for a Positive TST

4.2.4.1 Individual risk factors

From the logistic regression models fitted to determine risk factors which predict TST-positive at the individual level, age was the only significant predictor of TST-positive in the univariate and multivariate analysis for all participants with a greater odds in participants aged ≤15 years (Table 3.7 & 3.8, Appendix 9). This is consistent with our finding of increased risk of TB infection amongst younger participants (Figure 3.6). Prior studies equally identified age as an important predictor of a positive TST (5, 58, 62). Developing immune system and lack of prior exposure could be responsible for the increased risk in younger participants. Upper arm

circumference, residing in Klerksdorp, sex, BMI, past TB history and present cough were not significant predictors in all participants. Other studies have also documented that HIV status, sex, and BMI were not significant predictors of latent TB infection as seen in our study (13, 58, 66). From the multi-level analysis, standard deviation for household random intercept were different from zero indicating there were unexplained variations in TST positivity due to household level effect. The household variables are discussed in a separate section.

A further analysis based on HIV status indicated only age was a significant predictor in those aged ≤15 with a negative HIV status, while in those aged ≥16, no strong evidence of the effect of HIV status could be seen. These findings are similar to what was obtained from the logistic model and confirm the lack of association between HIV status and a positive TST as documented by other studies (13, 58). We also tried to ascertain if living in Klerksdorp affected the predictors significantly but no variable significantly predicted TST-positive amongst those aged ≥16 years who are resident in Klerksdorp (Appendix 7). Age was a significant predictor for those aged ≤15 years and a positive HIV status was marginally significant. So we conclude that our results are not significantly different for those living n Klerksdorp.

4.2.4.2 Household risk factors

At the household level, smoking within the household was the only significant predictor in the univariate analysis while the number of windows, smoking within the household and wealth index were significant predictors in the multivariate analysis (Table 3.9). Previous studies have demonstrated that smoking and exposure to tobacco smoke is an important predictor of TST-positive and where patients are being treated for TB, could lead to a poor outcome (25, 26, 71, 72). This is clearly demonstrated in this study and is consistent with our reported prevalence in households where residents smoke in comparison to those where they do not smoke (Table 3.5).

Overcrowding in a household has been shown in other studies to be an important predictor of TST-positivity (13, 70) but some other study conducted in Ethiopia showed this to be the reverse (63). Although we did not include crowding index in the final multivariate analysis because it was used in estimating wealth index, a separate model with crowding index showed it to be a significant predictor of TST-positivity in households. The high risk of infection due

to overcrowding has been associated with a possible increased contact time with infectious cases in a poorly ventilated environment (74).

Ventilation which is closely associated with overcrowding has been identified as a critical factor in the transmission of tuberculosis infection in an enclosed space where people who are potentially infected spend time with those who are not infected. This has been documented in different tuberculosis transmission studies in homes, hospitals, mines and other congested settings such as prisons (74-76). Our analysis showed that increasing number of windows significantly reduced the risk of TB infection by 10%, thus corroborating findings from the 2010 South African study by Middelkoop et al that increasing ventilation reduces the risk of TB infection in children aged 0-5 years (74). Some studies have linked increased risk of TB infection to use of biomass for cooking in a poorly ventilated kitchen or houses as this reduces the indoor air quality (77, 78).

The finding of an increased prevalence with wealth index in our multivariate analysis (socioeconomic status) is conflicting. Two previous studies, one conducted in Johannesburg - South Africa and the other in Zambia documented an increase in prevalence with higher socioeconomic status (13, 70), while other studies, including one conducted in Tanzania to identify how social activity patterns influence high rates of latent tuberculosis in adolescents, maintain that increased prevalence is associated with a lower socioeconomic status (62, 64, 71).

Employment, schooling, urban-dwelling with possible poor housing conditions and increased opportunities for social interactions are some of the factors thought to be responsible for the dose-response relationship between TST-positive prevalence and socioeconomic status (13, 58, 70). On the contrary, poor access to health services, poor living conditions, and homelessness, overcrowding, and malnutrition are some of the risk factors amongst people of low socioeconomic status responsible for the high risk of tuberculosis infection (71). A factor which could be responsible for the differences in the effect of socioeconomic status with regards to TB infection is the manner in which it is derived. Some studies utilize household income, ownership of certain household properties or availability of certain facilities within households. These proxy measures could be responsible for the different views and as such a standardized

means of measuring socioeconomic status with respect to TB research is required to overcome the differences observed in the effect of socioeconomic status (70).

4.3 Strength and Limitations of the study

4.3.1 Strength of the study

There was a very little restriction on participants selected for inclusion in the study which allowed for the determination of prevalence, the annual risk of infection and risk factors across all age groups without bias. This made it possible to identify subtle differences between children and adults. Estimation of prevalence using different TST cut-offs (5mm and 15mm) allowed for the comparison of prevalence obtained from these different cut-offs with the standard cut of ≥10mm. Data on HIV status allowed for assessment of the effect of HIV status on TST-positivity.

Availability of data linking participants to households made it possible to determine the household level prevalence and to model household clustering effect during the analysis. In addition, household-level data made it possible to identify household-level risk factors for TST-positive at different sites. Since we had data for two sites- rural and urban, it was possible to compare the prevalence and identify baseline risk factors between both sites. Most studies are conducted either in an urban or rural setting but this study combined both. Findings from this research are consistent with that of other studies conducted in South Africa and other parts of the world.

4.3.2 Limitations

This study involved the use of secondary data so the analysis was restricted to variables which were available within the data set. Based on this fact, certain TB risk factors could not be analyzed. It was not possible to assess if BCG vaccination had any effect on TST response. Incomplete follow up data made it impossible to determine TB incidence. Variables describing socioeconomic status, education, race, and occupation were missing so the effect of these could not be assessed. A proxy household socioeconomic status was rather derived from other household variables. It was also not possible to determine the impact of household TB contact on TST-positive as data on household TB contacts was not available. Additional data on the

sputum test was not available. As such, it was impossible to determine sensitivity, specificity, positive predictive value and negative predictive value of tuberculin skin test as a screening test. This would have been useful for TB programming in designing latent TB screening programs for at-risk populations.

Data for this research was obtained from a cohort study (PHIRST Study) which enrolled new participants annually from 2016 – 2018. Challenges with administration and collection of TST data in 2017 resulted in a fewer number of study participants in the 2017 cohort with TST measurements compared to other years. This limited yearly comparison of derived prevalence. The data for our study showed features of possible terminal digit preference at one of the sites. Digit preference is the inclination towards certain numbers by humans during measurements. This may be odd numbers, even numbers or rounding up to certain whole numbers. In the determination of ARTI, only the fixed cut off method was used. If some other method was used, it would have provided a basis for comparison of our obtained values of ARTI in our sampled population to mitigate the effect of age which was skewed in our sample.

5 CHAPTER FIVE: RECOMMENDATIONS AND CONCLUSION

5.1 Recommendations

This research showed the prevalence of TST-positive is very high in the sampled population and by extension South Africa. The high prevalence with a higher cut off point indicates an urgent need for intervention. The available latent TB screening and treatment services should be strengthened to increase access to these services by those in the high-risk population group and the general public. There should be an increased awareness and health education programs on LTBI so the general public is aware of the various risk factors and how they can protect themselves.

Our findings showed that children are at a higher risk of getting infected compared to adults with an increase in prevalence after the age of 10 years. This suggests that any vaccination strategy for prevention should target children below this age. Adequate measures should also be in place to protect children through health education on risk factors and periodic screening at schools. Prophylactic treatment suitable for children should be provided preferably targeting those leaving school after a high proportion would have been infected. Monitoring and treatment of adverse events related to prophylactic TB treatment should be improved through regular screening and reviews while on treatment.

Some occupations have been identified as high risk for TB infection. A periodic work place based screening for latent TB should be instituted, implemented and monitored in these occupations. This will help identify staff who will benefit from prophylactic treatment or full treatment for tuberculosis. The current trend of screening at the onset of employment without subsequent practical follow up in some occupations is not beneficial.

Smoking in households, ventilation, and overcrowding was identified as a prominent risk factor for a positive TST. Smoking cessation campaigns and campaigns to end smoking within households should form part of routine TB prevention services by the TB control programs and civil society organizations since this places children resident in those homes at a much greater risk compared to other children. Ventilation and overcrowding in homes, hospitals, and other

congregate settings should be improved upon. Children should ideally, sleep in separate bedrooms from adults to limit possible exposure. Use of biomass for cooking should be discouraged. Where this is not possible, biomass fuel should be used for cooking in open spaces and not within households to reduce indoor air pollution which can persist for a long time after cooking.

5.2 Conclusion

Despite advances in TB prevention and treatment services in South Africa, the prevalence of LTBI using TST as shown in this study is still high and consistent with previous prevalence studies conducted in South Africa. A higher prevalence seen amongst participants and households in Klerksdorp indicates its residents are exposed to more TB risk factors compared to Agincourt, particularly in the older age group. The highest annual risk of TB infection was seen amongst children aged <15 years and indicated a high rate of ongoing TB transmission within households and communities. Age was the single most important individual predictor of TST-positive in the study. HIV status, past TB history, and sex difference did not significantly increase the risks of TST-positive in adults and children. Smoking within households, ventilation (number of windows), overcrowding and wealth index were significant household predictors of TST-positive.

5.3 Recommendations for future research

1. Children were identified to have the highest risk of TST-positive in our study. Research to determine how best to protect young children in the South African context will be appropriate to minimize the incidence of childhood TB.

2. HIV positive status was not a significant predictor of TST-positive, though the impact of HIV on active TB disease is well documented. Further studies are required to ascertain if there is an association between the two and the most appropriate screening test for LTBI amongst PLHIV. This can minimize expenses on the provision of prophylactic treatment for LTBI.

REFERENCES

1. World Health Organisation. Global Tuberculosis Report 2018 Geneva: WHO; 2018 [accessed 2019, 05 January] Available from: http://http://apps.who.int/iris/bitstream/handle/10665/274453/9789241565646-eng.pdf?ua=1.

2. World Health Organisation. Tuberculosis Fact Sheet Geneva: WHO; 2018 [accessed 2019, 05 January]. Available from: http://www.who.int/news-room/fact-sheets/detail/tuberculosis.

3. Sulis G, Centis R, Sotgiu G, D'Ambrosio L, Pontali E, Spanevello A, et al. Recent developments in the diagnosis and management of tuberculosis. NPJ Primary Care Respiratory Medicine. 2016;26(1).

4. Maung HMW, Thein S, Aung ST, Mynit Z, Thu MK, Kumar AMV, et al. High treatment success rate among multidrug-resistant tuberculosis patients in Myanmar, 2012–2014: a retrospective cohort study. Transactions of The Royal Society of Tropical Medicine and Hygiene. 2018;111(9):410-7.

5. Wood R, Liang H, Wu H, Middelkoop K, Oni T, Rangaka MX, 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):406-12.

6. Mvusi L. Burden of TB Disease in South Africa South Africa: National Department of Health; 2017 [accessed 2019, 18 January]. Available from: http://www.health.gov.za/index.php/2014-03-17-09-09-38/strategic-documents/category/392-nepad-regional-meeting-midrand-2017?download=2253:burden-of-tb-disease-in-south-africa&start=20.

7. World Health Organisation. South Africa Tuberculosis Profile Geneva: WHO; 2019 [accessed 2019, 08 January]. Available from: https://extranet.who.int/sree/Reports?op=Replet&name=%2FWHO_HQ_Reports%2FG2%2FPROD%2FEXT%2FTBCountryProfile&ISO2=ZA&LAN=EN&outtype=html.

8. Statistics South Africa (STATSSA). Mortality and causes of death in South Africa, 2016: Findings from death notification. Statistical Release P0309.3. Pretoria: STATSSA; 2018. p. 146.

9. Robertson BD, Altmann D, Barry C, Bishai B, Cole S, Dick T, et al. Detection and treatment of subclinical tuberculosis. Tuberculosis (Edinb). 2012;92(6):447-52.

10. Hanifa Y, Grant AD, Lewis J, Corbett EL, Fielding K, Churchyard G. Prevalence of latent tuberculosis infection among gold miners in South Africa. The international journal of tuberculosis and lung disease : the official journal of the International Union against Tuberculosis and Lung Disease. 2009;13(1):39-46.

11. Lebina L. Latent Tuberculosis Infection in School Children and Contact Tracing in Matlosana, North West Province, South Africa. The international journal of tuberculosis and lung disease : the official journal of the International Union against Tuberculosis and Lung Disease. 2015;19(11):1290-2.

12. Salomon N, Perlman DC, Friedmann P, Ziluck V, Des Jarlais DC. Prevalence and risk factors for positive tuberculin skin tests among active drug users at a syringe exchange program. The international journal of tuberculosis and lung disease : the official journal of the International Union against Tuberculosis and Lung Disease. 2000;4(1):47-54.

13. Ncayiyana JR, Bassett J, West N, Westreich D, Musenge E, Emch M, 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.

14. World Health Organisation. Latent tuberculosis infection: updated and consolidated guidelines for programmatic management [Guideline]. Geneva: WHO; 2018 [accessed 2019, 8 January]. Available from: http://www.who.int/tb/publications/2018/latent-tuberculosis-infection/en/.

15. Rieder H. Annual risk of infection with Mycobacterium tuberculosis. The European respiratory journal. 2005;25(1):181-5.

16. Shanaube K, Sismanidis C, Ayles H, Beyers N, Schaap A, Lawrence K-A, et al. Annual Risk of Tuberculous Infection Using Different Methods in Communities with a High Prevalence of TB and HIV in Zambia and South Africa. PLOS ONE. 2009;4(11):e7749.

17. Styblo K. The relationship between the risk of tuberculous infection and the risk of developing infectious tuberculosis. Bull IUAT. 1985;60(3):117-9.

18. van Leth F, van der Werf M, Borgdorff M. Prevalence of tuberculous infection and incidence of tuberculosis; a re-assessment of the Styblo rule. Bulletin of the World Health Organization. 2008;86(1):20-6.

19. Nahid P, Pai M, Hopewell PC. Advances in the diagnosis and treatment of tuberculosis. Proc Am Thorac Soc. 2006;3(1):103-10.

20. Cohen A, Mathiasen VD, Schön T, Wejse C. The Global Prevalence of Latent Tuberculosis: A Systematic Review and Meta-Analysis. European Respiratory Journal. 2019:1900655.

21. Joshi R, Reingold AL, Menzies D, Pai M. Tuberculosis among health-care workers in low- and middle-income countries: a systematic review. PLoS medicine. 2006;3(12):e494.

22. Salehi M, Sharifi Mood B, Metanat M. Positive Tuberculin Skin Test Among Health Care Workers: Prevalence and Risk Factors in Teaching Hospitals of a Highly Endemic Region for Tuberculosis, Zahedan, Iran. Int J Infect. 2016;3(3):e36158.

23. Shapiro AE, Variava E, Rakgokong MH, Moodley N, Luke B, Salimi S, 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):1110-6.

24. Moran-Mendoza O, Marion SA, Elwood K, Patrick D, FitzGerald JM. Risk factors for developing tuberculosis: a 12-year follow-up of contacts of tuberculosis cases. The international journal of tuberculosis and lung disease : the official journal of the International Union against Tuberculosis and Lung Disease. 2010;14(9):1112-9.

25. Lin H-H, Ezzati M, Murray M. Tobacco Smoke, Indoor Air Pollution and Tuberculosis: A Systematic Review and Meta-Analysis. PLOS Medicine. 2007;4(1):e20.

26. Jafta N, Jeena PM, Barregard L, Naidoo RN. Childhood tuberculosis and exposure to indoor air pollution: a systematic review and meta-analysis. The International Journal of Tuberculosis and Lung Disease. 2015;19(5):596-602.

27. Arnadottir T, Rieder H, Trebucq A, Waaler H. Guidelines for conducting tuberculin skin test surveys in high prevalence countries. Tubercle and lung disease. 1996;77:1-19.

28. Malhamé I, Cormier M, Sugarman J, Schwartzman K. Latent Tuberculosis in Pregnancy: A Systematic Review. PLoS One. 2016;11(5).

29. Robin EH, Maybelle FS, John B. Bass, Jr. The Tuberculin Skin Test. Clinical Infectious Diseases. 1993;17(6):968-75.

30. Jasmer RM, Nahid P, Hopewell PC. Latent tuberculosis infection. New England Journal of Medicine. 2002;347(23):1860-6.

31. Diagnostic Standards and Classification of Tuberculosis in Adults and Children. American journal of respiratory and critical care medicine. 2000;161(4):1376-95.

32. Kleinert S, Tony H-P, Krueger K, Detert J, Mielke F, Rockwitz K, et al. Screening for latent tuberculosis infection: performance of tuberculin skin test and interferon-γ release assays under real-life conditions. Annals of the Rheumatic Diseases. 2012;71(11):1791-5.

33. Rangaka MX, Wilkinson KA, Glynn JR, Ling D, Menzies D, Mwansa-Kambafwile J, et al. Predictive value of interferon-γ release assays for incident active tuberculosis: a systematic review and meta-analysis. The Lancet Infectious Diseases. 2012;12(1):45-55.

34. Abubakar I, Drobniewski F, Southern J, Sitch AJ, Jackson C, Lipman M, et al. Prognostic value of interferon-γ release assays and tuberculin skin test in predicting the development of active tuberculosis (UK PREDICT TB): a prospective cohort study. The Lancet Infectious Diseases. 2018;18(10):1077-87.

35. Menzies D. What does tuberculin reactivity after bacille Calmette-Guerin vaccination tell us? Clinical Infectious Diseases. 2000;31(Supplement_3):S71-S4.

36. Rangaka MX, Wilkinson KA, Seldon R, Van Cutsem G, Meintjes GA, Morroni C, et al. Effect of HIV-1 infection on T-Cell-based and skin test detection of tuberculosis infection. American journal of respiratory and critical care medicine. 2007;175(5):514-20.

37. de Leon DP, Acevedo-Vasquez E, Alvizuri S, Gutierrez C, Cucho M, Alfaro J, et al. Comparison of an interferon-gamma assay with tuberculin skin testing for detection of tuberculosis (TB) infection in patients with rheumatoid arthritis in a TB-endemic population. The Journal of rheumatology. 2008;35(5):776-81.

38. Joos TJ, Miller WC, Murdoch DM. Tuberculin reactivity in bacille Calmette-Guerin vaccinated populations: a compilation of international data. The international journal of tuberculosis and lung disease : the official journal of the International Union against Tuberculosis and Lung Disease. 2006;10(8):883-91.

39. Wang L, Turner M, Elwood R, Schulzer M, FitzGerald J. A meta-analysis of the effect of Bacille Calmette Guerin vaccination on tuberculin skin test measurements. Thorax. 2002;57(9):804-9.

40. Swaminathan S, Subbaraman R, Venkatesan P, Subramanyam S, Kumar S, Mayer K, et al. Tuberculin skin test results in HIV-infected patients in India: implications for latent tuberculosis treatment. The international journal of tuberculosis and lung disease. 2008;12(2):168-73.

41. Nayak S, Acharjya B. Mantoux test and its interpretation. Indian dermatology online journal. 2012;3(1):2-6.

42. Eom JS, Kim I, Kim W-Y, Jo E-J, Mok J, Kim M-H, et al. Household tuberculosis contact investigation in a tuberculosis-prevalent country: Are the tuberculin skin test and interferon-gamma release assay enough in elderly contacts? Medicine. 2018;97(3):e9681.

43. Rathi S, Akhtar S, Rahbar M, Azam S. Prevalence and risk factors associated with tuberculin skin test positivity among household contacts of smear-positive pulmonary tuberculosis cases in Umerkot, Pakistan. The International Journal of Tuberculosis and Lung Disease. 2002;6(10):851-7.

44. Burhan E, Intani CN, Handayani H, Nurwidya F. Comparison of tuberculin skin test and interferon-gamma release assay in the diagnosis of latent tuberculosis infection among indonesian health-care workers. Journal of Natural Science, Biology and Medicine. 2019;10(1):53.

45. Stokols D. Translating social ecological theory into guidelines for community health promotion. American journal of health promotion : AJHP. 1996;10(4):282-98.

46. Kahn K, Collinson MA, Gómez-Olivé FX, Mokoena O, Twine R, Mee P, et al. Profile: Agincourt health and socio-demographic surveillance system. International journal of epidemiology. 2012;41(4):988-1001.

47. MRC/Wits - Agincourt Unit: Rural Public Health and Health Transitions Research Unit 2018 [accessed 2019, 19 January]. Available from: https://wwww.agincourt.co.za/?page_id=1805.

48. Anstey Watkins J, Goudge J, Gómez-Olivé F, Griffiths F. Mobile phone use among patients and health workers to enhance primary healthcare: A qualitative study in rural South Africa. Social Science & Medicine. 2018;198:139-47.

49. Gómez-Olivé FX, Angotti N, Houle B, Klipstein-Grobusch K, Kabudula C, Menken J, et al. Prevalence of HIV among those 15 and older in rural South Africa. AIDS care. 2013;25(9):1122-8.

50. City of Matlosana: 2016/17 Annual Report Matlosana:2018 [accessed 2019, 19 January]. Available from: http://www.matlosana.gov.za/Documents/AnnualReports/AnnualReport16-17/ANNUAL%20REPORT%202016-17%20FINAL%20COMPLETE%20REPORT.pdf.

51. Centres for Disease Control (CDC). Guidelines for Preventing the Transmission of *Mycobacterium tuberculosis* in Health-Care Settings, 2005. MMWR. 2005(54(No. RR-17)):1 - 141.

52. Gray A. Definitions of Crowding and the Effects of Crowding on Health: A Literature Review Prepared for the Ministry of Social Policy of New Zealand. 2001.

53. Goodyear R, Fabian A, Hay J. Finding the crowding index that works best for New Zealand. Statistics New Zealand Working Paper. 2011(11-04).

54. Kumaranayake L, Vyas S. Constructing socio-economic status indices: how to use principal components analysis. Health Policy and Planning. 2006;21(6):459-68.

55. Fry K, Firestone R, Chakraborty N. Measuring equity with nationally representative wealth quintiles. Washington, DC: PSI. 2014.

56. Long SS, Prober CG, Fischer M. Principles and Practice of Pediatric Infectious Diseases E-Book: Elsevier Health Sciences; 2017.

57. Khan PY, Glynn JR, Mzembe T, Mulawa D, Chiumya R, Crampin AC, et al. Challenges in the Estimation of the Annual Risk of Mycobacterium tuberculosis Infection in Children Aged Less Than 5 Years. American Journal of Epidemiology. 2017;186(8):1015-22.

58. Kizza FN, List J, Nkwata AK, Okwera A, Ezeamama AE, Whalen CC, et al. Prevalence of latent tuberculosis infection and associated risk factors in an urban African setting. BMC Infectious Diseases. 2015;15(1):165.

59. van Schalkwyk C, Variava E, Shapiro AE, Rakgokong M, Masonoke K, Lebina L, et al. Incidence of TB and HIV in prospectively followed household contacts of TB index patients in South Africa. PLoS One. 2014;9(4):e95372.

60. Andre M, Ijaz K, Tillinghast JD, Krebs VE, Diem LA, Metchock B, et al. Transmission network analysis to complement routine tuberculosis contact investigations. American journal of public health. 2007;97(3):470-7.

61. Cook VJ, Sun SJ, Tapia J, Muth SQ, Argüello DF, Lewis BL, et al. Transmission Network Analysis in Tuberculosis Contact Investigations. The Journal of Infectious Diseases. 2007;196(10):1517-27.

62. Mahomed H, Hawkridge T, Verver S, Geiter L, Hatherill M, Abrahams D, 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. 2011;15(3):331-6.

63. Teklu T, Legesse M, Medhin G, Zewude A, Chanyalew M, Zewdie M, et al. Latent tuberculosis infection and associated risk indicators in pastoral communities in southern Ethiopia: a community based cross-sectional study. BMC Public Health. 2018;18.

64. Maro II, Nakamura K, Seino K, Pallangyo K, Munseri P, Matee M, et al. Social Activity Patterns Drive High Rates of Latent Tuberculosis Infection among Adolescents in Urban Tanzania—Latent TB Infection in Adolescents, Tanzania. Journal of Tuberculosis Research. 2018;6(01):81.

65. Chen B, Gu H, Wang X, Wang F, Peng Y, Ge E, et al. Prevalence and determinants of latent tuberculosis infection among frontline tuberculosis healthcare workers in southeastern China: A multilevel analysis by individuals and health facilities. International Journal of Infectious Diseases. 2019;79:26-33.

66. 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 infectious diseases. 2011;11(1):89.

67. Martinez L, Arman A, Haveman N, Lundgren A, Cabrera L, Evans CA, et al. Changes in tuberculin skin test positivity over 20 years in periurban shantytowns in Lima, Peru. The American journal of tropical medicine and hygiene. 2013;89(3):507-15.

68. Hanrahan CF, Golub JE, Mohapi L, Tshabangu N, Modisenyane T, Chaisson RE, et al. Body mass index and risk of tuberculosis and death. AIDS (London, England). 2010;24(10):1501-8.

69. Zhang H, Li X, Xin H, Li H, Li M, Lu W, et al. Association of Body Mass Index with the Tuberculosis Infection: a Population-based Study among 17796 Adults in Rural China. Scientific reports. 2017;7:41933.

70. Boccia D, Hargreaves J, Ayles H, Fielding K, Simwinga M, Godfrey-Faussett P. Tuberculosis infection in Zambia: the association with relative wealth. The American journal of tropical medicine and hygiene. 2009;80(6):1004-11.

71. Duarte R, Lönnroth K, Carvalho C, Lima F, Carvalho ACC, Muñoz-Torrico M, et al. Tuberculosis, social determinants and co-morbidities (including HIV). Pulmonology. 2018;24(2):115-9.

72. du Preez K, Mandalakas AM, Kirchner HL, Grewal HM, Schaaf HS, van Wyk SS, et al. Environmental tobacco smoke exposure increases Mycobacterium tuberculosis infection risk in children. The international journal of tuberculosis and lung disease : the official journal of the International Union against Tuberculosis and Lung Disease. 2011;15(11):1490-6, .

73. Kritzinger FE, Boon SD, Verver S, Enarson DA, Lombard CJ, Borgdorff MW, et al. No decrease in annual risk of tuberculosis infection in endemic area in Cape Town, South Africa. Tropical Medicine & International Health. 2009;14(2):136-42.

74. Middelkoop K, Bekker L-G, Wood R, Lawn SD, Johnstone-Robertson S, Hargrove J, et al. Tuberculosis Transmission to Young Children in a South African Community: Modeling Household and Community Infection Risks. Clinical Infectious Diseases. 2010;51(4):401-8.

75. Fennelly KP, Nardell EA. The Relative Efficacy of Respirators and Room Ventilation in Preventing Occupational Tuberculosis. Infection Control & Hospital Epidemiology. 1998;19(10):754-9.

76. Robertson S, Lawn SD, Welte A, Bekker L-G, Wood R. Tuberculosis in a South African prison–a transmission modelling analysis. South African Medical Journal. 2011;101(11):809-13.

77. Singh AL, Jamal S. Unhealthy Cooking and Prevalence of Tuberculosis in Indian Women: A Case Study. Journal of Environmental Protection. 2012;Vol.03No.07:9.

78. Patel V, Mirsaeidi M, Schweitzer M, Foster A, Salem A, Kumar N. Does indoor air pollution enhance the risk of tuberculosis infection? B63 LATENT TUBERCULOSIS INFECTION AND EPIDEMIOLOGY OF TUBERCULOSIS DISEASE: American Thoracic Society; 2017. p. A3983-A.

APPENDICES

Appendix 1 Ethical clearance certificate

UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG

R14/49 Dr Victor Williams

HUMAN RESEARCH ETHICS COMMITTEE (MEDICAL)
CLEARANCE CERTIFICATE NO. M181005

NAME: (Principal Investigator)	Dr Victor Williams
DEPARTMENT:	School of Public Health Division of Epidemiology and Biostatistics Medical School University
PROJECT TITLE:	Prevalence of tuberculosis using the tuberculin skin test in two Human Immunodeficiency Virus endemic municipalities in South Africa
DATE CONSIDERED:	26/10/2018
DECISION:	Approved unconditionally
CONDITIONS:	
SUPERVISOR:	Dr K Otwombe
APPROVED BY:	Dr CB Penny, Chairperson, HREC (Medical)
DATE OF APPROVAL:	20/12/2018

This clearance certificate is valid for 5 years from date of approval. Extension may be applied for.

DECLARATION OF INVESTIGATORS

To be completed in duplicate and **ONE COPY** returned to the Research Office Secretary on 3rd floor, Phillip V Tobias Building, Parktown, University of the Witwatersrand, Johannesburg.
I/We fully understand the conditions under which I am/we are authorised to carry out the above-mentioned research and I/we undertake to ensure compliance with these conditions. Should any departure be contemplated from the research protocol as approved, I/we undertake to resubmit to the Committee. **I agree to submit a yearly progress report**. When a funder requires annual re-certification, the application date will be one year after the date of the meeting when the study was initially reviewed. In this case, the study was initially reviewed in **October** and will therefore reports and re-certification will be due early in the month of **October** each year. Unreported changes to the application may invalidate the clearance given by the HREC (Medical).

Principal Investigator Signature — Date: 8th January, 2019

PLEASE QUOTE THE PROTOCOL NUMBER IN ALL ENQUIRIES

Appendix 2 Plagiarism declaration form



SENATE PLAGIARISM POLICY: APPENDIX ONE

I **Victor Murphy Williams** (Student number: **1832945)** am a student registered for the degree of **MSc Epidemiology** in the academic year **2019**

I hereby declare the following:

- I am aware that plagiarism (the use of someone else's work without their permission and/or without acknowledging the original source) is wrong.
- I confirm that the work submitted for assessment for the above degree is my own unaided work except where I have explicitly indicated otherwise.
- I have followed the required conventions in referencing the thoughts and ideas of others.
- I understand that the University of the Witwatersrand may take disciplinary action against me if there is a belief that this is not my own unaided work or that I have failed to acknowledge the source of the ideas or words in my writing.
- I have included as an appendix a report from "Turnitin" (or other approved plagiarism detection) software indicating the level of plagiarism in my research document.

Signature: _____ Date: 12th August 2019

Appendix 3 Approval from PHIRST study principal investigator

NATIONAL HEALTH LABORATORY SERVICE
NATIONAL INSTITUTE FOR COMMUNICABLE DISEASES

Centre for Respiratory Diseases and Meningitis (CRDM)
1 Modderfontein Road, Sandringham, 2031 Tel: +27 (0)11 386 6593; Fax: +27 (0)11 882 9979

3 December 2018

Prof Cheryl Cohen
Centre Head: Centre for Respiratory Diseases and Meningitis
NICD
1 Modderfontein Road
Sandringham

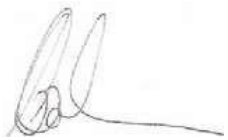
Human Research Ethics Committee (Medical)
Wits Health Consortium
8 Blackwood Road
Parktown

Re: master student permission to analyze data from the PHIRST project: reference 150808.

Title: A Prospective Household observational cohort study of Influenza, Respiratory Syncytial virus and other respiratory pathogens community burden and Transmission dynamics in South Africa (The PHIRST Study).

I have reviewed Victor Williams Master's project protocol (student number: 1832945) and given him permission to continue with the analysis, once he has HREC approval.

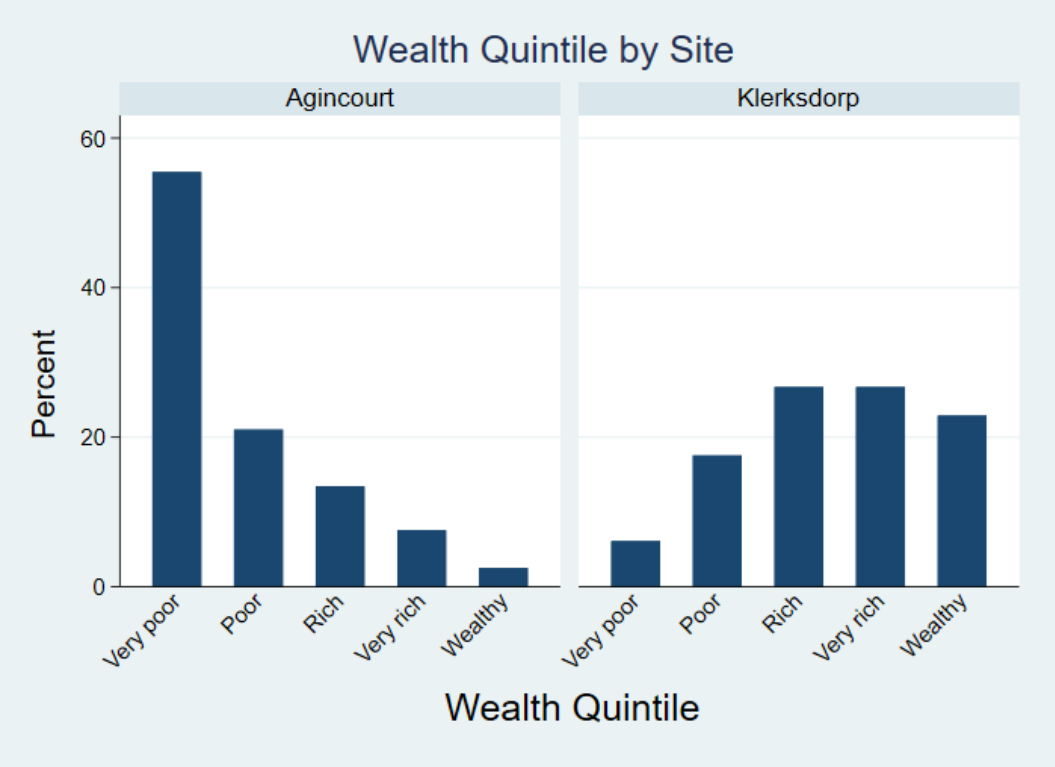
Yours Sincerely



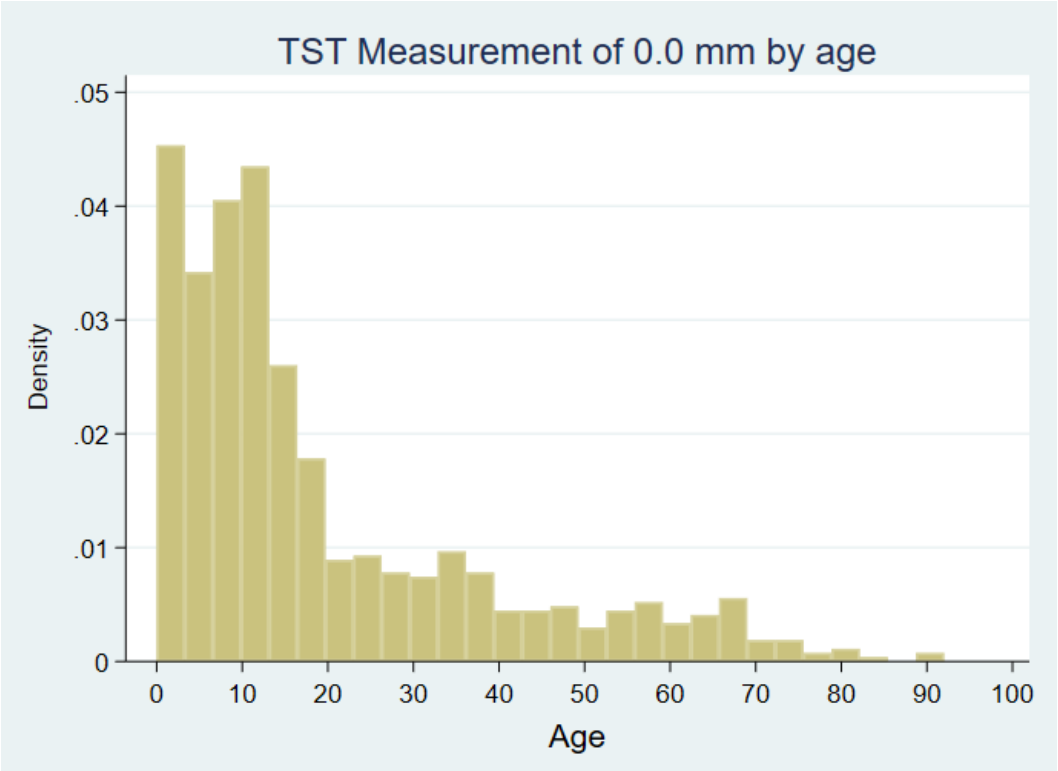
Prof Cheryl Cohen

Principal Investigator PHIRST study

Appendix 4 Graphical description of wealth quintile by site



Appendix 5 Graphical description of 0 mm TST measurement by age



Appendix 6 Smoothed graphs describing TST-Positive prevalence

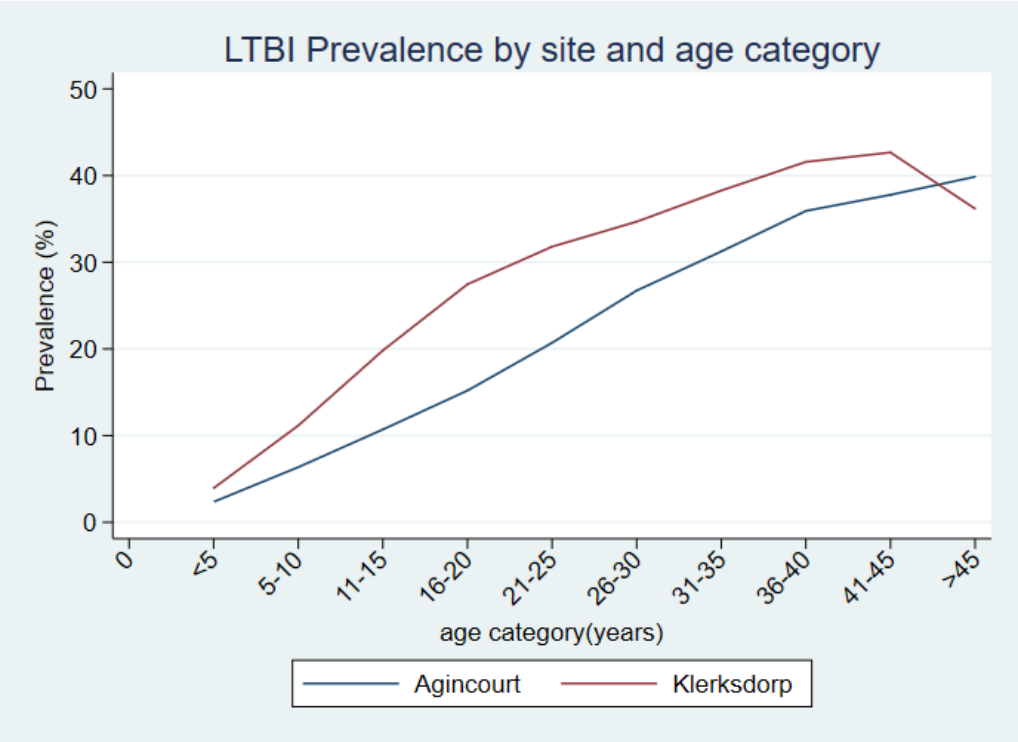


Figure 3.3: TST-Positive prevalence by site and age category

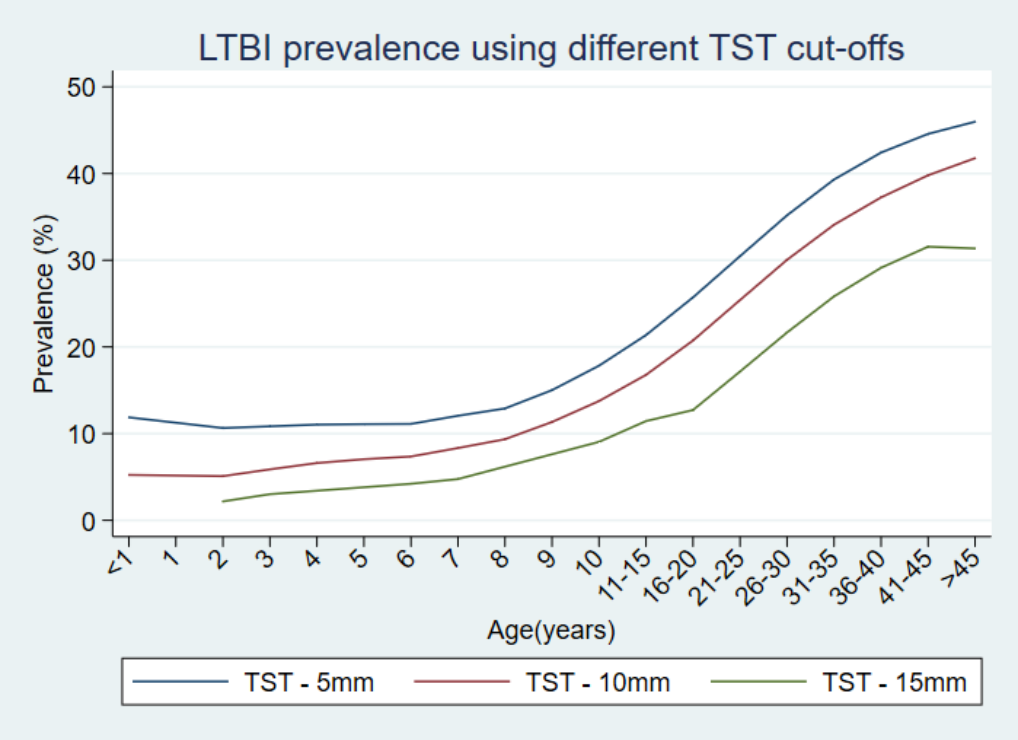


Figure 3.5: TST-Positive Prevalence at 5mm, 10mm and 15mm TST cut-off by age

Appendix 7 Logistic model of individual-level factors which predict TST-Positivity within households in Klerksdorp

Table for Participants ≥16 years of age

Variable	Univariate analysis		Multivariate analysis	
	Odds (95% CI)	P-Value	Odds (95% CI)	P-Value
Sex				
Male	1	-	1	
Female	0.79 (0.50, 1.26)	0.328		
Age (yrs)	1.002 (0.99, 1.02)	0.650	1.002 (0.99, 1.01)	0.802
BMI	0.99 (0.97, 1.03)	0.878		
BMI Category				
Underweight	1	-		
Normal	2.02 (0.75, 5.39)	0.162		
Overweight	1.86 (0.66, 5.26)	0.242		
Obese	1.53 (0.56, 4.16)	0.407		
HIV Status:				
Negative	1	-	1	
Positive	0.60 (0.35 1.04)	0.071	0.67 (0.38, 1.18)	0.162
Unknown	0.74 (0.29, 1.87)	0.519	0.90 (0.35, 2.33)	0.832
Past TB History				
No	1	-	1	
Yes	1.19 (0.50, 2.82)	0.686	1.34 (0.55, 3.30)	0.518
Present Cough				
No	1	-	1	
Yes	1.39 (0.58, 3.36)	0.463		

Hosmer-Lemeshow test statistic = 12.08, p=0.1477

Table for Participants ≤15 years of age

Variable	Univariate analysis		Multivariate analysis	
	Odds (95% CI)	P-Value	Odds (95% CI)	P-Value
Sex				
Male	1			
Female	0.89 (0.42, 1.89)	0.770		
Age (yrs)	1.16 (1.05, 1.28)	0.003	1.14 (1.01, 1.28)	0.037
Upper-arm Circum. (cm)	1.11 (1.02, 1.20)	0.012	1.04 (0.94, 1.16)	0.400
HIV Status:				
Negative	1		1	
Positive	4.97 (1.13 21.93)	0.034	4.08 (0.89, 18.77)	0.071
Unknown	-			
Present Cough				
No	1			
Yes	0.66(0.08, 5.22)	0.690		

Hosmer-Lemeshow test statistic = 7.14, p=0.5216

Appendix 8 Logistic models of individual predictive factors based on HIV status

Table for Negative HIV status (Age ≤15 years)

Variable	Univariate analysis		Multivariate analysis	
	Odds (95% CI)	P-Value	Odds (95% CI)	P-Value
Site				
Agincourt	1	-	1	
Klerksdorp	1.89 (1.01, 3.55)	0.046	1.82 (0.96, 3.44)	0.066
Sex				
Male	1	-	1	
Female	0.80 (0.43, 1.48)	0.484		
Age (yrs)	1.13 (1.05, 1.22)	0.002	1.13 (1.02, 1.24)	0.014
Upper arm circumference	1.07 (1.001, 1.142)	0.047	1.01 (0.93, 1.09)	0.885
Present Cough				
No	1	-	1	
Yes	0.66 (0.09, 5.11)	0.692		

Hosmer-Lemeshow test statistic = 5.70, p = 0.6808

Table for Negative HIV status (Age ≥16 years)

Variable	Univariate analysis		Multivariate analysis	
	Odds (95% CI)	P-Value	Odds (95% CI)	P-Value
Site				
Agincourt	1	-	1	
Klerksdorp	1.64 (1.08, 2.49)	0.019	1.58 (1.04, 2.41)	0.032
Sex				
Male	1	-	1	
Female	0.79 (0.51, 1.20)	0.267	0.79 (0.51, 1.22)	0.302
Age (yrs)	1.01 (0.99, 1.02)	0.055	1.01 (0.99, 1.02	0.058
BMI	1.01 (0.98, 1.04)	0.346		
BMI Category				
Underweight	1	-		
Normal	1.49 (0.59, 3.74)	0.401		
Overweight	1.45 (0.55, 3.85)	0.451		
Obese	1.58 (0.62, 4.02)	0.342		
Past TB History				
No	1	-	1	
Yes	0.65 (0.17, 2.45)	0.525		
Present Cough				
No	1	-	1	
Yes	1.61 (0.62, 4.19)	0.326		

Hosmer-Lemeshow test statistic = 10.36, p = 0.2405

Table for those who were HIV-infected at the time of TST testing (Age ≥16 years)

Variable	Univariate analysis		Multivariate analysis	
	Odds (95% CI)	P-Value	Odds (95% CI)	P-Value
Site				
Agincourt	1	-	1	
Klerksdorp	0.87 (0.43, 1.74)	0.697		
Sex				
Male	1	-	1	
Female	0.89 (0.37, 2.14)	0.792		
Age (yrs)	1.02 (0.99, 1.05)	0.085	1.02 (0.99, 1.05)	0.099
BMI	0.99 (0.94, 1.03)	0.553		
BMI Category				
Underweight	1	-		
Normal	0.52 (0.12, 2.21)	0.377		
Overweight	0.58 (0.13, 2.54)	0.467		
Obese	0.49 (0.11, 2.08)	0.330		
Past TB History				
No	1	-	1	
Yes	2.21 (0.75, 6.52)	0.151	2.15 (0.72, 6.42)	0.169
Present Cough				
No	1	-	1	
Yes	1.16 (0.21, 6.59)	0.865		

Appendix 9 Multi-level modeling (three levels) of individual risk factors of TST-Positivity considering clustering within households and site

Table for participants Aged ≥16 years

Variable	Univariate analysis		Multivariate analysis	
	Odds (95% CI)	P-Value	Odds (95% CI)	P-Value
Sex				
Male	1	-	1	
Female	0.87 (0.52, 1.45)	0.593		
Age (yrs)	1.02 (1.01, 1.03)	0.007	1.02 (1.004, 1.03)	0.012
BMI	1.02 (0.98, 1.05)	0.370		
BMI Category				
Underweight	1	-		
Normal	1.34 (0.47, 3.83)	0.586		
Overweight	1.29 (0.42, 3.93)	0.658		
Obese	1.53 (0.52, 4.53)	0.440		
HIV Status:				
Negative	1			
Positive	0.83 (0.45, 1.51)	0.531	0.85 (0.45, 1.58)	0.602
Unknown	0.88 (0.26, 2.99)	0.844	1.08 (0.32, 3.70)	0.897
Past TB History				
No	1	-	1	
Yes	1.65 (0.52, 5.23)	0.393	1.59 (0.48, 5.19)	0.446
Present Cough				
No	1	-	1	
Yes	1.59 (0.50, 5.07)	0.430		

Table for participants Aged ≤15 years

Variable	Univariate analysis		Multivariate analysis	
	Odds (95% CI)	P-Value	Odds (95% CI)	P-Value
Sex				
Male	1	-		
Female	0.63 (0.30, 1.31)	0.215		
Age (yrs)	1.16 (1.06, 1.26)	0.001	1.15 (1.03, 1.28)	0.019
Upper arm circumference	1.09 (1.01, 1.18)	0.031	1.001(0.90, 1.11)	0.980
HIV Status:				
Negative	1		1	
Positive	2.65 (0.57, 12.41)	0.216	1.97 (0.41, 9.39)	0.396
Unknown	0.28 (0.03, 2.53)	0.256	0.40 (0.04, 3.82)	0.429
Present Cough				
No	1	-		
Yes	0.39 (0.03, 4.81)	0.464		