LATENT TUBERCULOSIS INFECTION PREVALENCE, SPATIAL CLUSTERING AND RISK FACTORS IN A SOUTH AFRICAN URBAN INFORMAL SETTLEMENT

Jabulani R. Ncayiyana

A dissertation submitted to the faculty at the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Epidemiology in the Gillings School of Global Public Health.

Chapel Hill
2015

Approved by:

Annelies Van Rie

Daniel Westreich

Audrey Pettifor

Michael Emch

Eustasius Musenge

ABSTRACT

Jabulani R. Ncayiyana: Latent Tuberculosis Infection Prevalence, Spatial Clustering and Risk Factors in a South African urban informal settlement
(Under the direction of Annelies van Rie)

This dissertation investigated the burden spatial clustering and risk factors of latent tuberculosis infection (LTBI), in a South African urban informal settlement. Using data from a large community-based household survey with random sampling and from the 2011 South African census as disseminated by Statistics South Africa (STATSSA), we estimated the prevalence of LTBI in the general population, the annual risk of infection (ARI) in children, and investigated individual-, household- and neighborhood-level factors associated with LTBI (paper 1). We assessed spatial heterogeneity of LTBI prevalence and the association between community-level factors and LTBI clusters (paper 2).

In paper 1, we observed that the overall prevalence of LTBI was 34.3% (95% CI, 30% – 39%), the annual risk of infection among children age 0-14 years was 3.1% (95% CI: 2.1 - 5.2). In multivariable logistic regression analysis, LTBI was associated with age, male gender, marital status, and higher socio-economic status.

In paper 2, we investigated the spatial clustering and spatial heterogeneity of LTBI prevalence and predictive community-level factors. One statistically significant cluster of high LTBI prevalence was found using the spatial scan statistic. Higher socio-economic status (SES) was associated with higher LTBI prevalence in both a non-spatial regression model and a geographically weighted regression (GWR) model. However, only a small part of the spatial

heterogeneity in LTBI prevalence was explained by variation in community-level SES, suggesting that further research is needed to better understand the determinants of LTBI in such settings. Overall, this dissertation suggests that spatial analysis of LTBI can identify clusters within a single community and that LTBI prevalence is not associated with HIV status but may be associated with higher SES, in contrast to the well-established association between TB disease, HIV, and poverty.

In loving memory of late parents, Fulumaka and Ntombizodwa Ncayiyana. For teaching me "ukuthi umuntu ngumuntu ngabantu"

ACKNOWLEDGEMENTS

I would like to acknowledge and thank Dr. van Rie, the chair of my dissertation committee. Throughout my program, she has been an exceptional advisor and mentor. She has always been supportive and encouraging and I am extremely appreciative of her willingness to work with me throughout this exploration and research. I am especially grateful for her support, patience and understanding with my personal struggles with my health and academic programs.

I would also like to thank the members of my dissertation committee: Dr. Emch, Dr. Westreich, Dr. Pettifor and Dr. Musenge for their time, expertise, guidance, helpful suggestions and comments, and support throughout this research program.

In addition, I would also like to acknowledge and thank Dr. Charles van der Horst, for giving me an opportunity to pursue a PhD program in Epidemiology in one of the best schools in USA. He was supportive of my interest and application to this program at UNC and was always encouraging and supportive throughout this journey.

A special thank you to Nancy Colvin, for all your support and encouragement through difficult times.

Thanks to my classmates in the UNC Department of Epidemiology. Especially Maganiso Chagomerana, Emily Smith, Liz Cromwell, Joann Gruber, Nalyn Siripong, Stephanie de Long, William Miller, Shakia Hardy for always being there for me in a hour of need in and out of the classroom.

I would also like to thank Fogarty International Center (NIH) (grant number TW007373) and South African National Research Foundation (NRF) for the support and funding of my

program. The Diepsloot Household Mapping survey was funded by USAID (award number AID-674-A-12-00033, PI: Dr. Jean Bassett).

Finally, I am grateful to the inhabitants of Diepsloot for their participation, the survey teams (study nurses, lay HIV counsellors, drivers and interns), Clinical HIV Research Unit (CHRU) and the Witkoppen Health and Welfare Centre research team for their contributions.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ARI	annual risk of infection
ART	anti-retroviral treatment
CHWs	community health care workers
DOH	department of health
GIS	geographic information system
GPS	global positioning system
GWR	geographic weighted regression
HCWs	health care workers
TB	tuberculosis
HBCs	high TB burden countries
LTBI	latent tuberculosis infection
OLS	ordinary least squares
SES	socio-economic status
IGRAs	interferon-gamma release assays
IPT	isoniazid preventive therapy
NTM	non-tuberculous mycobacteria
PPD	purified protein derivative
SADHS	South African Demographic and Health Surveys
HIV	human immune-deficiency syndrome
WHO	World Health Organization
DOTS	directly observed therapy short-course
MDGs	Millennium Development Goals
TST	tuberculin skin test
BCG	Bacille Calmette Guerin

ESAT-6	early secreted antigenic target 6
CFP-10	culture filtrate protein

CHAPTER 1: BACKGROUND

1. Burden of active TB disease and Latent TB Infection (LTBI)

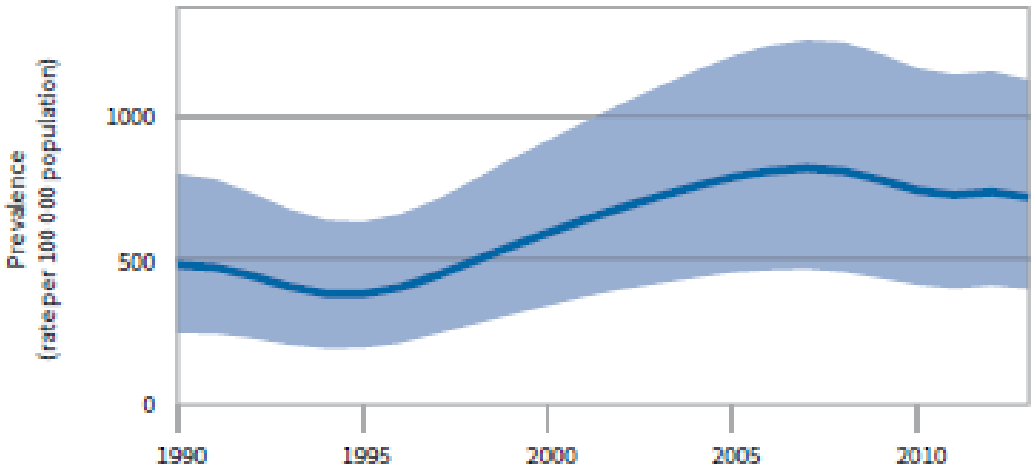
1.1 Burden of TB

Tuberculosis (TB) remains a significant public health problem worldwide, especially in sub-Saharan Africa: In 2013, there was an estimated 9 million new cases of TB and 2.6 million, with is over one quarter (29%) of all incident cases reported in sub-Saharan Africa.[1] Of the 22 high TB burden countries (HBCs) that account for about 80% of the world's TB cases, nine are in sub-Saharan Africa region.[1] Of the 9 HBCs countries in the sub-Saharan Africa region, more than half (5) are in southern Africa.[1] Overall TB incidence has been steadily decreasing in sub-Saharan Africa region, however TB incidence rates vary widely between regions and countries, with around 1350 per 100 000 people in Swaziland, and fewer than 100 per 100 000 population in western African countries such as Burkina Faso, Ghana and Togo.[2]

South Africa has the sixth highest burden of TB burden in the world, with 450 000 (410,000 – 520,000) of incident TB cases reported in 2013,[1] Over the last two decades, annual incidence of TB has increased drastically from 300 per 100 000 population in 1990 to 860 per 100 000 population in 2013. TB prevalence has also increased from 475 per 100 000 population to 715 per 100 0000 population over the same period (Figure 1).[1] These estimates conceal the individual-level disparities and geographical variations of TB burden in South Africa, between provinces[3, 4] and even between communities. Two South African Demographic and Health Surveys (SADHS) showed that the burden of TB varies greatly by age, gender, rural or urban area with higher prevalence in urban communities[5-7] compared to rural communities.[7]

Community-based prevalence surveys have shown that TB burden varies significantly and that the TB prevalence in some communities was 2-fold that of national estimates. [5, 6, 8-11]

Figure 1.1 Estimated TB prevalence trends in South Africa, 1990–2013.



Source: (WHO, 2014)[1]

1.2 TB/HIV co-infection burden

There is a significant public health burden of HIV-associated TB; especially in sub-Saharan Africa. In 2013, 1.1 million (12%) of the 9 million people who developed TB worldwide were HIV-positive.[1] Overall, 870, 000 (790 000-960 000) of TB cases were estimated to be co-infected with HIV were reported from sub-Saharan Africa in 2013.[1] These cases accounted for 79% of TB/HIV co-infected cases reported worldwide, which increased by 10% increases from 69% TB/HIV co-infected cases reported in 2011.[1, 2] Epidemiological studies in countries with high HIV prevalence have also shown that spatial and temporal variation in TB incidence is strongly associated with the prevalence of HIV infection.

South Africa alone accounts for almost one-quarter of the global burden of HIV-associated TB. Among the 22 high-burden countries, South Africa has the highest number of HIV-co-infected TB cases with 60% of new TB cases co-infected with HIV.[1] Whilst most of TB

burden in South Africa is associated with HIV infection, there is an indication that HIV infection alone doesn't explain the TB burden in South Africa. KwaZulu-Natal, with the highest HIV sero-prevalence rate nationally, has the lowest TB prevalence while Western Cape with the lowest HIV sero-prevalence rate nationally, has the highest TB prevalence.[3, 4]

1.3 Limitations of current TB control strategy

Since WHO declared TB a global emergency in 1993, the directly observed therapy short-course (DOTS) strategy has been the key public health intervention global TB control.[12, 13] The performance targets of tuberculosis (TB) control programs adopted by WHO are to detect 70% of new sputum smear-positive cases of TB and cure rates of more than 85% of detected cases.[14] The Stop TB Partnership has set additional targets related to the Millennium Development Goals (MDGs): to reverse TB incidence, to halve TB prevalence and mortality between 1990 and 2015.[15, 16] It was predicted that if the above targets were reached by 2005, it would be possible to achieve the TB-related MDGs.[17-20]

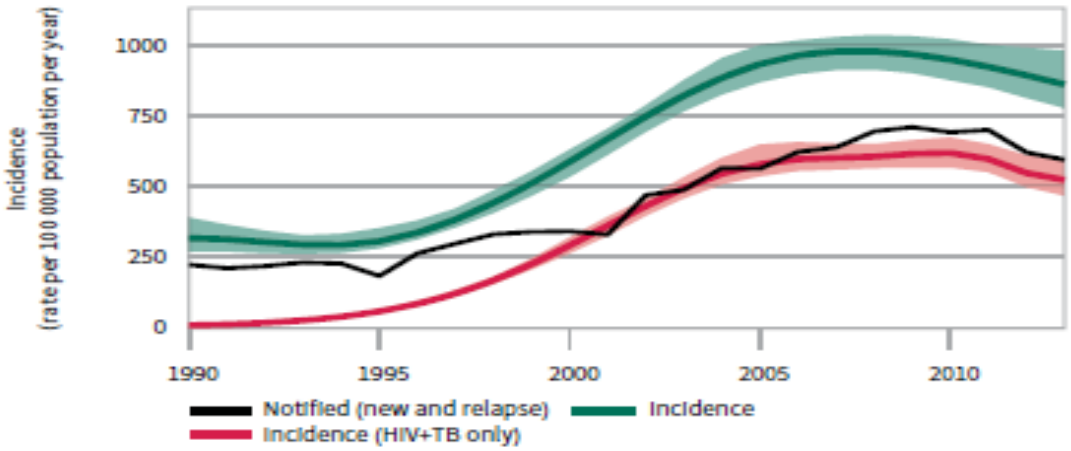
The target of an 85% treatment success rate for sputum smear-positive cases was first achieved in 2007 and case detection rate increased substantially between 1995 and 2008.[16, 21] However, TB incidence has only been falling at a very slow rate of less than 1% per year since 2004.[20, 22] Moreover, among the 22 HBCs, 11 are not on track to reduce incidence, prevalence and mortality in line with targets.[1] Although DOTS strategy has been effective in most regions of the world, resulting in a sustained downward trend in global TB incidence, it has failed to bring down the prevalence of TB in eastern Europe and Africa especially in the southern Africa where HIV prevalence is highest.[23, 24]

The impact of HIV on TB control was predicted by Karel Styblo who developed DOTS strategy for WHO in 1990.[25, 26] Styblo warned that DOTS strategy will not prevent an increase in TB in regions hardest hit by HIV if the prevalence of LTBI is high. DOTS strategy does not prevent the progress of LTBI to active TB, which is where HIV primarily exerts its strongest impact.[25, 26] Another major impediment to achieving targets of TB control is changing population

risk factors.[27-29] There is a growing recognition that the rates of TB burden decline might also be more strongly related to social and economic factors and general population health than the performance of national tuberculosis control programs.[27, 28, 30, 31]

Despite boasting about a well-organized and high coverage of DOTS strategy, South Africa has consistently failed to achieve these targets. Treatment success rate for sputum smear-positive cases was 77%, and case detection rate was 69%, falling substantially short of the 85% target.[1] Achievement of MDGs for TB is unlikely in South Africa, though the incidence of TB has started to slightly decrease in 2013 (Figure 2), South Africa is not track on meeting 50% reduction of TB prevalence and mortality.[1] It is evident that the current TB control strategy had very limited impact on the burden of TB in South Africa. TB epidemic in South Africa is thought to be fueled by very high rates of LTBI and high ongoing transmission rates of infection.[22]

Figure 1.2 Estimated TB incidence rates, estimated incidence rates of HIV-positive TB and TB notification rates in South Africa, 1990–2013.



Source: (WHO, 2014)[1]

1.4 Latent M. tuberculosis infection (LTBI): burden and risk factors

Individuals with LTBI are infected with *M. tuberculosis*. They do not have symptoms of

TB, are not infectious, but are at risk of progressing to active disease and thus becoming infectious.[22, 32-35] Without HIV co-infection, the average lifetime risk of infected individuals developing tuberculosis is 5 to 20%, the highest risk being within the first five years of infection.[36] With HIV co-infection, the annual risk of progressing to active is 5-15%,[37, 38] with average lifetime risk as high as 50%.[39] While antiretroviral treatment reduces the risk of progressing from latent infection to active disease, it does not eliminate this risk and the risk in individuals on ART remains higher than the risk in HIV negative individuals.[40] The risk of developing TB following infection also changes with age. Infants and young children up to the age of five years who are infected with *M. tuberculosis* are at high risk.[41-43]

An estimated 2 billion persons worldwide have latent tuberculosis infection (LTBI) and approximately 200 million are at risk of progression to active TB disease during their lifetime.[17, 44, 45] The burden of LTBI varies greatly across the world, in developing countries, LTBI estimates have ranged from 5% in Kenya to 60% in Ethiopia.[46-49] In North America and Europe, the prevalence of LTBI in the general population is less than 10%,[50, 51] but high burden of LTBI occurs among high-risk populations.[50, 52, 53] One such high-risk population is household contacts of TB cases. A systematic review of 203 studies of LTBI among household contact reported great geographic LTB prevalence of 51.5% (95% CI= 47.1–55.8%) in low- and middle-income settings compared to 28.1% (95% CI=24.2–32.4%) in high-income settings.[54] Another high-risk population are people living with HIV.[55] Of all people living with HIV worldwide, it is estimated that one-third are co-infected with LTBI, varying from 14% in Europe to 46% in Southeast Asia, and more than 50% in sub-Saharan Africa.[55, 56]

Estimates of LTBI prevalence in South Africa are scarce. In early 1980s, national Tuberculin Skin Test (TST) surveys, reported an LTBI prevalence of 10-20%.[57] Since then, no national surveys have been performed recently. Recent LTBI studies in South Africa have focused on smaller regions or communities. In a high-density predominantly black townships of Cape Town, up to 50% of 15-year-olds[58] and in a large urban gold mining populations in

Johannesburg, 77%–89% of adults had evidence of LTBI.[59] In an urban township of Johannesburg, the LTBI prevalence was 33% (95% CI [21%–32%].[60] A recent study in a black township of Cape Town has reported the prevalence of LTBI to be 52.7% among HIV-infected patients.[61] This suggests that there is a heterogeneous in the burden of LTBI with South Africa.

1.5. Annual risk of infection with M. tuberculosis

A measure related to the prevalence of LTBI is the annual risk of TB infection (ARI), which is an averaged measure of risk of TB infection over the lifetime of individuals.[62] ARI is typically measured in school-age children and used as the indicator of recent transmission of *M. tuberculosis* in the community.[58] In low burden TB countries ARI is less than 1% while recent ARI estimates in sub-Saharan range from 1.5-4%.[63, 64] South Africa had very high ARI (5-8%) in early 1980s national TST surveys.[57] More recent studies in one of the high TB burden setting of South Africa reported ARI of 4%.[58, 65]

1.6 Diagnosis of LTBI

Because LTBI is the sub-clinical stage of TB disease continuum, culturing and isolation of *M. tuberculosis* bacilli is not possible during LTBI stage.[32, 66, 67] Instead, the diagnosis of LTBI is based on immunological response to *M. tuberculosis*.[32, 66] Tuberculin skin test (TST) and interferon-gamma release assays (IGRAs), are the test currently available used to diagnose LTBI.[66]

For over 100 years, the hallmark of LTBI diagnosis has been TST, which is also called the Mantoux skin test.[32, 66, 67] TST requires the intradermal injection of purified protein derivative (PPD).[66] TST measures cell-mediated hypersensitivity to tuberculin PPD, which contain a mixture of the antigens found in several species of mycobacteria.[66] The TST's result is recorded as the diameter size of transverse induration in millimeters (mm) 48-72 hours after TST has been administered.[66] Interpretation of the TST result varies depending on the prevalence of and the risk for progression to TB in different groups.[66]

TST has several important limitations. TST has poor sensitivity (75-90%), especially in immnuno-compromised populations.[66] To address this limitation, in HIV-infected individuals, a positive TST is defined as an induration of at least 5 mm.[68] TST specificity range from 70-95% with lower specificity in Bacille Calmette Guerin (BCG)-vaccinated individuals, especially in the first years following initial BCG vaccination or in individuals who have received repeated BCG vaccinations.[66] There is also some cross reactivity with non-tuberculous mycobacteria (NTM). Despite its limitations, TST continues to be recommended and used as there is a moderate to strong association between TST positivity and risk of active TB during follow-up.[66] A review of 11 studies showed that the largest TST positivity reactions are associated with an increased risk of TB than the smallest TST positivity reactions. The risk ratio of TB among the largest TST positivity reactors ranged from 2.2 to 26.3.[69] There is also strong evidence showing the benefits of treating TST positive individuals in reducing the risk of progressing from latent infection to active TB.[66] A study showed a 62% reduction in risk if active TB among HIV-infected patients treated with IPT.[70] TST therefore remains a useful tool, both for epidemiologic research and the control and prevention of clinical tuberculosis.[71]

Alternative test for LTBI diagnosis are IGRAs, which are available as commercial assays.[32, 66] IGRAs are in vitro tests of whole blood or mononuclear cells that are based on Interferon-gamma release after T-cell stimulation by *M. tuberculosis*–specific proteins. Two of the most commonly used *M. tuberculosis*–specific proteins are early secreted antigenic target 6 (ESAT-6) and culture filtrate protein 10 (CFP-10), which are absent from BCG vaccine strains and most NTM.[66, 67] The positive result is based on either the concentration of Interferon-gamma or the number of Interferon-gamma producing cells (spots) depending on the type of the assay used.[67]

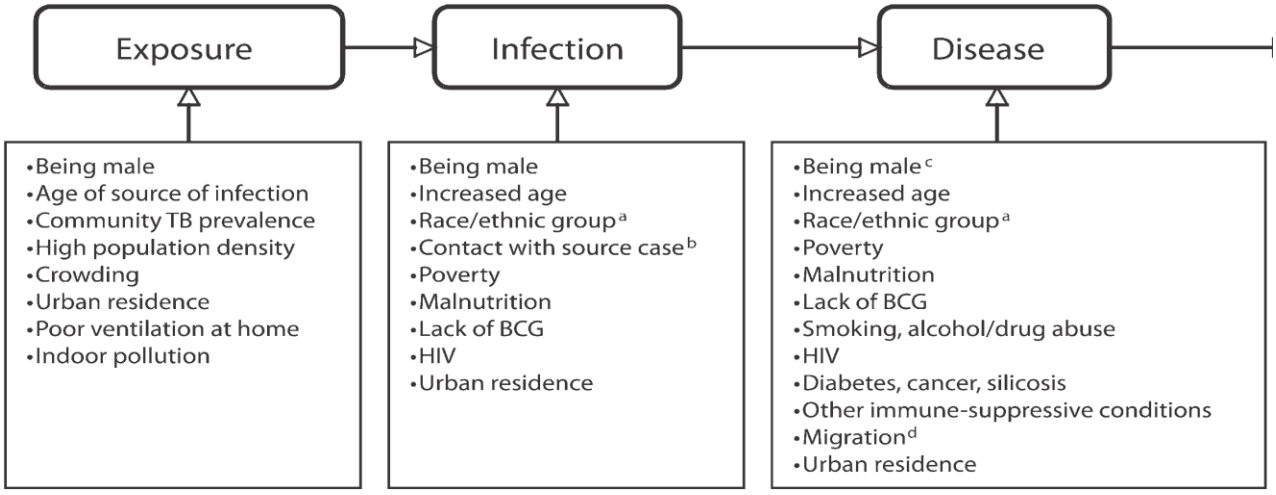
IGRAs have slight advantage over TST in terms of higher sensitivity (75-95%) and specificity (90-100%), and specificity is not affected by BCG vaccination history.[66, 67] However, they also share similar limitations with TST as IGRAs' sensitivity is poorer in immune-

compromised populations and, similar to TST, IGRAs cannot distinguish between LTBI and active TB.[66, 67] Despite their promise, IGRAs have not replaced TST is the main diagnosis test for LTBI, especially in poor resourced settings.[67] This is partly due to their high costs and need for the laboratory infrastructure, which is not available in most of poor resourced settings.[72] IGRAs are not recommended by WHO for screening individuals who will be eligible to receive IPT.[73]

1.7 Risk factors for LTBI

In order to appreciate the importance of risk factors for LTBI and rationale for mitigating against them, it is necessary to first understand that the development of TB is a two-stage phenomenon.[74] First, a person becomes infected after being exposed to *M. tuberculosis* bacilli and secondly, infected persons may develop active TB after an interval ranging from few months to decades.[74, 75] Since the development of active TB is often distant from the acquisition of infection, some of the risk factors for LTBI are different from the risk factors for active TB.[74, 75] However, most of the risk factors are shared for both LTBI and active TB as illustrated in Figure 3.[76] Furthermore, many factors are shared by communities across the world. Common risk factors include close contact with a TB case, older age, HIV infection, and race/ethnicity. Other factors may be specific to certain settings. For example, being foreign born is a main risk factor for LTBI in developed countries but less so in high burden countries,[50, 53, 77] while HIV and urban residence are relevant for developing countries.[77]

Figure 1.3 Risk factors for different stages of TB disease continuum.



(Hargreaves,2011)[76]

1.7.1 Close contact of TB case

It is well established that contacts of TB cases are at higher risk of LTBI than the general population.[78] Many studies across the world, have found that being a household contact of TB is the strongest predictor for LTBI.[48, 61, 79-82] A systematic review of 41 studies conducted in 17 countries (50% in Africa) reported a pooled LTBI prevalence of 51.8% (95% CI=50.9-52.8) among household contacts of TB cases.[78] The level of infectiousness of the index case (sputum positive vs negative case), physical proximity to the index case, and duration of contact with a TB case, modify the association between being a close contact and the risk of LTBI.[83] LTBI prevalence tends to be higher in contacts exposed to smear positive TB cases compared to smear negative ones.[84, 85] Studies in the Gambia found that an association between LTBI and physical proximity (OR=2.3, 95% CI=1.7-3.1),[75, 86] and prolonged exposure to a TB case (OR=2.4. 95% CI=1.7-3.3).[75] In a South African study conducted in high TB burden area found that close contacts had 2.5 times the odds of LTBI (OR=2.5, 95% CI=2.2-2.9) compared to those were not household contacts.[87]

1.7.2 Age

Children under 2 years acquire LBTI from the household index TB case, while, children older than 2 years, mostly acquire LTBI from the community.[9, 88] However, household index TB case remains an important source of infection for children up to 10 years of age.[89, 90] A study in Philippines among found that household contacts older than 5 years old had higher odds of LTBI (OR 3.17, 95%CI 1.43-7.01).[91] Studies in sub-Saharan Africa also reported increased age as independent risk factor for LTBI.[47, 75, 92, 93] This association may be due to two factors. First, children young than 2 year of age progress to active disease shortly after infection. Second, their immune responses to PPD is weak due to underdeveloped immunity. As a result young children may be misclassified as not having LTBI, resulting in an underestimation of the prevalence of LTBI in younger children.[89] A study among household contacts, found that compared with contacts 4 years of age or younger, contacts had increasing odds of LTBI with increasing age: OR 2.60 (95% CI: 1.32–5.01) among 5–14 years, OR 4.45 (95% CI: 1.98–9.97) in 15–24 years and OR = 7.16 (95% CI: 3.39–15.10) in those 25 years or more.[94] In South Africa, two studies found that increased age was associated with LTBI (OR=2.0, 95% CI=1.7-2.3).[87, 95]

1.7.3 HIV

HIV infection may not only increase the risk of progression from infection to disease but may also increase the risk of LTBI in those exposed *M. tuberculosis*.[96] HIV infection weakens cell-mediated immunity due to reduction and dysfunction of CD4 cells, which plays a critical role in the host immune defense mechanisms against *M. tuberculosis*.[96] HIV infected individuals tend to have higher risk of acquiring LTBI than HIV non-infected individuals.[54, 55] Epidemiological studies in high HIV burden settings investigating the association between HIV status and risk of LTBI have reported conflicting results. A cross-sectional study across 24 communities in Zambia and South Africa found that HIV infected individuals had lower odds of LTBI (OR=0.61; 95% CI=0.46–0.82).[85] A study among pregnant women in Tanzania did not

found an association between HIV status and LTBI.[97] A study among South African miners found that HIV-infected miners had lower odds of have LTBI (OR= 0.41, 95%CI =0.17-0.96).[59]

1.7.4 Gender

Tuberculin-testing surveys conducted during the 1950s and early 1960s show an age-dependent pattern of LTBI prevalence between men and women.[98] The prevalence of LTBI was equal between men and women until beginning in adolescence, after which male prevalence began to exceed female prevalence.[98] A community-based study in Ethiopia reported, that being male was associated with higher odds of LTBI (OR=1.8, 95% CI=1.2-2.7).[49] Other studies across sub-Saharan Africa have reported similar associations.[75, 93] In South Africa, studies have consistently reported the association between risk of LTBI and being male.[65, 87] A differential increase of social contacts is often presented as an explanation for this association.[65, 99]

1.7.5 Race or ethnic group

Poor and marginalized populations or ethnic minorities carry a disproportionately higher burden of LTBI.[100] Similar to gender, race/ethnicity risk for LTBI is not only biological but occurs within a social, economical and cultural context. Even in developed countries, there are racial disparities with regards to the risk of LTBI. In USA, blacks were 7 times as likely to have LTBI (OR=7.5, 95% CI=4.0–13.9).[50] A study in Laos found that ethnic minorities were 5 times as likely to have LTBI (OR: 5.4, 95% CI: 2.2-13.6).[79] A study in Tanzania found that certain ethnic groups were found to be less vulnerable to LTBI as compared to others. In South Africa, early study reported that LTBI vary by racial groups.[57] A recent study in South Africa reported higher odds of LTBI for blacks (OR=4.2, 95% CI=3.1–5.7) and mixed race group (OR=3.9, 95% CI=2.8–5.4).[87]

1.7.6 Poverty (SES)

It has long been established that TB is a disease of poverty, however there has been little attention paid to poverty or socio-economic status (SES) as the risk for LTBI.[28, 101] In recent

years, poverty/SES has been recognized as a one of important drivers, which need to be address in TB control.[16, 31] The prevalence of LTBI is higher in poor countries but, even within rich countries, individuals living in poverty tend to have higher burden of LTBI.[50] A study in the USA found that the odds of developing tuberculosis were 1.9 times higher in the poorest individuals than in the richest.[50] By contrast, studies from developing countries reported conflicting results. A study in Zambia found that higher SES, rather than lower, was associated with significantly higher odds of LTBI.[102] A study in Viet Nam found that higher SES, was associated with lower odds of LTBI (OR=0.88. 95% CI=0.80–0.97).[48] One possible explanation of this discrepancy is that the difference in how poverty/SES indicators are measured in the different regions may produce different pathways of association between poverty and LTBI.[103] For example, poverty is measured by household income, while in developing countries poverty/SES is usually measured by household items.[104]

1.7.7 Urban residence

Rapid urbanization witnessed in developing countries has also been shown to have influence on a person's susceptibility to infection.[105-107] Poor high-density urban settlements are often the breeding ground for infectious diseases.[107] Urban poor settlements are even more relevant for TB transmission since the factors that facilitate transmission tend to be more prevalent in these settings.[28, 108] In 19th-century Europe growing urbanization due to massive industrialization, resulted in the highest TB burden ever seen in modern Europe.[28, 108] Two the countries (India and China) with highest TB burden in the world, have undergone unprecedented urbanization in modern history.[109] In India, the risk of TB infection is higher in urban than in rural areas.[110] In sub-Saharan Africa, LTBI is highly prevalent among urban residents.[59, 111] A study in Viet Nam found that urban residence, was associated with higher odds of LTBI (OR=1.47, 95% CI=1.34–1.62).[48] Urbanization is confounded with other risk factors such as overcrowding, and poverty, malnutrition, lack of proper housing and access to health care.[28, 108] The worst crowding occurs in urban and peri-urban slum areas, due to rapid

urbanization and overpopulation.[112] In Thailand, a study found that people living in crowded households had 2.6 times (OR = 2.63, 95%CI = 1.18-5.85) the odds of LTBI.[113] A study in Zambia found that people living in crowded households had higher odds of LTBI (OR=3.0, 95% CI=1.2–7.4).[102]

1.8 Why LTBI is significant in TB control

The importance of LTBI is not limited to the individual, but is also an important public health consideration. Studies show that 5–20% of those with LTBI will develop active TB at some point in their lifetime, with the majority developing TB disease within 5 years of the initial infection.[36, 114] Therefore, LTBI burden represents a large reservoir of new TB cases. Particularly HIV-infected individuals and younger children with LTBI who are at highest risk of developing TB.[22, 43] Modeling studies have shown that screening and treatment of LTBI may play a key role in TB elimination.[115, 116] TB elimination is aimed at reducing the prevalence of LTBI, so that future cases of TB will be prevented.[117] Given the significant public health impact and costs that are associated with active TB disease, public health interventions aimed at reducing the burden of LTBI and risk of progression to active TB, are being implemented.[1, 2] The main intervention aimed at reducing the burden of TB is screening and provision of isoniazid preventive therapy (IPT) treatment to HIV-infected individuals and younger children with LTBI. However, its implementation though improving is still very poor in high TB burden countries.[1, 2]

In low TB burden countries, targeted testing for LTBI and provision of IPT has become key part of the TB control strategy as the focus has shifted to elimination. However, in high TB burden settings, DOTS is still the main strategy for national TB control programs.[22] Therefore, in high-burden settings, testing for LTBI and IPT is limited to selected high risk groups such as childhood household contacts and all people living with HIV infection.[22] As TB incidence decreases in most of these countries, countries may decide to expand testing for LTBI and IPT to other at-risk populations.

2. Spatial epidemiology of TB

2.1 Spatial clustering and patterns of TB burden

Transmission of infectious disease is closely linked to the concepts of spatial and spatial-temporal proximity, as transmission is likely to occur if the at-risk populations are close in a spatial and temporal sense.[118] The recognition that infectious diseases and their risk factors are spatial and temporal related is growing rapidly. Consequently, geographical information systems (GIS) and spatial analysis are increasingly being used to describe distribution and patterns of infectious disease.[119-123] The use of GIS and spatial analysis to describe the pattern of TB has gained momentum in recent years.[124-130] A study in India found significant hotspots of TB in three areas of the Almora district.[130] Studies conducted in China have found significant TB clustering in urban settings.[131-133] In West African country, spatial scan statistic was used to assess purely spatial clusters of TB in an urban setting.[134] In South African, few studies used GIS and spatial analysis to investigate TB burden patterns in a high-incidence area.[124, 128]

2.2 Factors associated with spatial patterns of TB

Clustering of a disease or disease hotspots occurrence is closely linked to the clustering of a risk factors.[118] This is especially true for TB since factors which facilitate transmission are largely influence by spatial characteristics such as migration, crowding, and poverty. A study in USA identified poverty, age, race, and foreign born as factors associated with TB incidence clusters.[135] In China, studies using spatial analysis, found migration, poverty/SES and housing type was associated TB hotspots.[133, 136-138] Studies done in Brazil found that TB incidence hotspots were associated with low SES.[139, 140] In Madagascar, spatial TB clustering was associated with low SES and migration.[129] In South Africa, an early study in a small urban (3.4 km^2) community showed that there was uneven spatial distribution of notified TB cases.[124] Another study in the same small urban community found that the clustering of TB notified cases were associated with unemployment, overcrowding and number of shebeens per enumerator sub-district.[128] Another study in rural community found a

low-risk cluster for incident TB was associated with patients living near the local health facility. The study did not find a high-risk cluster of incident TB.[141]

2.3 Application of spatial analysis for the control of TB

The success of public health intervention depends on a broad and accurate understanding of the risk factors that determine the occurrence of disease.[142] The challenge facing the TB control program is that the disease burden is not homogenous but varies geographically. There is growing evidence that the use of GIS and spatial analysis to inform TB control strategies may aid in more effectively reducing TB burden. A study in a Smith county, Texas in USA used GIS to identify neighborhoods with high TB burden that could be targeted for LTBI screening and IPT treatment.[143] The study reported a dramatic reduction of TB incidence in those neighborhoods, which were identified by GIS and targeted for intervention.[143] The results from study further demonstrated that the application of GIS and spatial analysis in TB control program is not only effective but also cost-effective.[143]

A mathematical model explored the impact of targeting TB hotspots in the community compared to targeting the whole community.[116] The study showed that achieving TB control targets in a hotspot containing 6% of a city's population can have similar impact on community TB incidence as achieving the same targets throughout the remaining community.[116] The projected impact of hotspot-targeted interventions depends strongly on the rate of TB transmission from cases in the hotspot to members of the general community. Previous studies have shown that TB transmission is not only limited within the household,[88, 144, 145] but significant amount of TB transmission occurs in other settings outside the household in high TB burden areas.[146, 147] These settings include crowded and poorly ventilated informal alcohol drinking places (sheebens/taverns), public transportation, community halls, and churches.[88, 144, 147] Therefore, TB transmission occurring in these settings is more likely to take place within geographically defined boundaries.[145, 146]

3. Limitations of Current Research

3.1 LTBI prevalence studies

Previous studies estimating the prevalence of LTBI, ARI and associated risk factors have provided a valuable insight on the burden of LTBI. However, these studies have key limitations. Most studies estimating prevalence of LTBI and its risk factors have been conducted on children and adolescents in schools.[87, 95, 148] While this is a convenient and easy population to study, studies done on children and adolescents may not represent the true burden of LTBI and transmission patterns at the community level. Studies have shown that the exposure and risk of LTBI is age-dependent.[47, 75, 92, 93] In young children, the main risk factor for LTBI is adult TB case in the household. In adults,[9, 88] exposure TB bacilli is not limited to the household with a significant amount of TB transmission occurs in other settings outside the household.[146, 147] Studies done in adults population have usually been performed in specific groups of adults such as health care workers (HCWs), miners or HIV infected individuals.[59, 149-153] These specific groups do not represent the general population and their risk factors for LTBI may be different.[59, 149-153] Studies estimating the prevalence of LTBI and ARI across all age groups in the general population may better estimate the true burden of LTBI in communities.

To date, many LTBI prevalence studies have been conducted at national level.[46, 48, 79, 154] This means that the reported estimate of LTBI prevalence will not give an insight of the geographic variations of LTBI prevalence at community level. Considering that TB control intervention are implemented at district and community level, national LTBI prevalence studies are not useful in identifying local high burden areas and the corresponding individual, household, and community level risk factors. Community-based LTBI prevalence studies are better positioned to inform targeted community-based interventions.

Despite the recognition that factors associated with LTBI burden are multifaceted (individual-, household- and community/neighborhood levels), there is limited number of studies assessing risk factors for TB using multilevel analysis.[155-157] A South African study using

multilevel analysis to explore individual-, household- and community-level factors associated with TB prevalence, found that community-level income inequality was independently associated with increased odds of TB prevalence (OR= 2.37, 95% CI: 1.59–3.53).[156] There is only one study to date of risk factors for LTBI using multilevel analysis. A study conducted in Pakistan using multilevel analysis found that both individual-level and household-level factors were associated with LTBI prevalence.[94] LTBI prevalence studies measuring the associations with risk factors at individual, household and community level are needed to identify those individual communities and areas within communities that need to be prioritized for the most effective interventions.

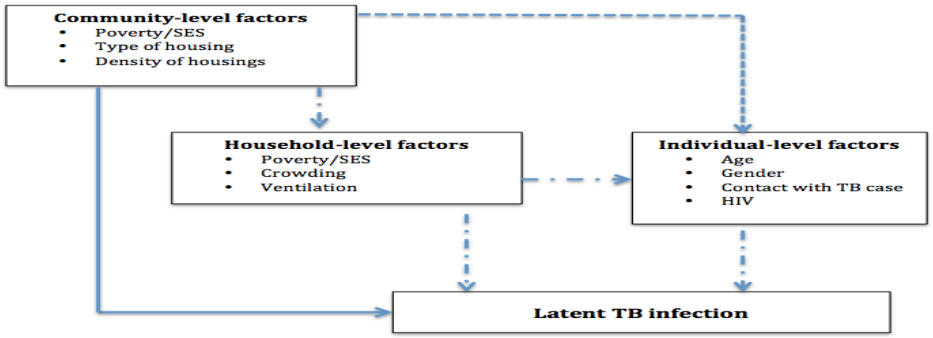
3.2 Potential application of spatial analysis for LTBI

The impact of any public health intervention is critically dependent on disease prevalence and factors driving the transmission in the population.[142] Considering the fact that prevalence of LTBI and its risk factors vary by geographical location, GIS may help identify local LTBI hotspots.[158] Furthermore, spatial analysis of risk factors associated with LTBI hotspots may help improve the targeting of scarce resources for public health interventions. Indeed, local targeted public health interventions have been shown to be more efficient and cost-effective than community-wide public health interventions.[116, 143] In addition, the success of any public health intervention depends on accurate understanding of the individual, socio-economic, and environmental factors that drive disease transmission.[142] Most studies using GIS and spatial analysis have investigated the spatial distribution of active TB and its risk factors.[124-128, 131, 136, 145] There are very few studies in South Africa investigating LTBI hotspots using spatial tools.[158] The study found that LTBI in younger children was strongly associated with presence of an adult case on the residential location.[158] A recent follow-up study in the same area reported similar findings.[159] Studies using GIS and spatial analysis to identify LTBI hotspots and their risk factors could be highly informative to the development of targeted public health interventions.

4. Conceptual framework of factors associated with LTBI in the proposed study

TB control efforts until recently were focused almost exclusively on the diagnosis and successful treatment of infectious TB cases. Growing awareness of the importance of social determinants of health in other areas has stimulated interest in the role of these determinants for TB.[160] We hypothesize that in addition to individual-level factors, household- and community-level factors will be associated with LTBI burden. Figure 4 illustrate possible pathways though which household- and community-level factors may affect LTBI prevalence. It has been previously shown that social interactions result in a substantial proportion of TB transmissions at the community level.[65, 99] While factors such crowding and poor ventilation facilitate TB transmission at the household level.[102, 113]

Figure 1.4 Conceptual framework of factors associated with LTBI



CHAPTER 2: METHODS

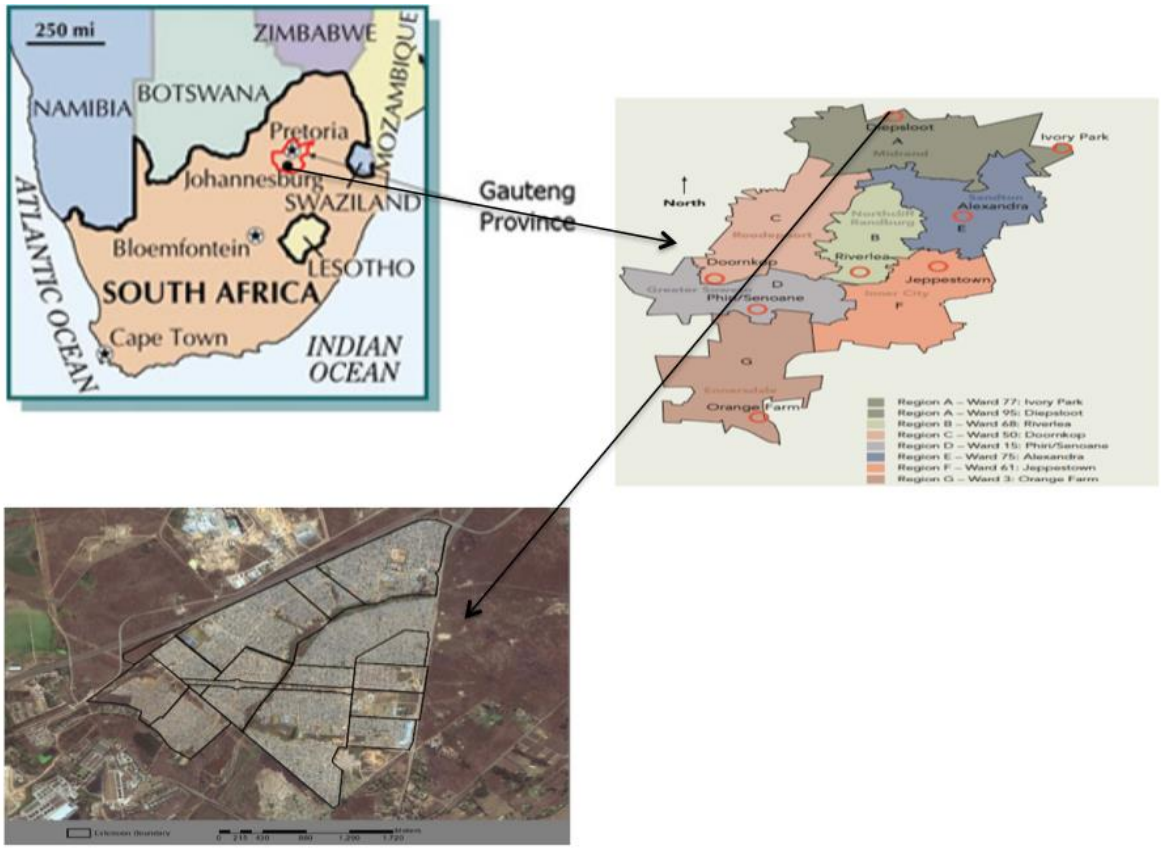
1. Study overview

This is an ancillary study of a community-based cross-sectional survey to estimate the burden of LTBI at the community level and to investigate the spatial epidemiology of LTBI. We estimated the LTBI prevalence and ARI, overall and stratified by age and HIV status. We used a predictive multilevel logistic and normal logistic regression risk models to assess individual-, household- and neighborhood- level factors associated with LTBI prevalence. We used SaTScan statistics to detect significant spatial clusters or hotspots of LTBI. To assess neighborhood-level factors associated with LTBI prevalence, we used ordinary least squares (OLS) linear regression and Geographically Weighted Regression (GWR) models.

2. Study setting

The study area is the Diepsloot township located in Region A, one of seven regions in the city of Johannesburg, South Africa (Figure 5). The township covers a small area, the size of 12 km^2 and is densely population, with a total population of about 150,000 or 12,500 people per km^2. Diepsloot township is comprised of 13 demarcated extensions or ‘‘communities’’ within the township. The study area is typical of many urban South African townships consisting of informal settlements in the form of a mix of series of high-density shacks and government-subsidized brick houses. The Johannesburg Poverty and Livelihoods Study conducted in 2006 among the 8 poorest urban informal settlements including Diepsloot in the City of Johannesburg. Out of 8 settlements, study found that Diepsloot had the highest TB burden per household.[109]

Figure 2.1 Map of study area of Diepsloot township, Johannesburg in Gauteng province, South Africa



3. Parent Study

The study is an ancillary study of a large community-based cross-sectional household survey that was performed between May 2013 and March 2014 using a random sampling framework. The aim of the parent study was to assess unmet health needs and public health priorities in the township order to guide efforts by the department of health (DOH) efforts to improve the health or citizens of Region A.

Geographic coordinates were generated from an aerial map of the 13 digital geo-referenced extensions of the township. Geographic coordinates were randomly selected within each extension and the number of coordinates per extension was proportional to the population size of the extension.

The randomly selected coordinates were then located by the study team using a hand-held geographic positioning system (GPS) device (eTrex 10, Garmin). The household nearest to

but within 30m of each randomly selected geographic coordinate was eligible for study participation. If multiple households were equidistant from the geo-coordinate, the survey team randomly selected one household using a random number generator. Following this method, survey teams approached 2006 households. Households where no-one could be found home despite up to five repeat visits were considered missing and not replaced.

At time of the home visit, the exact latitude and longitude coordinates of the house was geocoded. When the household member agreed for the household to participate in the survey, all household members were enumerated. one of the enumerated adult (≥15 years) household members was randomly selected for study participation using the Kish grid method.[161] This procedure was implemented to avoid the selection bias that would have occurred had the adult household member at home at the time of the survey been systematically selected for study participation. If the adult household member selected for study participation was not home, then the survey team made up to 4 attempts before the household member was considered unreachable. Selected adults who could not be reached were not replaced. All childhood household members were invited to participate in a health assessment if the selected adult household member consented for their study participation. If a child <15 was not in household at the time the selected adult participant was interviewed, no return home visits were made for the child.

4. Study population

The study population of the **parent study** consisted of 1231 adults and 167 children residing in the Diepsloot township. Given the geographically-weighted random sampling framework, the study population is likely a random sample of residents of the Diepsloot township.

The study population for our study is a subset of the parent study population. Data from parent study participants was included in the ancillary study if a TST was performed and the result was read between 48 and 72 hours. Data from parent study participants was excluded

from the analysis if no TST result was available. A total of 446 participants were included in the analysis.

5. Predictor Variables assessment

5.1 Individual-level factors:

- **Age:** self-reported and recorded in years.
- **Gender:** self-reported and defined as male or female
- **HIV Status:** HIV-infected, HIV-uninfected or unknown.

In the parent study, adults of unknown HIV status were offered HIV testing. Children age 12-17 were offered HIV testing with parental consent and child assent. Children <12 years of age were offered HIV testing *only* if the mother was the randomly selected adult household participant and adult consent was provided.

- **Household contact with TB:** defined as contact with an adult or child in the household who was diagnosed with TB
- **Body Mass Index (BMI):** was calculated using the formula: weight (kg) / [height (m)]^2 and classified as underweight (<18.5 kg/m^2), normal (18.5 - 24.9 kg/m^2), overweight (25 - 29.9 kg/m^2), and obese (≥30 kg/m^2).
- **Anemia:** was be measured using hemoglobin level and categorized into non-anemia, and anemia. Anemia was defined as a hemoglobin (Hb) value below 13.0 g/dl (men), 12.0 g/dl (women and 12-15yrs), or 11.0 g/dl(< 5yrs), 11.5 g/dl (5-12yrs); down-adjusted by 0.65 g/dl because of altitude.[162]
- **Alcohol use:** self-reported and was defined as non-alcohol user (never drinks), or alcohol user (drink at least once a week).
- **Smoking:** self-reported and was categorized into non-smoker (never smoke), current smoker (currently smoking), and past smoker (smoking in the past).

- **Educational attainment:** self-reported as no formal education, completed primary education, secondary or post-matric education.
- **Employment Status:** self-reported as unemployed, or employed.
- **Marital Status:** self-reported marital as living with partner, not living with a partner.

5.2 Household-level factors:

- **Household socioeconomic status (SES):** was measured as a composite index of SES based on ownership of durable goods, source of drinking water, toilet facilities and hunger, rather than the current inflow as provided by household income.[163] A principal component analysis (PCA) was applied to these variables, which showed relevant contributions to the combined SES score factor.[164] The factor of the PCA with the highest eigenvalue was used as the variable to describe SES of a household. Households were ranked by SES score and then divided into tertiles of wealth; Low SES, Medium, High SES.
- **Household size**: defined as the self-reported total number of individuals reported to be staying in the participant's household at the time of study.
- **Number of rooms in the household:** self-reported number of rooms.
- **Crowding**: defined as the number adults per bedroom in a household
- **Sleep with windows open**: reported sleeping with windows open served as a proxy measure of household ventilation.
- **Type of dwelling:** Stand alone house Reconstruction and Development Programme (RDP), Stand alone house non-RDP, Informal dwelling/shack in back yard, and Informal dwelling/shack not in back yard.
- **Material of main walls of the household**: observed type of dwelling was categorized into: Plastic/Cardboard, Mud, Bricks/Cement blocks, Corrugated iron.
- **Mobility**: Length of time lived in the household: ≤12 months or > 12 months

- **Food security:** frequency that there was no for the household; never, rarely, sometimes or often. This served as the proxy for malnutrition.

5.3 Neighborhood-level factors:

- **Neighborhood socioeconomic status (SES):** was measured as a mean composite index of household level SES from each neighborhood. Neighborhoods were ranked by this index and then divided into tertiles of wealth; Low SES, Medium, High SES.
- **Percentage under 5 years population:** proportion is the proportion of under 5 years at each community neighborhood.
- **Percentage female headed household:** proportion is the proportion of female headed household at each community neighborhood.
- **Population density:** population per km^2 at each community neighborhood, and
- **Household size density:** household per km^2 at each community neighborhood

6. Outcome variable assessment

6.1 Latent TB Infection

The outcome variable of interest is LTBI. In the parent study, LTBI was measured by TST. The TST was placed by a trained nurse, and read between 48 and 72 hours after placement by a trained community health worker. The CHW marked the edges of the induration with a pen and used a ruler to measure the widest transverse diameter in millimeters (mm). In HIV infected individuals, an induration of $\geq$5 mm was considered positive whereas an induration of $\geq$10 mm was considered positive for HIV uninfected individuals and those with unknown HIV status.

7. Spatial data collection

The survey team collected spatial data of the households in the study area using mobile phone devices with global positioning system (GPS). Digital geographic data from the City of

Johannesburg's GIS department was used to define community boundaries and generate random sampling points using geospatial software (ArcGIS 10.0).

8. Quality Assurance and data management

Standardized questionnaires were used to collect data on all adults on demographics, household characteristics, mobility, disease history, mental health, violence and injury, food security, and substance use, health assessments and built environment characteristics. The questionnaires were routinely checked for missing data and errors.

The data was entered in Microsoft (MS) Access databases. To ensure data quality, the user interface of the databases was very similar to the questionnaires, which eased data entry and reduced the risk of data entry errors. MS Access data quality controls such as validation rules to restrict entries in a given field to a range or type –text vs. numeric were used to further limit data entry errors. The databases were backed up daily in a server to prevent data loss. Further, we made use of queries in MS access for the process of data cleaning. Queries were regularly to identify missing data, inconsistencies and outliers. The results of the queries were returned to the data entry for checking and corrections in the data set.

9. Analytic approach for aim 1

9.1 Prevalence of LTBI

Prevalence of LTBI was calculated using the following formula:

$$\frac{Number\ of\ TST\ positives}{Total\ number\ of\ participants\ TST\ tested\ and\ read}$$

The overall prevalence LTBI was calculated, and by age groups and HIV status. All 95% confidence intervals were also calculated.

9.2 Annual risk of infection (ARI)

Annual risk of infection (ARI) with *M. tuberculosis* in children age 0 to 14 years was calculated using the formula: $ARI = 1 - (1 - P)^{1/a}$;

where P is the observed prevalence of LTBI, and *a* the mean age of participating children.

9.3 Factors associated with LTBI prevalence

We opted for a multilevel (hierarchical) structure of our data with Individuals and households (first level) nested into 20 township neighborhoods (second level). We calculated the intraclass correlation coefficient (ICC) to assess the magnitude of variability due to the covariates at the neighborhood level in order to determine whether multilevel logistic models were appropriate.[165-167] The ICC was calculated by fitting a "null model" using the Stata command *gllamm*.[168]

We used bivariate and multivariable logistic regression to identify individual level factors associated with LTBI. Starting from a full model with all potential predictors, we employed a stepwise backward elimination approach removing the least significant factor one at a time to reach a parsimonious final model that only included the factors significantly associated with LTBI. We repeated the model building procedures using stepwise forward selection to check whether this yielded the same final model. Associations between predictors and LTBI are summarized in odds ratio (OR) along with 95% CIs and the associated p-value. Data analysis was conducted using Stata version 13.1 (Stata Corp, College Station, TX).

10. Statistical analysis for Aim 2:

To investigate the spatial patterns of LTBI and their spatial risk factors in the urban informal settlement of Diepsloot, South Africa.

10.1 Detection of spatial patterns

First, we identified spatial hotspots of LTBI prevalence using the spatial scan statistics (SaTScan).[169] SaTScan is widely used method for spatial hotspots detection in the epidemiology because of its efficiency and accuracy.[170] SaTScan finds the locations with higher or lower number of LTBI cases than expected under spatial randomness by creating circular windows of various sizes to range (scan) across the study area. For each location, the number of LTBI cases inside the window is compared with area outside of it. The radius of

the circular window is changed continuously so that it can take any value from 0 up to a pre-specified maximum value. A maximum circular window of 0.5 km was used for scanning potential hotspots with high or low numbers of LTBI in this study area.

The SatScan software uses a likelihood ratio test to evaluate the statistical significance of each potential hotspot by calculating a likelihood ratio assuming that the number of LTBI in each circular window is an independent Bernoulli random variable, with the likelihood L_0 under null hypothesis of spatial randomness. For each circular window, the number of observed and expected cases of LTBI within and outside the circular window is calculated. The circular window with the highest likelihood ratio values is identified as a LTBI hotspot. P-value associated with likelihood ratio test was determined using Monte Carlo simulations and used to evaluate whether the LTBI hotspot is significantly different form the neighboring space.

After identifying statistically significant spatial clusters, we determined if these areas change when the model is adjusted for known risk factors for LTBI including age, HIV status, sex, conducting separate analyses for each covariate. When adjusting for categorical covariates, the SaTScan program will search for clusters above and beyond that which is expected due to these covariates. We adjusted for covariates when all three of the following are true:

- The covariate is related to the LTBI.
- The covariate is not randomly distributed geographically.
- It is of interest to identify clusters that cannot be explained by that covariate. For example, it is of interest to find clusters of LTBI that are not due to geographic differences in socio-economic status

Covariates were introduced into the spatial scan in an iterative manner and the model was controlled for not more than two covariates at a time to avoid that partition of the data results in unreliable p-values generated by the scan statistic when locations have categories with no data.

For the Poisson model, the expected number of LTBI cases in each area under the null-hypothesis was calculated using indirect standardization. Without covariate adjustment the expected number of cases in a location is

$$E[c] = p*C/P$$

where c is the observed number of cases and p the population in the location of interest, while C and P are the total number of cases and population respectively.

Let c_i, p_i, C_i and P_i be defined in the same way, but for covariate category i. The indirectly standardized covariate adjusted expected number of cases (spatial analysis) is:

$$E[c] = \Sigma i \Sigma i_{E[c} i_{]} = p i * C i / P i$$

By adding one covariate at a time, the analysis allows for assessment of how the underlying geographic distribution of that covariate affects the distribution of LTBI prevalence.

10.2 Assessment of spatial factors associated with LTBI prevalence clusters

To investigate factors associated with LTBI hotspots, we first fitted OLS regression model.

$$\underline{\text{OLS model:}}\ Y_i = \text{LTBI}_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots\ldots + \varepsilon_i$$

where Yi is the dependent variable (LTBI) measured at some location i, Xi is the independent variable, εi is a random error term assumed to be normally distributed, $\beta 0$ is the intercept and $\beta 1$ is the parameter to be estimated.

We tested for the normality, multicollinearity and spatial independence assumptions of the OLS model. Multicollinearity was assessed through variance inflation factor (VIF) values. VIF greater than 10 indicates multicollinearity.[171] The OLS is a global model and expects variable relationships to be constant (stationary) across the study area. The underlying assumption of a multivariate regression model is thus that the relation under study is spatially constant.[172] To adjust for multiple comparisons, we applied the false discovery rate method which was implemented using the *smileplot* add-on module to Stata 13.[173] The spatial independence of residuals was evaluated using the spatial autocorrelation coefficient, Moran's I index.[174] using ArcGIS software (version 0.9.3). A positive spatial correlation (i.e. absence of

spatial independence) will occur when high values of a variable in location i tends to be clustered with high values of the same variable in locations that neighbors of I and vice versa.

The association between risk factors and disease of interest are unlikely to be stationary and more likely to vary over space. The parameter estimates might thus demonstrate significant spatial variation.[172] In that case, we investigated the association between LTBI hotspots and covariates using a geographical weighted regression (GWR) model.[172] A GWR model accounts for and examines the presence of spatial nonstationarity in the association between variables. The GWR is a local model that takes into account how the relationship between the dependent variable (LTBI) and each explanatory variable fluctuates geographically. As such, the GWR model produces more informative results regarding parameters over spatial area under investigation.[172]

GWR model: $Y_{ij} = LTBI_{ij} = \alpha_0 (u_i, v_i) + \Sigma_k \beta_k (u_i, v_i) x_{ik} + \varepsilon_{ij}$

where (ui, vi) denotes the geographic (GPS) coordinates of the ith point and βk (ui, vi) is a realization of the continuous function βk (u, v) at point i. to allow a continuous surface of parameter values and measurements of this surface taken at certain points to denote the spatial variability of the surface.

We used the Akaike Information Criterion (AIC) to determine which covariates improve model fit and to determine whether the spatial perspective significantly improves the model fit. A reduction of >3 in the AIC between the global OLS model and the local GWR signified better model fit.[172]

CHAPTER 3: AIM 1 RESULTS

1. Introduction

Tuberculosis (TB) remains a significant public health problem worldwide with an estimated 9 million cases and 1.3 million deaths in 2013.[1, 175] In 2013, South Africa had the second highest burden of TB in the African region and was ranked fifth among the 22 countries classified by the World Health Organization as high TB burden countries.[1] In 2013, more South Africans died of TB, predominantly HIV-associated TB, than any other disease.[176] These statistics suggest that the current TB control strategy is unable to control the TB epidemic in South Africa, which is fueled by both progression from LTBI to active disease, in large part due to HIV co-infection, and ongoing transmission of *Mycobacterium tuberculosis (M. tuberculosis)*.[177]

Globally, about 2.6 billion people are infected with *M. tuberculosis*, representing a large reservoir of people at risk of progression to active TB disease.[1, 17, 44, 45, 178] About 5-10% of people with LTBI progress to active TB disease in their lifetime, the majority within 2 years of infection.[179] Those at highest risk of progression to active TB disease are young children and immunocompromised individuals.[22, 178, 180] To date, studies of the burden of LTBI in South Africa have mainly focused on high-risk populations such as young children, adolescents, household contacts of TB cases, people living with HIV, gold miners and health care workers.[58, 59, 95, 181-183] These studies observed LTBI prevalence ranging from 26% up to 89%. The only community-based study, performed in an urban township of Cape Town, observed a very high (88.0%) LTBI prevalence, but the study was limited to healthy HIV-negative individuals.[65]

The goal of this study was to describe the burden of LTBI in a representative sample of

all residents of an urban South Africa township and determine factors associated with LTBI.

2. Materials and methods

2.1 Study site and Study population

The study was conducted in Diepsloot, a densely populated, urban township located in northern Johannesburg, South Africa. The community covers an area of 12 km^2 and has an estimated population of 136,289, corresponding to a very high population density of 11,357 people/km^2.[184] The area is typical of urban South African townships, consisting of informal settlements with a mix of high-density shacks and government-subsidized brick houses. According to the 2006 Johannesburg Poverty and Livelihoods Study, Diepsloot is one of the poorest urban informal settlements in Johannesburg.[185]

This analysis represents a sub-study of a large community-based household health survey conducted between May 2013 to March 2014 using a random sampling framework. Geographic coordinates were generated from an aerial map of the 13 digital geo-referenced extensions of the township. Geographic coordinates were randomly selected within each extension and the number of coordinates per extension was proportional to the size of the extension. The randomly selected coordinates were then located by the study team using a hand-held geographic positioning system (GPS) device (eTrex 10, Garmin). The household nearest to but within 30m of each randomly selected geographic coordinate was eligible for study participation. If multiple households were equidistant from the geo-coordinate, the survey team randomly selected one household using a random number generator. Following this method, survey teams approached 2006 households. Households where no-one could be found home despite up to five repeat visits were considered as attempts exhausted and not replaced.

At time of the home visit, the exact latitude and longitude coordinates of the house were geocoded. When a household member agreed for the household to participate in the survey, all household members were enumerated. One of the enumerated adult (≥15 years) household members was randomly selected for study participation using the Kish grid method.[161] This

procedure was implemented to avoid the selection bias that would have occurred had the adult household member at home at the time of the survey been systematically selected for study participation. If the adult household member selected for study participation was not home, then the survey team made up to four additional attempts before the household member was considered unreachable. Selected adults who could not be reached were not replaced. All childhood household members were invited to participate in a health assessment if the selected adult household member was their parent or legal guardian and consented for their study participation. If a child <15 was not at the household at the time the selected adult participant was interviewed, no return home visits were made for the child.

Using a structured questionnaire offered in English, Sesotho or IsiZulu, data on socio-demographics and household characteristics, education and employment, history of TB or contact with TB, and alcohol and smoking habits were collected from all adult participants. A health assessment was performed in all adult and child participants. Weight and height were measured, and blood was collected for hemoglobin and HIV testing by a trained lay HIV counsellor. Participants were assessed for symptoms of active TB and a tuberculin skin test (TST) was placed by a trained nurse. A quantity of 0.1 ml (5TU) of purified protein derivative (PPD) (Aplisol and Tubersol) was injected in the fore arm; the size of induration was read 48 to 72 hours later. Because of adverse events observed in HIV negative individuals, including blistering and ulceration, in consultation with the ethics committee overseeing the study, we decided to restrict placement of TST to HIV positive individuals and children <5 years of age starting from October 2013.

2.2 Study variables

The outcome of LTBI was based on TST positivity, with a TST considered positive if the induration was ≥5 mm in people living with HIV or ≥10 mm in those with unknown or HIV negative status.[186]

Individual covariates included age (<15, 15-24, 25–34, 35–44 or ≥45 years); sex (male or female), HIV status (positive or negative/unknown), body Mass Index (BMI; underweight/normal if BMI ≤18.5 - 24.9 kg/m^2, overweight if BMI 25 - 29.9 kg/m^2, or obese if BMI ≥30 kg/m^2) presence of anemia (with anemia defined as hemoglobin value below 13.0 g/dl for men, <12.0 g/dl for women and children aged 12 to15 years, <11.0 g/dl or children under 5 years, or <11.5 g/dl for children aged 5 to12 years; all down-adjusted by 0.65 g/dl because of altitude),[162] education (primary or less vs. secondary or higher); marital status (living with partner or not living with a partner); employment status (unemployed or employed); household contact with TB (yes or no); smoking status (ever or never), and alcohol consumption (yes or no).

The household-level covariates included were household socioeconomic status (SES), household ventilation and household exposure to smoking. Household SES was calculated as a composite index developed by factor analysis based on household ownership of durable goods (car, motorcycle, bicycle, refrigerator, television, radio, and mobile phone), house ownership, source of drinking water, and type of toilet facilities.[163, 187] Household SES indices were categorized into tertiles of highest, median and lowest household SES. Household ventilation was defined based on the frequency household members sleep with the window open (always, only when warm enough, never, no windows in the house), household exposure to secondary smoking as (yes or no).

We created 20 neighborhoods from the 13 extensions by further subdividing 5 largest extensions of Diepsloot township. Neighborhood-level factors included neighborhood SES which was obtained by summarizing household SES by 20 neighborhood, population density defined as the number of people per square kilometer (low, medium or high) and household density defined as the number of households per square kilometer (low, medium or high). Population and household density data was retrieved from the 2011 South African census as disseminated by Statistics South Africa (STATSSA) using the SuperCROSS software.[184]

3. Statistical analysis

LTBI prevalence was calculated by dividing the number of participants with a positive TST by the total number of participants with a TST result and 95% confidence intervals (95% CI) were estimated. Annual risk of infection (ARI) with *M. tuberculosis* in children age 0 to 14 years was calculated using the formula $ARI = 1 - (1 - P)^{1/a}$; where P is the observed prevalence of LTBI, and *a* the mean age of participating children.[188, 189]

We opted for a multilevel (hierarchical) structure of our data with Individuals and households (first level) nested into 20 township neighborhoods (second level). We calculated the intraclass correlation coefficient (ICC) to assess the magnitude of variability due to the covariates at the neighborhood level in order to determine whether multilevel logistic models were appropriate.[165-167] The ICC was calculated by fitting a "null model" using the Stata command *gllamm*.[168]

We used bivariate and multivariable logistic regression to identify individual level factors associated with LTBI. Starting from a full model with all potential predictors, we employed a stepwise backward elimination approach removing the least significant factor one at a time to reach a parsimonious final model that only included the factors significantly associated with LTBI. We repeated the model building procedures using stepwise forward selection to check whether this yielded the same final model. Associations between predictors and LTBI are summarized in odds ratio (OR) along with 95% CIs and the associated p-value. Data analysis was conducted using Stata version 13.1 (Stata Corp, College Station, TX).

4. Ethics statement

This study was approved by the institutional review board of University of North Carolina at Chapel Hill and by the University of the Witwatersrand's Human Research Ethics Committee. Written consent was obtained from all adult participants; written parental consent was obtained

for all children <18 years old and child assent was obtained for participating children ages 7-17 years old.

5. Results

5.1 Study participants

Of the 2006 randomly selected households, 1620 could be enumerated. Of the 1620 randomly selected adults, 1581 (97.6%) could be contacted and 1230 agreed to participate. (Fig 3.1) In addition, 169 children living in the same household as the participating parent were enrolled. TST was offered to 626 participants (all participants until October 2013, only HIV positive individuals and children <5 thereafter). Of these, 144 refused and TST was not placed in a further 23 due to logistical considerations (e.g. the participant would not be available to have the TST read). Of the 459 participants in whom a TST was placed, the TST was read in 446 (97%). The remaining 13 could not be traced within 48-72 hour of TST placement.

Among the 446 participants with TST result, mean age was 35 years, 11% were 0 to 15 years of age, 17% were 15 to 24 years, 33% were 25 to 34 years, 18% were 35 to 34 years and 21% were 45 years or older (Table 3.1). Sixty percent were female, 44% were married or living with a partner, two thirds (67%) were unemployed and the majority (72%) had at least some secondary education. Self-report of smoking (26%) and alcohol use (37%) was relatively low. Overall, 18% of the 446 participants with TST result were HIV positive, 35% were anemic, 23% were underweight and 27% obese. Only 6% of participants reported a history of contact with a TB case. Almost all (93.4%) participants either lived in a house without windows or never slept with windows open and 20% were exposed to household secondary smoking.

5.2 Distribution of TST results, LTBI Prevalence and Annual Risk of infection

The size of induration observed in 446 participants was 1-4mm in 7.2%, 5-9mm in 11.6%, and ≥ 10-mm in 33.4% (Fig 2.2). Off 77 HIV positive participants, 4 had 1-4mm induration size, 6 had 5-9mm induration size and 19 had ≥ 10mm induration size. Using HIV-specific definitions for LTBI, the overall prevalence of LTBI was 34% [95% CI, 30% – 39%].

LTBI prevalence increased with age, from 19% in the 0–14 age group to 45% in the 45 and older age group (p=0.002), was higher in women (37%) than men (32%) (p=0.273), but similar in HIV positive (36%) and HIV negative (32%) participants (p=0.553) (Table 3.2). Based on changes in TST prevalence with age among children age 0 to 15 years, the ARI was estimated at 3.1% (95% CI: 2.1 - 5.2).

5.3 Factors associated with LTBI at individual and household level

In univariable logistic regression, age showed a strong association with LTBI with increasing odds of LTBI for every year increase in age (OR =1.17, 95% CI = 1.08 – 1.26) (Table 3.3). Other variables associated with LTBI were marital status, with individuals living with a partner being twice as likely to have LTBI compared with those living with a partner (OR =2.00, 95% CI: 1.06 - 3.80); history of household contact with TB , with those reporting such history being twice as likely to have LTBI compared with those not in household contact with a TB case (OR = 2.33, 95% CI: 1.03 - 5.28); and number of room in the house, with people living in dwellings with 3 or more rooms being more likely to have LTBI compared to people living in dwellings with less than 3 rooms (OR = 1.62, 95% CI: 1.05 - 2.50). People of the highest tertile of SES were 1.5 times more likely to have LTBI as compared to the lowest SES tertile, but the 95% CI crossed 1 (95% CI 0.91 - 2.47). In multivariable logistic regression, age (OR =1.03, 95% CI = 1.01 – 1.05), male gender (OR =1.77, 95% CI = 1.10 – 2.86), being married/cohabitating (OR =2.00, 95% CI = 1.13 – 3.54) and living in a household that belongs to the highest tertile SES of the community (OR 2.11, 95% CI 1.04 – 4.31) were independently associated with a diagnosis of LTBI.

5.4 Factors associated with LTBI at neighborhood level

None of the neighborhood level factors were associated with LTBI. The multilevel ‘‘null’’ model showed that ICC was 0.01032 (p=0.4005), meaning that only 1% of the variance in LTBI was explained by differences in neighborhood factors. Fitting a multilevel logistic regression model was thus not indicated for the analysis of our data. Small variability at the neighborhood

level may have been due to the sparsity of level 2 clusters with only 20 neighborhoods (level 2 clusters), smaller than recommendation of 50 level 2 clusters.[190]

6. Discussion

The burden of LTBI in this urban informal settlement community of northern Johannesburg, South Africa, was high with an overall prevalence of 34.3% and an annual risk of infection of 3.1%. Risk factors independently associated with LTBI prevalence were older age, male gender, living with a partner, and high SES.

While the LTBI burden observed was high, the 34.3% prevalence was lower than has been observed in the few prior population-based studies previously performed in urban townships. In a Peruvian shantytown and a Ugandan urban population, the LTBI prevalence was higher, with half of all residents living with LTBI (52%; 95% CI: 48-57 in Peru and 49%; 95% CI: 44-55 in Uganda).[191, 192] A study of 8 South African urban communities however showed that LTBI prevalence among household contacts can be highly variable between communities in the same region, as they documented a range of LTBI prevalence from 24% to 77%.[193] The ARI in our study fell within the range of ARI estimates from prior South African studies (2.8% - 5.8%).[58, 194] Taking together, these results suggest that the LTBI prevalence in urban settlements is high, but shows substantial variation.

Individual risk factors for LTBI were household contact with a TB case, (OR 2.96 95% CI 1.21 – 7.24) increasing age (OR for each 1 year increase 1.03; 95% CI: 1.01-1.05), male gender (OR 1.77; 95% CI: 1.10-2.86), and living with a partner (OR =2.00, 95% CI = 1.13 – 3.54). Exposure to a household TB case is well established risk factor for LTBI,[32] resulting in a large proportion of LTBI among children and young adults being due to household exposure to TB.[195] The increasing prevalence of LTBI with age reflects the cumulative exposure to TB through social interaction in high TB burden settings[159, 196-199] and is consistent with findings of other LTBI studies in urban populations.[58, 65, 183, 192, 193] Data on the association between male gender and increased LTBI prevalence are conflicting. A higher LTBI prevalence among males was

also observed in a rural area of Ethiopia[49] and a Peruvian peri-urban shantytown[191] but not in an urban population in Ugandan[192] nor in prior South Africa studies.[183, 193, 200] The higher rate of LTBI in urban males we observed may be due to the high risk of TB transmission in social gathering places, such as informal alcohol drinking establishments (shebeens),[196, 201] which are more frequented by men than women.

HIV infection was common (18%) but not associated with LTBI prevalence in this population. Other LTBI prevalence studies in high HIV burden settings have reported similar observations.[102, 192] The lack of association between HIV and LTBI may be due poor sensitivity of TST in HIV-infected individuals,[66] however we addressed this by decreasing the TST cut-off to 5mm.[68] In addition, some other risk factors such as poor ventilation, smoking and exposure to household secondary smoking[202, 203] were not associated with LTBI prevalence in our study. TB disease has clearly been established a as disease of poverty.[28, 204] It is therefore surprising that we observed a higher LTBI prevalence among people with higher household SES. An association between higher SES rather than lower SES and greater LTBI prevalence was also observed in a study in Zambia [102], and in a population-based multicenter study in China.[205]

Taken together, these findings suggest that SES may have a differential effect on the risk of LTBI acquisition and risk of progression from infection to active TB disease. Boccia et al suggested that “it is possible that, especially in urban settings, higher SEP is associated with housing characteristics that reduce ventilation and life-styles that increase social mixing and therefore the likelihood of contact between cases and susceptible people. We could not find an association between ventilation and LTBI, and higher SES was not associated with poorer ventilation in our sample. Given that we did not assess use of public transportation or social mixing in our study we could not assess whether these factors can explain the observation of higher LTBI prevalence in people of higher SES within urban settlements. These hypotheses thus warrant further in-depth investigations.[196, 206, 207]

Our study had many strengths, including the population-based design with geographically weighted random sampling of the general population, including people living with and without HIV and both adults and children, and a standardized approach to define SES tertiles. Our study does have some limitations. First, the cross-sectional nature of the study does not allow for establishment of temporality or causality between LTBI and associated factors. Second, even though some of well-known risk factors such occupation, crowding, and ventilation were not measured, the proxy measures of these factors were not associated with LTBI. BCG vaccination status, which can reduce the specificity of TST, was also not documented.[208, 209] Third, the sample size was relatively small, especially for children under 12 years of age since we did not made more attempts to find this group of participants if they were not at home during interview of the adult participants. Finally, as 19% of the targeted households were not enrolled due to failure to find someone at home despite multiple attempts or refusal to participate, our aim to enroll a representative sample of the population may not have been fully achieved.

In conclusion, the prevalence of LTBI and the annual risk of infection with *M. tuberculosis* are high in urban populations, especially in men, but independent of HIV infection status. The unexpected association between higher LTBI and higher household SES suggest that the differential association between SES as risk factors for acquisition of TB infection and progression from LTBI to active disease is not yet fully understood. A better understanding of individual, household and community-level risk factors for LTBI will be important for the development of efficient, targeted LTBI interventions in high TB burden settings.

Figure 3.1 Flow chart of study participants

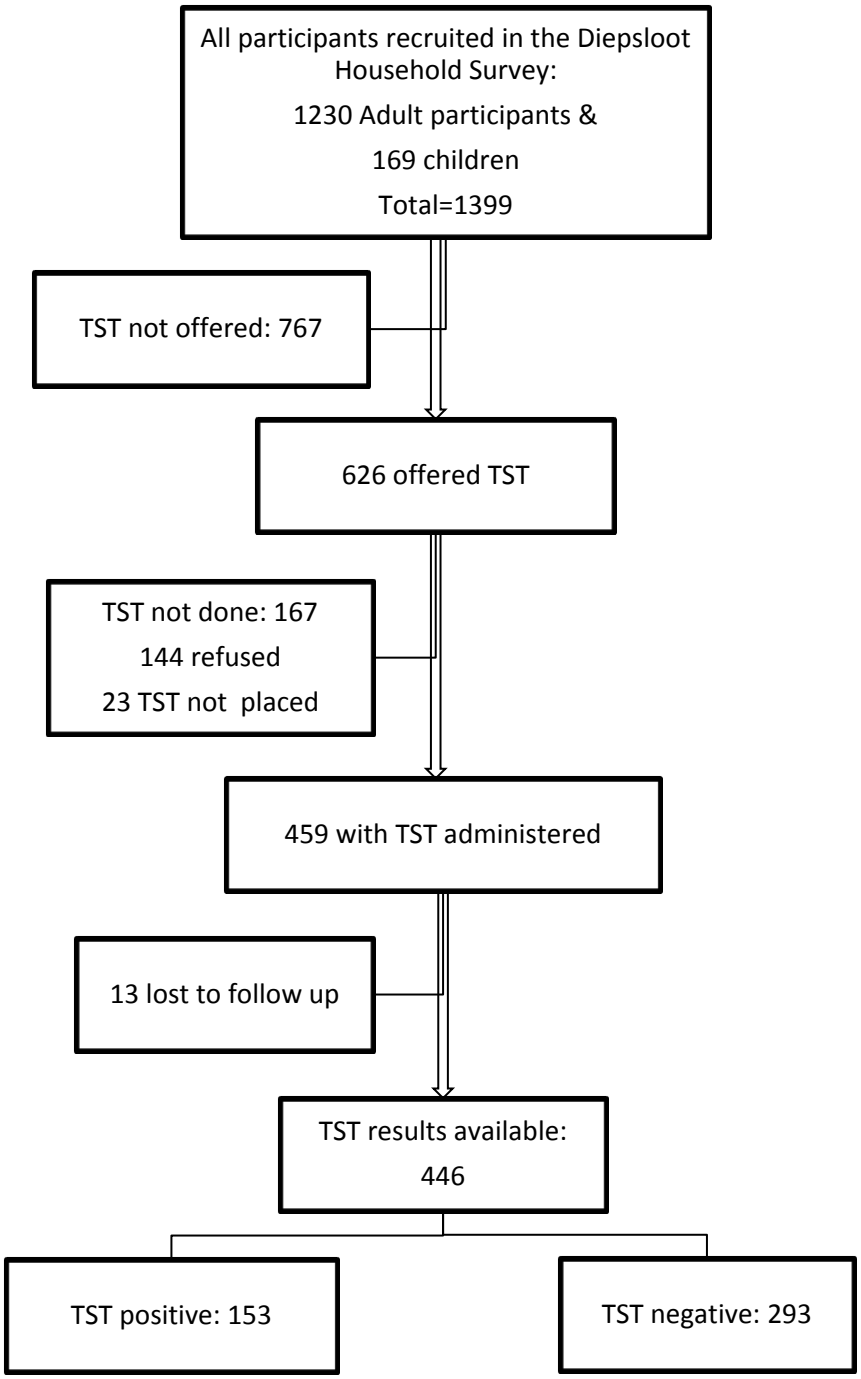


Figure 3.2 Frequency distribution of indurations (in mm) in 446 residents of an urban township of Diepsloot.

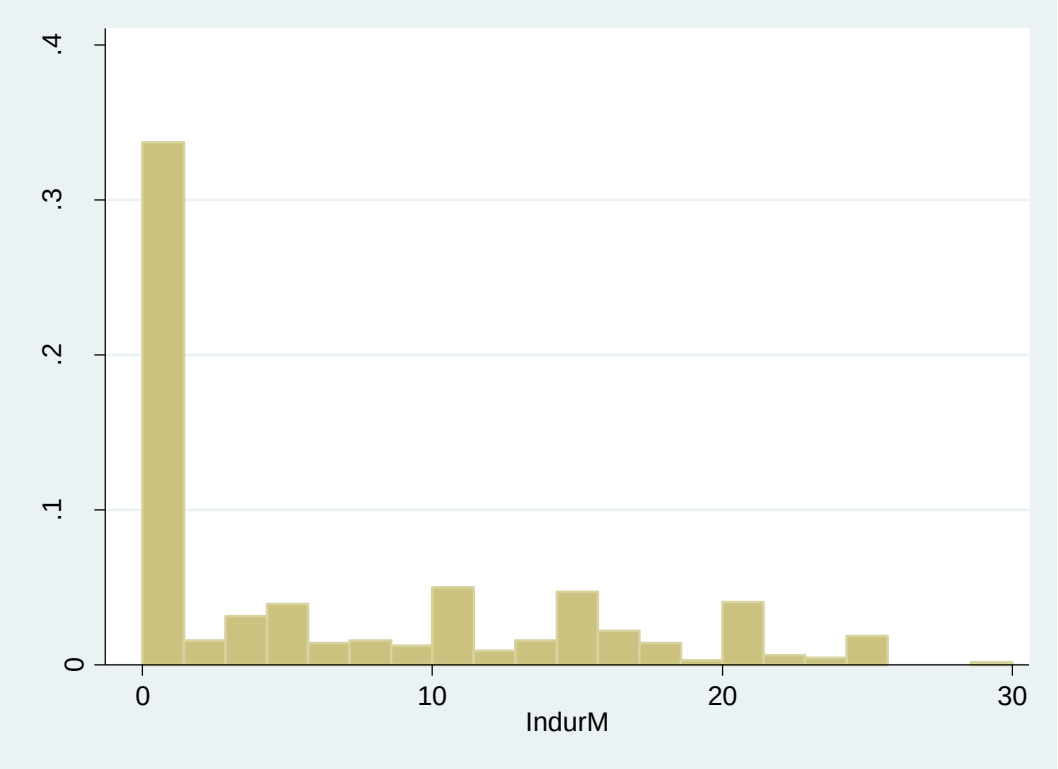


Table 3.1 Characteristics of 446 participants with positive and negative TST results.

Characteristics		TST Positive N (row %)	TST Negative N (row %)	Total N (column %)
Age in Years median (IQR)		**35 (27 – 45)**	**29 (22 – 38)**	**32 (23-41)**
Age group	0-14	9 (18.7)	39 (81.3)	48 (10.8)
	15-24	20 (26.3)	56 (73.7)	76 (17.0)
	25-34	45 (30.6)	102 (69.4)_	147 (33.0)
	35-44	37 (45.1)	45 (54.9)	82 (18.4)
	≥45	42 (45.2)	51 (54.8)	93 (20.8)
Sex	Female	86 (32.2)	181 (67.8)	267 (60.4)
	Male	65 (37.1)	110 (62.9)	175 (39.6)
HIV status	Positive	23 (32.9)	47 (67.1)	70 (18.1)
	Negative	115 (36.3)	202 (63.7)	317 (81.9)
BMI (kg/m^2) median (IQR)		**25 (21 – 29)**	**23 (20 – 29)**	**24 (20 – 29)**
BMI categories	Normal/underweight	70 (34.8)	131 (65.2)	201 (49.9)
	Overweight	39 (41.5)	55 (58.5)	94 (23.3)
	Obese	37 (34.3)	71 (65.7)	108 (26.8)
Anaemia	Anaemic	51 (32.3)	107 (67.7)	158 (35.4)
	Non-anaemic	102 (35.4)	186 (64.6)	288 (64.6)
Education level	≤ Primary	44 (40.0)	66 (60.0)	110 (28.1)
	≥ Secondary	100 (35.4)	181 (64.6)	281 (71.9)
Employment status	Unemployed	96 (36.8)	165 (63.2)	261 (66.6)
	Employed	48 (36.6)	83 (63.4)	131 (33.4)
Marital status	Not living with a partner	34 (29.6)	81 (70.4)	115 (61.6)
	Living with a partner	75 (44.4)	94 (55.6)	169 (38.4)
Household contact with TB	Yes	14 (58.3)	10 (41.7)	24 (6.2)
	No	129 (35.3)	236 ((64.7)	365 (93.8)
Smoking	No	101 (35.0)	188 (65.0)	289 (73.7)
	Yes	43 (41.8)	60 (58.2)	103 (26.3)
Alcohol consumption	No	87 (35.2)	160 (64.8)	247 (62.7)
	Yes	57 (38.8)	90 (61.2)	147 (37.3)
Household ventilation (sleep with window open)	Always/Only when warm	13 (50.0)	13 (50.0)	26 (6.6)
	Never/No windows	131 (35.6)	237 (64.4)	368 (93.4)
Household SES	Low	43 (29.9)	101 (70.1)	144 (34.0)
	Medium	50 (33.8)	98 (66.2)	148 (35.0)
	High	51 (38.9)	80 (61.1)	131 (31.0)
Household exposure to secondary smoking	No	115 (36.4)	200 (63.6)	315 (79.9)
	Yes	29 (36.7)	50 (63.3)	79 (20.1)
Household number of rooms	<3	102 (31.3)	224 (68.7)	326 (73.1)
	≥3	51 (42.5)	69 (57.5)	120 (26.9)
Household density	Low (<300/km^2)	53 (34.0)	103 (66.0)	156 (35.0)
	Medium (300-600/km^2)	56 (38.9)	88 (61.1)	144 (32.3)
	High (>600/km^2)	44 (30.1)	102 (69.9)	146 (32.7)

Table 3.2 Estimated prevalence of infection by age, sex, and HIV status.

Characteristics	Mean age years	Prevalence, % % (95% CI)	p-value
Overall	32.2	34.3 (30.0 - 38.8)	
Age group years			
0-14	6.2	18.8 (10.0 - 32.5)	0.002
15-24	20.7	26.3 (17.6 -37.4)	
25-34	29.5	30.6 (23.7 -38.8)	
35-44	39.1	45.1 (34.6 -56.1)	
45+	53.4	45.2 (35.3 - 55.4)	
Sex			
Male	33.3	37.1 (30.3 - 44.8)	0.273
Female	31.8	32.1 (26.7 -37.9)	
HIV status			
Positive	38.1	32.4 (22.5 - 44.2)	0.553
Negative	32.7	36.1 (31.0 - 41.5)	

Table 3.3 Logistic regression analysis of risk factors associated with LTBI.

Characteristics		Unadjusted odds ratio (95% CI)	Adjusted odds ratio (95% CI)
Individual-level characteristics			
Age in years		1.17 (1.08 – 1.26)	1.03 (1.01 – 1.05)
Sex	Female	1.00	1.00
	Male	1.25 (0.84 - 1.86)	2.70 (1.55 – 4.70)
HIV status	Negative	1.00	
	Positive	0.85 (0.49 - 1.46)	
BMI categories	Normal/Underweight	1.00	
	Overweight	1.33 (0.80 - 2.19)	
	Obese	0.97 (0.60 - 1.59)	
Anaemia	Non-anaemic	1.00	
	Anaemic	0.87 (0.58 - 1.31)	
Education level	≤ Primary	1.00	
	≥ Secondary	0.83 (0.53 - 1.30)	
Employment status	Unemployed	1.00	
	Employed	0.99 (0.64 - 1.54)	
Marital status	Not living with a partner	1.00	1.00
	Living with a partner	1.90 (1.50 - 3.14)	2.00 (1.13 - 3.54)
Household contact with TB	No	1.00	1.00
	Yes	2.33 (1.03 - 5.28)	2.27 (0.76 – 6.82)
Smoking	No	1.00	
	Yes	1.33 (0.84 – 2.11)	
Alcohol consumption	No	1.00	
	Yes	1.16 (0.76 - 1.78)	
Household- and neighbourhood-level characteristics			
Household exposure to secondary smoking	No	1.00	
	Yes	1.01 (0.61 – 1.69)	
Household number of rooms	<3	1.00	
	≥3	1.62 (1.05 - 2.50)	
Household ventilation (sleep with window open)	Always/Only when warm	1.00	
	Never/No windows	1.37 (0.29 – 6.53)	
Household SES	Low	1.00	1.00
	Medium	1.20 (0.73 - 1.96)	1.73 (0.85 – 3.52)
	High	1.50 (0.91 - 2.47)	2.11 (1.04 – 4.31)
Household density	Low (<300/km^2)	1.00	
	Medium (300-600/km^2)	1.24 (0.77 - 1.98)	
	High (>600/km^2)	0.84 (0.52 - 1.36)	

CHAPTER 4: AIM 2 RESULTS

1. Introduction

Latent tuberculosis infection (LTBI) is a significant public health problem worldwide affecting people of all ages.[210] About 2.6 billion people are estimated to be infected with *Mycobacterium tuberculosis,* resulting in an estimated 200 million people progressing to active tuberculosis (TB) disease during their lifetime.[1, 17, 44, 45] People latently infected with TB thus represent a large reservoir for future TB disease.[22, 180]

In many low-burden settings, targeted testing for LTBI among high-risk individuals and provision of treatment for LTBI is a key aspect of the TB control strategy, with TB elimination as the ultimate goal.[180] In high burden settings, the focus of TB control efforts is on achieving high case detection rates, especially of infectious cases of pulmonary TB, and high cure rates. Since 2004, the WHO recommends treatment of LTBI in people living with HIV in high burden settings,[211, 212] but the uptake of this policy has been slow.[2] The most recent Global Plan to stop TB (2016-2020) aims to have 90% of people living with *M. tuberculosis* infection, and 90% of children exposed to TB, on treatment for latent *M. tuberculosis* infection.[213]

Mathematical models have long identified LTBI treatment as an important tool for long-term reductions in TB incidence and TB elimination.[108, 214-218] There is also a growing recognition that accelerated progress in TB control in high burden settings may only be achieved by expanding LTBI treatment beyond those co-infected with HIV.[219-221] A trial of community-wide mass Isoniazid Preventive Therapy (IPT) in a mining population in South Africa however failed to show a significant impact in reducing TB incidence.[222] Using a mathematical model, Dowdy et al. showed that the impact of LTBI treatment in reducing TB incidence can

vary greatly in communities due to spatial clustering of LTBI,[116] suggesting that successful implementation of LTBI interventions may require knowledge of the spatial patterns of LTBI in order to identify those sub-groups that should be prioritized for interventions. In the US, a study used spatial analysis to identify the neighborhoods with the highest LTBI incidence for targeted LTBI screening and treatment. Ten years later, TB incidence declined from 15 to zero cases in the targeted neighborhoods, and from 128 to 75 cases in the entire county.[143] Findings from this study support the results from the mathematical models and demonstrate the potential role of spatial analysis in the planning of targeted LTBI interventions.

To date, most studies using geographic information systems (GIS) and spatial analysis have investigated the distribution of active TB and its risk factors, predominantly in low burden countries,[126, 127, 130-133, 223-225] with few studies in sub-Saharan Africa.[134, 226-228] To our knowledge, no studies have assessed the spatial distribution of LTBI in sub-Saharan African communities. This study explores the spatial distribution of LTBI, identifies potential LTBI hotspots, and investigates associations between community-level factors and LTBI prevalence in an urban township of Johannesburg, South Africa.

2. Methods

2.1 Study area

Diepsloot, a densely populated, geographically well-demarcated urban township located in northern Johannesburg, South Africa covers an area of 12 km^2 in size, is divided in 13 extensions, and has an estimated total population of 136,289 people (Figure 4.1).[184] The study area is typical of many urban South African townships, consisting of informal settlements with a mix of high-density shacks and government-subsidized brick houses. According to the 2006 Johannesburg Poverty and Livelihoods Study, Diepsloot is one of the poorest urban informal settlements in the City of Johannesburg.[185]

2.2 Study population and data collection

This is a sub-study of a large community-based household health survey conducted in Diepsloot using a random spatial sampling framework, from May 2013 to March 2014. Survey teams consisting of nurses and lay HIV counsellors aimed to visit all of the 2006 randomly selected households. At time of the home visit, the latitude and longitude coordinates were geocoded using a hand-held global positioning system receivers and all household members were enumerated. At each participating household, one household member (≥15 years) was randomly selected for study participation using the Kish grid method to minimize bias from the survey team and to prevent only sampling family members that are the easiest to find at home.[229] A questionnaire for socio-demographics, medical history and health seeking behavior was administered, and a health assessment included anthropometrics, HIV testing, and active and latent TB investigation. All children (<15 years) living in the same household of an adult participant were invited to participate. A health assessment was performed in participating children.

To determine LTBI, a tuberculin skin test (TST) was administered by a trained nurse. A quantity of 0.1 ml (5TU) of purified protein derivative (PPD) (Aplisol or Tubersol) was injected in the participant's forearm and the size of induration was read by study staff 48 to 72 hours later. All participants with TST results were included in this analysis. LTBI was defined as a TST with induration of ≥5 mm in people living with HIV or ≥10 mm in those with unknown or HIV negative status.

3. Data analysis

Population data was retrieved from the 2011 South African census as disseminated by Statistics South Africa (STATSSA) using the SuperCROSS software.[184] We split the five largest extensions to create 20 neighborhoods of 2,669 to 13,738 inhabitants. For each of the 20 neighborhoods, the LTBI prevalence was calculated as the proportion of people with a positive

TST result among all individuals who had a TST placed and 95% confidence intervals (95% CI) were estimated.

A composite index for socioeconomic status (SES) was developed by factor analysis based on household ownership of durable goods (car, motorcycle, bicycle, refrigerator, television, radio, and mobile phone), home ownership, source of drinking water (piped water in house, yard or public tap), and type of access to toilet facilities (Table 4.1).[39, 40] High indices reflect higher SES (lower poverty) while low indices reflect lower SES (higher poverty). Household SES indices were summarized by neighborhood to obtain community-level SES. Data for the other community-level factors were retrieved from the 2011 South African census including percentage under 5 years population, percentage female headed household, population density, and household size density.[184]

To represent the spatial distribution of LTBI prevalence, we generated choropleth maps of the 20 neighborhoods using the *spmap* module written for Stata.[230] To identify LTBI clusters of higher or lower than expected number of LTBI cases, a discrete Poisson model spatial scan statistic (SaTScan™ software version 9.4.1) was employed using the population size for each neighborhood.[231] The radius of the circular window varied from zero to a maximum of 1, 0.8, 0.5, 0.4, and 0.2 km to restrict the maximum size of the window from exceeding 50% 40%, 25%, 20%, to 10% of the total study population, respectively.[169, 231] Using Monte Carlo hypothesis testing, the primary cluster was identified as the circle with the maximum likelihood among all radius sizes explored at all point locations. Using SaTScan, we also identified the secondary clusters with significantly large likelihood ratios.

To investigate factors associated with LTBI hotspots, we built global ordinary least squares (OLS) models and tested the assumption of multicollinearity among all factors investigated using variance inflation factor (VIF) values, with VIF greater than 10 indicating multicollinearity.[171] Using the spatial autocorrelation coefficient Moran's I index, we tested the spatial independence assumption of the global OLS model, which assumes that the relation

under study (SES and LTBI) is constant over space.[172, 174] To adjust for multiple comparisons, we applied the false discovery rate method which was implemented using the *smileplot* add-on module to Stata 13.[173] Where significant spatial variation was observed, we built local geographical weighted regression (GWR) models to investigate the spatial association between LTBI and covariates. GWR model was as follows:

$$Y_i = \beta_{i0} + \sum_{j=1} \beta_{ij} X_{ij} + \varepsilon_i$$

Here, *i* is the neighborhood of a study area, *Y* represents LTBI at neighborhood *i*, X_{ij} represents the value of jth covariate at neighborhood *i*, β_{i0} represents the intercept and the regression coefficients of jth covariate and error term are represented as β and ϵ, respectively.

We used the Akaike Information Criterion (AIC) to determine whether the spatial model improves the fit. A better model fit was defined as a reduction of > 3 AIC between the global and local model.[172] Stata (IC version 13.1, StataCorp LP) and ArcGIS software (version 10.2.2) were used for analyses and data visualization.

4. Ethics

This study was approved by the institutional review board (IRB) of University of North Carolina at Chapel Hill and by the University of the Witwatersrand's Human Research Ethics Committee. Written consent was obtained from all adult participants (≥18 years old), written parental consent obtained for all children <18 years old, and written assent was obtained from all participating children age 7-17 years old.

5. Results

5.1 Descriptive statistics

The summary statistics of the study neighborhood-level variables are shown in Table 4.2. The population size of the 20 neighborhoods ranged from 2669 to 13738 (mean = 6814.45, standard deviation = 2692.25), with population density ranging from 15023 to 49566 per km^2. The average neighborhood SES index was 2.88 (Min= 1.14, Max=4.52).

The neighborhood LTBI prevalence was based on 153 TST positive results, corresponding to an overall prevalence of 34.3% (153 positive TST results among 446 TST's reading, 95% CI 30.0%, 38.8%). Prevalence of LTBI in Diepsloot at a neighborhood-level ranged from 15.0% (95% CI 4.7%, 38.4%) to 59.1% (95% CI 37.6%, 77.6%) (Figure 4.2).

5.2 Spatial scan statistics: LTBI hotspots

One primary and 3 secondary clusters of LTBI were detected (Table 4.3, Fig 4.3). Based on Monte Carlo hypothesis testing, the SaTScan software identified a primary cluster (Relative Risk= 2.06, log likelihood ratio = 40.59, radius = 0.98 km, p = 0.03) that included 3 neighborhoods in the northern part of Diepsloot. The secondary clusters were smaller in size, and had relative risks between 1.43 to 2.33, but were not statistically significant ($p \geq 0.47$). As a sensitivity analysis we tested the impact of different population size (i.e., 50%, 40%, 30%, 20%, and 10% of total population) on the spatial pattern of LTBI but did not find any significant difference in identification of primary and secondary clusters.

5.3 OLS model

In global (OLS) regression analysis, only SES level was associated with LTBI prevalence, with higher LTBI prevalence in neighborhoods with higher SES (coefficient = 0.039, p = 0.048; Table 4.4). Proportion of female headed households and household size had negative coefficients, suggesting that LTBI prevalence may be higher in neighborhoods with lower proportion of female headed households and low household density, though neither of these associations were statistically significant (p=0.10 and p=0.33, respectively). Population density and proportion of population below 5 years of age had positive coefficients, suggesting that higher LTBI prevalence may occur in areas of higher population density and higher proportion population below 5, but these associations were again not statistically significant (p =0.07 and p=0.34, respectively). The OLS model explained only 7.5% of the variance in neighborhood-level LTBI prevalence. The VIF values ranged from 1.45 to 2.36, suggesting no signs of strong multicollinearity (Table 4.4). The residuals of the OLS model were spatially random.

5.4 Local GWR model

Compared with the global OLS model, the local GWR model improved model fit (AIC 18.21 vs 10.84). The R^2 for the local GWR model ranged from 0.23 to 0.53, suggesting that the R^2 of the global OLS model (0.32) summarizes a wide distribution of local associations between the neighborhood-level covariates and LTBI prevalence. The local GWR model also explained a greater proportion (16% vs 7.5%) of the variance in neighborhood-level LTBI prevalence (Table 4.5).

Figure 4.4 shows the maps for the GWR regression residuals, R-squared values, intercept and coefficients of the predictors across the study area. The regression coefficients show both the direction and size of the spatial relationship between LTBI prevalence and predictors. The neighborhood SES, population under 5 years and population density were positively associated with LTBI prevalence (Fig 4.4D-4E and 4G). By contrast, proportion of female headed households and household density were inversely associated with LTBI prevalence (Fig 4.4F and 4.4H). The SES had the strongest positive association on LTBI prevalence in the south eastern part, and the lowest positive association on LTBI prevalence in the northern and eastern part of Diepsloot. LTBI prevalence was higher in neighborhoods that had high proportions of the population less than 5 years and population density and these neighborhoods were mainly located in south and north western Diepsloot. On the other hand, LTBI prevalence was lower in neighborhoods that had high proportions of female headed households and household density and these neighborhoods were mainly located in south eastern Diepsloot.

6. Discussion

South Africa has one of the highest burdens of TB globally, but little is known about its burden of LTBI, the spatial heterogeneity of LTBI, and factors associated with the spatial patterns of LTBI at community level. In Diepsloot, one of the poorest townships of Johannesburg, South Africa, we found a high burden of LTBI (34%) and observed that the LTBI prevalence was not homogenous in its spatial distribution, with clusters occurring within this

relatively small (12 km^2) community. We observed a positive association between LTBI and SES, with higher LTBI prevalence in areas with higher SES, and the local GWR model showed that this association varied across space.

The observed LTBI prevalence corresponds with the WHO LTBI prevalence estimate for the Africa region (35%),[17] but was much lower compared to the findings of other recent studies of LTBI prevalence in South Africa. In a poor, densely populated black township of Cape Town, 88% of HIV-uninfected young adults and 53% of HIV-infected adults were TST positive.[65, 232] In the gold mines in the North West and Gauteng provinces, 77%–89% of miners were found to be infected be *M. tuberculosis*.[59] While miners are known to be at high risk of LTBI acquisition due to a combination of high rates of active TB, poor housing conditions and occupational risk, the reason for the stark difference in LTBI prevalence between township residents in Cape Town and Johannesburg is unclear.

To date, spatial epidemiology studies of TB in South Africa mainly aimed to visualize the spatial distribution of active TB cases.[124, 128, 233] To date, only two studies have explored the spatial distribution of LTBI in an urban African setting, but these studies limited inclusion to children and adolescents in a Cape Town township, did not investigate spatial clustering of LTBI, and excluded the area (approximately 20%) of the township with informal housing.[158, 159] To our knowledge, this is the study assessing community-level spatial clustering of LTBI prevalence in urban Africa. Our results of the spatial distribution of LTBI mirror the findings of the three spatial cluster analysis studies of active TB in sub-Saharan Africa (Antananarivo, Madagascar; Banjul, the Gambia; and Douala, Cameroon) which showed significant clustering of active TB within a single city.[134, 226, 228]

The association between TB disease and lower SES is well established,[27, 28] and spatial analysis studies have demonstrated that the association between TB disease and SES varies by geographic locations.[125, 128, 228, 234, 235] In contrast, studies of the association between LTBI and SES provided conflicting results, with studies in the Gambia, South Africa and Peru not

finding an association,[11, 75, 236] and a study in Zambia observing a higher risk of LTBI with higher SES.[102] Similar to the Zambia study, we observed that, at the community level, higher SES was associated with higher LTBI prevalence. Taken together, these findings are surprising given that tuberculosis is a disease of poverty. Both the GWR and OLS regression models had low R^2, suggesting that there are community level factors not assessed in our study that influence LTBI prevalence not present in our data.

While our study is one of the first to employ spatial methods to assess the spatial distribution of LTBI in a country with a high burden of active TB, some limitations should be noted. First, while great care was taken to have a random sample of the community, our sample size was affected by the decision of the ethics committee overseeing the study. The ethics committee restricted the placement of TST only to HIV positive individuals and children under 5 years old in October 2014 after high rate of adverse events observed in HIV negative individuals, including blistering and ulceration. However, this restriction did not affect the representation of HIV negative individuals in the study. Second, causality cannot be inferred given the cross-sectional nature of the study. Third, the determinants of LTBI were assessed at the community level, which is appropriate given that LTBI risk is driven by prevalence of active TB in the community, but precludes inference of associations at the individual level. Finally, the low R^2 of the regression model suggests that factors other than those assessed in this study may be important determinants of the spatial distribution of LTBI prevalence within a community.

7. Conclusion

We observed high LTBI prevalence to varying degrees across the neighborhoods. Our study also showed spatial clustering of LTBI. Spatial cluster analysis has the potential to detect LTBI hotspots within a small geographic area and inform LTBI interventions, potentially allowing more efficient tailoring and targeting of the intervention to neighborhoods in greatest need. Only a part of the spatial heterogeneity in LTBI prevalence in high burden communities was

explained by the spatial variation of community-level SES, suggesting that further research is needed to better understand the community-level variance in LTBI and other contextual factors which may explain the spatial variation of LTBI.

Figure 4.1 Map of the study area Diepsloot, Johannesburg, South Africa.

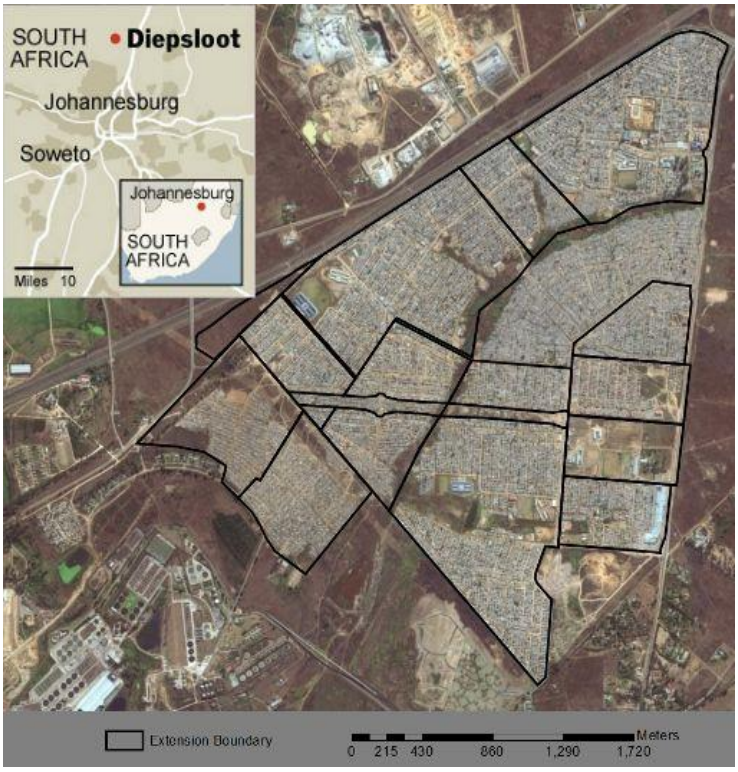


Figure 4.2 Spatial distribution of neighborhood-level LTBI prevalence in Diepsloot.

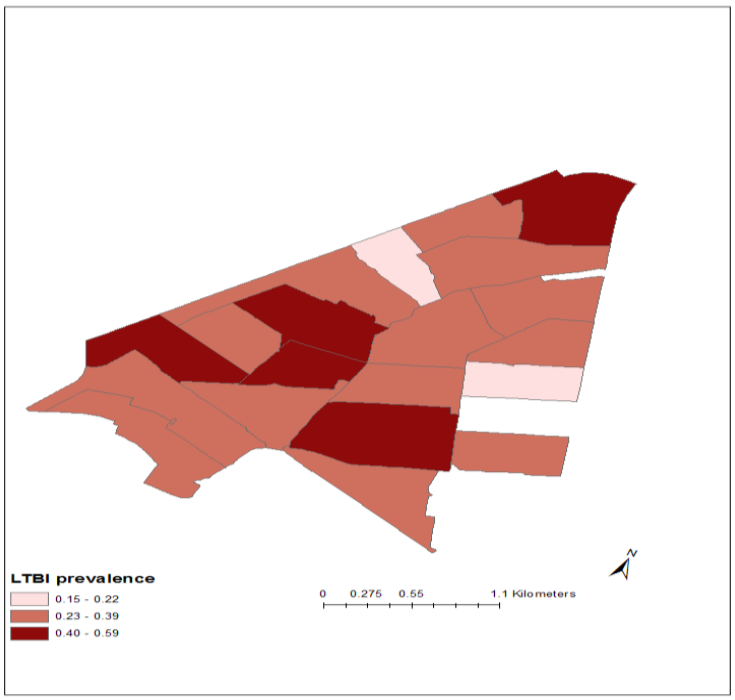


Figure 4.3 Distribution of spatial clusters of LTBI prevalence in Diepsloot.

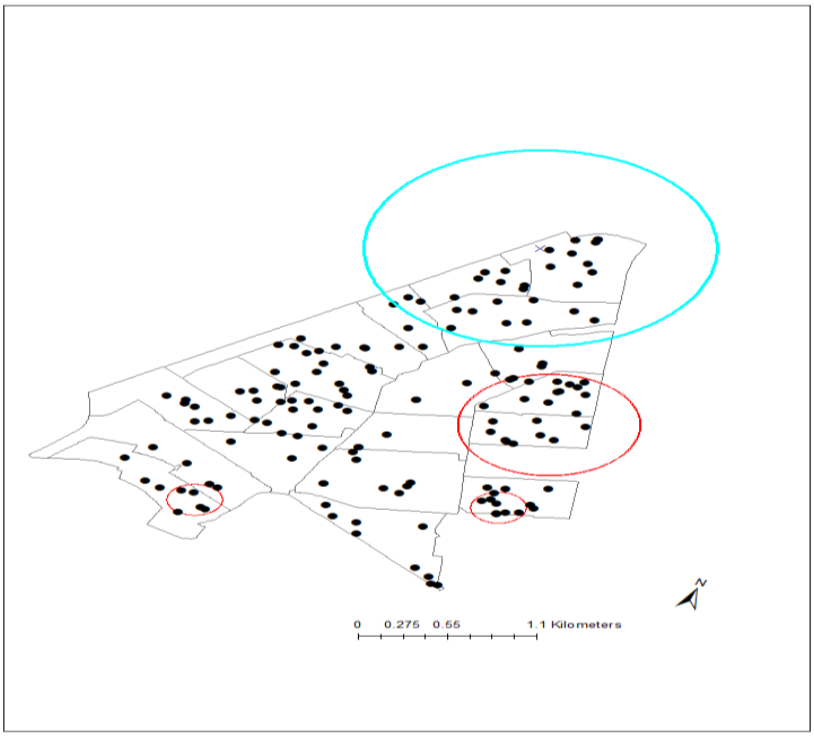


Figure 4.4 The spatial variation of residual, local R^2 and the parameter estimates from GWR model.

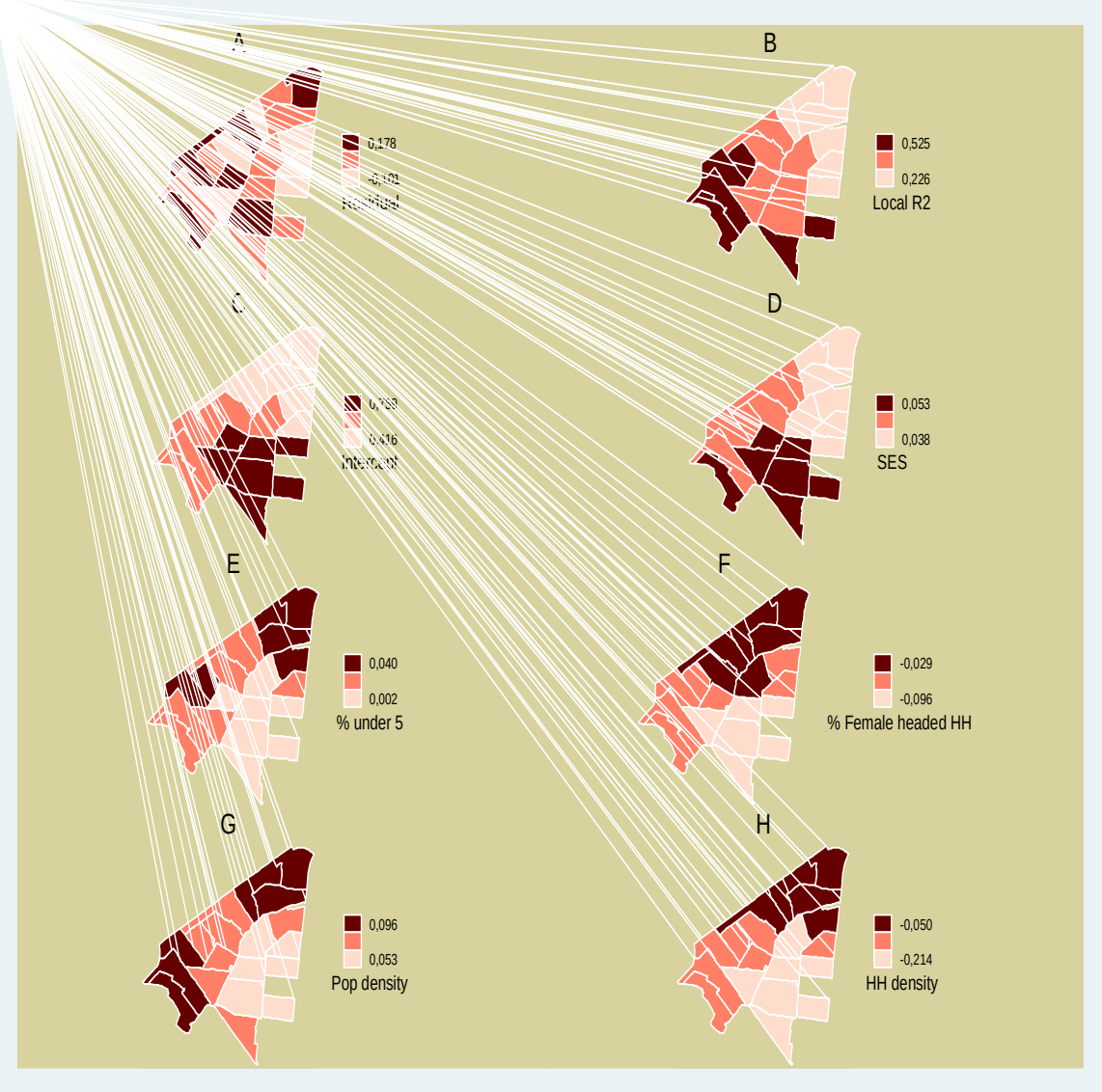


Table 4.1 Household asset-based variables included in a principal components analysis.

Household Assets	Source of drinking water
1 = yes; 0 = no	3=Piped water in house
Car ownership	2=Piped water in yard
Motorcycle ownership	1=Public Tap/Other
Bicycle ownership	
Refrigerator ownership	
TV ownership	
Radio ownership	
Mobile phone ownership	
House ownership	
Sharing toilet facilities with other households	

Table 4.2 Summary values of dependent and independent variables for 20 Diepsloot neighborhood used in the analysis.

Parameter	Min	Max	Mean	SD
LTBI prevalence	0.15	0.59	0.34	0.10
Socio-economic status	1.14	4.52	2.88	0.99
% of population under 5 years	10	10	12	1.00
% Female headed household	24	36	29	3.00
Population density, per km^2	15023	49566	28831	8767
Household size density	211	689	427	151
Population size, per neighborhood	2669	13738	6814	2692

Table 4.3 Characteristics of detected clusters of LTBI prevalence: Diepsloot, 2013-2014.

Cluster	No of areas in clusters	Radius (km)	No of observed cases	No of expected cases	RR	p-value
1	3	0.98	27	14.42	2.06	0.0332*
2	1	<0.50	8	3.53	2.33	0.4743
3	2	0.51	26	18.80	1.46	0.7615
4	1	<0.50	13	9.33	1.43	0.9855

*statistically significant at level of 0.05.

Table 4.4 Summary of Results from Ordinary Least Square regression model.

Parameter	Coefficient	SE	t-statistic	p-value	VIF[a]
Intercept	0.487	0.127	3.85	0.002*	-
Socio-economic status	0.039	0.018	2.17	0.048*	2.36
% of population under 5 years	0.042	0.042	1.00	0.335	1.70
% of female headed household	-0.056	0.056	-1.00	0.333	1.71
Population density	0.089	0.046	1.96	0.070	1.98
Household size density	-0.088	0.050	-1.76	0.101	1.45
Multiple R^2	0.319				
Adjusted R^2	0.075				
AICc	18.21				

*statistically significant at level of 0.05.
[a]Variance Inflation Factor

Table 4.5 Summary of results from Geographically Weighted Regression model.

Parameter	Min	1st Quartile	Median	3rd Quartile	Max	Mean
Intercept	0.416	0.476	0.537	0.580	0.759	0.543
Socio-economic status	0.037	0.039	0.043	0.045	0.053	0.043
% of population under 5 years	0.002	0.013	0.024	0.031	0.040	0.023
% of female headed household	-0.096	-0.059	-0.052	-0.046	-0.029	-0.055
Population density	0.052	0.073	0.082	0.089	0.096	0.080
Household size density	-0.214	-0.126	-0.104	-0.078	-0.050	-0.105
Multiple R^2	0.526					
Adjusted R^2	0.160					
AICc	10.84					

CHAPTER 5: DISCUSSION

1. Introduction

This study investigated the LTBI prevalence, spatial clustering and risk factors in a South African urban informal settlement. Specifically, we estimated the prevalence of LTBI in the general population, the ARI in children and investigated individual-, household- and neighborhood-level factors associated with LTBI (Aim 1). We hypothesized that, in this community, LTBI prevalence will be higher than 20%, ARI will be higher than 2%, and LTBI prevalence will be higher in older and HIV positive individuals. We further assessed spatial heterogeneity of LTBI prevalence and the association between community-level factors and LTBI clusters (Aim 2).We hypothesized that LTBI prevalence is not homogenous across this community and that LTBI prevalence is associated with neighborhood-level characteristics such as type of housing and median size of the household.

2. Summary of findings

For Aim 1, our study observed LTBI prevalence of 34% and the ARI of 3%, confirming our hypothesis that the prevalence of LTBI is high in an urban informal settlement in South Africa. We first explored individual-, household- and neighborhood-level factors associated with LTBI using multilevel logistic regression model. The multilevel ‘‘null’’ model showed that only 1% of the variance in LTBI was explained by differences in neighborhood factors, negating the need for a multilevel logistic regression model. When we fitted normal logistic regression, no neighborhood-level factors were associated with LTBI.

Individual- and household-level factors age, sex, marital status, being TB household contact and SES were associated with LTBI. HIV infection was common (18%) but not

associated with LTBI prevalence in this population. In addition, some other risk factors such as poor ventilation, smoking and exposure to household secondary smoking were not associated with LTBI prevalence in our study.

For aim 2, we observed neighborhood-level LTBI prevalence ranging from 15.0% to 59.1% in the Diepsloot community. Based on Monte Carlo hypothesis testing, the SaTScan software identified one primary cluster and three secondary clusters. In global (OLS) regression analysis, only SES level was associated with LTBI prevalence, with higher LTBI prevalence in neighborhoods with higher SES. The residuals of the OLS model were spatially random, which warranted fitting a local (GWR) model to further explore whether associations between the neighborhood-level covariates and LTBI prevalence observed in a global model vary spatially.

Compared with the global OLS model, the local GWR model improved model fit and explained a greater proportion of the variance in neighborhood-level LTBI prevalence. The GWR regression estimates showed both the direction and size of the spatial relationship between LTBI prevalence and predictors. While SES was the strongest factor associated with LTBI prevalence, the strength of the association varied between areas of the Diepsloot community. SES had the strongest positive association on LTBI prevalence in the south eastern part, and the lowest positive association on LTBI prevalence in the northern and eastern part of Diepsloot.

3. Interpretation of findings

LTBI prevalence and TB transmission in the Diepsloot community was high but lower than LTBI prevalence observed in previously published studies in South Africa. The ARI observed fell within the range of ARI previously reported in South Africa. Taking our findings together with those of previous studies suggests that LTBI prevalence in urban settlements is high but can show substantial variation, even within the same country. We were not able to demonstrate an association between neighborhood-level factors and LTBI prevalence, which

may have been due to the sparsity of level 2 clusters and the small number of 20 neighborhoods assessed (20 instead of the recommended minimum of 50 clusters).

We observed an association between LTBI and age, male gender, marital status and being a household close contact of index TB case, individuals-level risk factors that were also observed in prior studies. We were surprised to observe a higher LTBI prevalence among people with higher household SES. An association of higher SES rather than lower SES associated with higher LTBI prevalence was also observed in a study in Zambia,[102] and in a population-based multicentre study in China.[205] Taken together, these findings suggest that SES may have a differential effect on the risk of LTBI acquisition and risk of progression from infection to active TB disease. Boccia et al suggested that "it is possible that, especially in urban settings, higher SES is associated with housing characteristics that reduce ventilation and life-styles that increase social mixing and therefore the likelihood of contact between cases and susceptible people". We could not find an association between ventilation and LTBI, and higher SES was not associated with poorer ventilation in our sample. Other LTBI prevalence studies in high HIV burden settings have reported similar observations.[102, 192]

We were similarly surprised by a lack of association between poor ventilation, smoking and exposure to household secondary smoking[202, 203] were not associated with LTBI prevalence in our study. Lack of association between HIV and LTBI was (OR= 0.85, 95 % CI 0.49 – 1.46) was also found in a recently study in another urban African setting (OR= 0.91, 95 % CI 0.52 – 1.62).[192] Two other studies conducted in high burden HIV settings in sub-Saharan African reported lower odds of LTBI among HIV-infected individuals.[59, 193] The lack of association between HIV and LTBI may be due poor sensitivity of TST in HIV-infected individuals[66], however we addressed this by decreasing the TST cut-off to 5mm.[68]

We observed that the LTBI prevalence at a neighborhood level was not homogenous in its spatial distribution, with spatial clusters occurring within this relatively small (12 km^2) community. This demonstrated the utility of SATScan statistics in identifying LTBI hotspots even

in relative small area such as our study site. We observed a positive association between LTBI and SES, with higher LTBI prevalence in areas with higher SES, and the local GWR model showed that this association varied across space.

4. Strengths and limitations

Our study contributes valuable contributions to the epidemiologic literature on LTBI in high TB burden settings. Our study used a population-based representative sample to estimate the burden of LTBI. Very few studies in the literature have estimated the burden of LTBI in a general population. Most studies have either been conducted exclusively in children or healthy adults, or in high-risk populations such miners and health care workers. Our study may provide a better estimate of the true burden of LTBI in communities since our sample included all people living in the community, including those with and without HIV and both adults and children. Another strength of the study is that our population-based sample was drawn using random sampling, which resulted in a representative sample with broader generalizability to other urban poor resourced settlement in South Africa. Community-based LTBI prevalence studies are better positioned to inform targeted community-based interventions.

In addition, this study employed novel methodological tools to investigate the risk factors associated with LTBI prevalence at community level. Multilevel models are rarely applied in epidemiologic research of LTBI burden. Application of SATScan and GWR methods to examine the spatial distribution of LTBI burden is also rare. To our knowledge, our study is one of the first to employ spatial methods to assess the spatial distribution of LTBI in a country with a high burden of active TB in sub-Saharan Africa. We show that these methods can identify clusters with high risk of LTBI. This finding is helpful researchers who aim to understand the factors associated with clusters of high LTBI risk, and to policy makers who aim to address the public health implications of these clusters on TB control and prevention. The findings of our study may assist health planners and policy makers in identifying high-risk areas of LTBI and developing more effective targeted TB control and prevention strategies.

Moreover, we used a standardized approach to measure SES and to define SES tertiles. Traditionally, SES is measured using occupation, education and income. However, these measures reflect distinct aspects of SES and have many limitations.[237, 238] We used asset-based measures to measure SES and applied a principal components analysis (PCA) to the asset-based measures data to derive a SES tertiles,[164] an approach previously used in a South African context.[187, 239]

There are a few limitations to our study. First, the cross-sectional nature of the study does not allow for establishment of temporality or causality between LTBI and associated factors. However, it was not our aim to assess the causal relationship between risk factors and LTBI. We were rather interested in identifying factors that independently predict LTBI prevalence in a community. To address this limitation, we used predictive risk models instead of a causal inference model.

Second, some risk factors such occupation, crowding, and ventilation were not measured in great detail and BCG vaccination status, which can reduce the specificity of TST, was not documented.[208, 209] Third, the sample size was relatively small, especially for children under 12 years of age since we did not made exhaustive attempts to find this group of participants if they were not at home during the interview with the adult participants. Our sample size was further affected by the decision of the ethics committee overseeing the study. In response to a number of adverse events, the ethics committee restricted the placement of TST to HIV positive individuals and children under 5 years old. Finally, 19% of the targeted household was not enrolled due to failure to find someone at home despite multiple attempts or refusal to participate. Some selection bias may this have occurred and our aim to enroll a representative sample of the population may not have been fully achieved.

For our spatial analysis, the determinants of LTBI were assessed at the community level, which is appropriate given that LTBI risk is driven by prevalence of active TB in the community, but precludes inference of associations at the individual level. The low R^2 of the global and

local regression models suggests that factors other than those measured may be important determinants of the spatial distribution of LTBI prevalence within a community.

5. Conclusion

The prevalence of LTBI and the annual risk of infection with *M. tuberculosis* are high in urban populations, especially in men and older individuals, but independent of HIV infection status. The unexpected association between higher LTBI and higher household SES suggest that the differential association between SES as risk factors for acquisition of TB infection and progression from LTBI to active disease is not yet fully understood. We did not assess whether differences in use of public transportation or social mixing between SES could explain the observation of higher LTBI prevalence in people of higher SES within urban settlements. These hypotheses thus warrant further in-depth investigations. A better understanding of individual, household and community-level risk factors for LTBI will be important for the development of efficient, targeted LTBI interventions in high TB burden settings.

REFERENCES

1. WHO., *Global Tuberculosis Report 2014*. 2014: World Health Organization.

2. WHO., *Global Tuberculosis Report 2012*. 2012, WHO/HTM/TB.

3. Department of Health, M.R.C., OrcMacro, *South Africa Demographic and Health Survey 1998.* 1998.

4. Department of Health, M.R.C., OrcMacro, *South Africa Demographic and Health Survey 2003*. 2007, Department of Health Pretoria.

5. Den Boon, S., et al., *High prevalence of tuberculosis in previously treated patients, Cape Town, South Africa.* Emerging infectious diseases, 2007. **13**(8): p. 1189.

6. Shapiro, A.E., 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): p. 1110-1116.

7. Claassens, M., et al., *High prevalence of tuberculosis and insufficient case detection in two communities in the Western Cape, South Africa.* PloS one, 2013. **8**(4): p. e58689.

8. Godfrey-Faussett, P., et al., *Tuberculosis control and molecular epidemiology in a South African gold-mining community.* The Lancet, 2000. **356**(9235): p. 1066-1071.

9. Verver, S., et al., *Transmission of tuberculosis in a high incidence urban community in South Africa.* International journal of epidemiology, 2004. **33**(2): p. 351-357.

10. Wood, R., et al., *Burden of new and recurrent tuberculosis in a major South African city stratified by age and HIV-status.* PLoS One, 2011. **6**(10): p. e25098.

11. Claessens, N., et al., *High frequency of tuberculosis in households of index TB patients.* The international journal of tuberculosis and Lung disease, 2002. **6**(3): p. 266-269.

12. Arnadottir, T., *Tuberculosis: trends and the twenty-first century.* Scandinavian journal of infectious diseases, 2001. **33**(8): p. 563-567.

13. Arnadottir, T., *The Styblo model 20 years later: what holds true?[State of the art series. Tuberculosis. Edited by ID Rusen. Number 8 in the series].* The International Journal of Tuberculosis and Lung Disease, 2009. **13**(6): p. 672-690.

14. WHO, *STOP TB Partnership. The Stop TB Strategy. Building on and enhancing DOTS to meet TB-related millenium development goals.* World Health Organization, 2006. **37**: p. 1-20.

15. WHO, *STOP TB Partnership The Global Plan to Stop TB 2011-2015.* World Health Organization, 2010.

16. Organization, W.H., *The Global Plan to Stop TB 2006–2015: progress report 2006–2008.* Geneva: World Health Organization, 2009.

17. Dye, C., et al., *Global burden of tuberculosis: estimated incidence, prevalence, and mortality by country.* Jama, 1999. **282**(7): p. 677-686.

18. Dye, C., et al., *Targets for global tuberculosis control [Short Communication].* The International Journal of Tuberculosis and Lung Disease, 2006. **10**(4): p. 460-462.

19. Dye, C., *Tuberculosis 2000–2010: control, but not elimination [The Comstock Lecture].* The International Journal of Tuberculosis and Lung Disease, 2000. **4**(12s2): p. S146-S152.

20. Borgdorff, M.W., K. Floyd, and J.F. Broekmans, *Interventions to reduce tuberculosis mortality and transmission in low-and middle-income countries.* Bulletin of the World Health Organization, 2002. **80**(3): p. 217-227.

21. Dowdy, D.W. and R.E. Chaisson, *The persistence of tuberculosis in the age of DOTS: reassessing the effect of case detection.* Bulletin of the World Health Organization, 2009. **87**(4): p. 296-304.

22. Kasprowicz, V.O., et al., *Diagnosing latent tuberculosis in high-risk individuals: rising to the challenge in high-burden areas.* Journal of Infectious Diseases, 2011. **204**(suppl 4): p. S1168-S1178.

23. Lawn, S.D. and G. Churchyard, *Epidemiology of HIV-associated tuberculosis Running Head: Epidemiology of TB/HIV.* Current Opinion in HIV and AIDS, 2009. **4**(4): p. 325.

24. Mitka, M., *Group targets surging TB in Africa.* JAMA, 2005. **293**(22): p. 2707-2707.

25. Styblo, K., *The potential impact of AIDS on the tuberculosis situation in developed and developing countries.* Bulletin of the International Union against Tuberculosis and Lung Disease, 1988. **63**(2): p. 25.

26. Styblo, K., *The impact of HIV infection on the global epidemiology of tuberculosis.* Bulletin of the International Union against Tuberculosis and Lung Disease, 1991. **66**(1): p. 27-32.

27. Lönnroth, K., et al., *Tuberculosis control and elimination 2010–50: cure, care, and social development.* The Lancet, 2010. **375**(9728): p. 1814-1829.

28. Lönnroth, K., et al., *Drivers of tuberculosis epidemics: the role of risk factors and social determinants.* Social science & medicine, 2009. **68**(12): p. 2240-2246.

29. Lönnroth, K. and M. Raviglione. *Global epidemiology of tuberculosis: prospects for control*. in *Seminars in respiratory and critical care medicine*. 2008.

30. Squire, S.B., A. Obasi, and B. Nhlema-Simwaka, *The Global Plan to Stop TB: a unique opportunity to address poverty and the Millennium Development Goals.* The Lancet, 2006. **367**(9514): p. 955-957.

31. Lancet, T., *Tackling poverty in tuberculosis control.* The Lancet, 2006. **366**(9503): p. 2063.

32. Mack, U., et al., *LTBI: latent tuberculosis infection or lasting immune responses to M. tuberculosis? A TBNET consensus statement.* European Respiratory Journal, 2009. **33**(5): p. 956-973.

33. Esmail, H., C.E. Barry, and R.J. Wilkinson, *Understanding latent tuberculosis: the key to improved diagnostic and novel treatment strategies.* Drug discovery today, 2012. **17**(9): p. 514-521.

34. Druszczyńska, M., et al., *Latent M. tuberculosis infection--pathogenesis, diagnosis, treatment and prevention strategies.* Pol J Microbiol, 2012. **61**(1): p. 3-10.

35. Dye, C., et al., *Tuberculosis*. Disease and mortality in sub-Saharan Africa, ed. D.T. Jamison. 2006: World Bank Publications.

36. Sutherland, I., *Recent studies in the epidemiology of tuberculosis, based on the risk of being infected with tubercle bacilli.* Advances in tuberculosis research. Fortschritte der Tuberkuloseforschung. Progres de l'exploration de la tuberculose, 1975. **19**: p. 1-63.

37. Aaron, L., et al., *Tuberculosis in HIV-infected patients: a comprehensive review.* Clinical microbiology and infection, 2004. **10**(5): p. 388-398.

38. Pawlowski, A., et al., *Tuberculosis and HIV co-infection.* PLoS Pathog, 2012. **8**(2): p. e1002464.

39. Selwyn, P.A., et al., *A prospective study of the risk of tuberculosis among intravenous drug users with human immunodeficiency virus infection.* New England journal of medicine, 1989. **320**(9): p. 545-550.

40. Kranzer, K., et al., *High prevalence of self-reported undiagnosed HIV despite high coverage of HIV testing: a cross-sectional population based sero-survey in South Africa.* PLoS One, 2011. **6**(9): p. e25244.

41. Marais, B., H. Rabie, and M. Cotton, *TB and HIV in children–advances in prevention and management.* Paediatric respiratory reviews, 2011. **12**(1): p. 39-45.

42. Marais, B.J. and H.S. Schaaf, *Childhood tuberculosis: an emerging and previously neglected problem.* Infectious disease clinics of North America, 2010. **24**(3): p. 727-749.

43. Loeffler, A.M. *Pediatric tuberculosis*. in *Seminars in respiratory infections*. 2003. [Orlando, Fla.]: Grune & Stratton, c1986-2003.

44. Diel, R., et al., *Old ideas to innovate tuberculosis control: preventive treatment to achieve elimination.* European Respiratory Journal, 2013. **42**(3): p. 785-801.

45. Zumla, A., et al., *Tuberculosis.* New England Journal of Medicine, 2013. **368**(8): p. 745-755.

46. Bosman, M., et al., *National tuberculin survey of Kenya, 1986–1990.* The International Journal of Tuberculosis and Lung Disease, 1998. **2**(4): p. 272-280.

47. Gustafson, P., et al., *Risk factors for positive tuberculin skin test in Guinea-Bissau.* Epidemiology, 2007. **18**(3): p. 340-347.

48. Hoa, N., et al., *First national tuberculin survey in Viet Nam: characteristics and association with tuberculosis prevalence.* The International Journal of Tuberculosis and Lung Disease, 2013. **17**(6): p. 738-744.

49. Legesse, M., et al., *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): p. 89.

50. Bennett, D.E., et al., *Prevalence of tuberculosis infection in the United States population: the national health and nutrition examination survey, 1999–2000.* American journal of respiratory and critical care medicine, 2008. **177**(3): p. 348-355.

51. Horne, D.J., et al., *Association between smoking and latent tuberculosis in the US population: an analysis of the National Health and Nutrition Examination Survey.* 2012.

52. Diez, M., et al., *Prevalence of M. tuberculosis infection and tuberculosis disease among HIV-infected people in Spain.* The international journal of tuberculosis and lung disease: the official journal of the International Union against Tuberculosis and Lung Disease, 2007. **11**(11): p. 1196-1202.

53. Kall, M.M., et al., *Latent and subclinical tuberculosis in HIV infected patients: a cross-sectional study.* BMC infectious diseases, 2012. **12**(1): p. 107.

54. Fox, G.J., et al., *Contact investigation for tuberculosis: a systematic review and meta-analysis.* European Respiratory Journal, 2013. **41**(1): p. 140-156.

55. Corbett, E.L., et al., *The growing burden of tuberculosis: global trends and interactions with the HIV epidemic.* Archives of internal medicine, 2003. **163**(9): p. 1009-1021.

56. Corbett, E.L., et al., *Tuberculosis in sub-Saharan Africa: opportunities, challenges, and change in the era of antiretroviral treatment.* The Lancet, 2006. **367**(9514): p. 926-937.

57. Fourie, P.B., *The prevalence and annual rate of tuberculous infection in South Africa.* Tubercle, 1983. **64**(3): p. 181-92.

58. Middelkoop, K., et al., *Rates of tuberculosis transmission to children and adolescents in a community with a high prevalence of HIV infection among adults.* Clinical Infectious Diseases, 2008. **47**(3): p. 349-355.

59. Hanifa, Y., et al., *Prevalence of latent tuberculosis infection among gold miners in South Africa.* The International Journal of Tuberculosis and Lung Disease, 2009. **13**(1): p. 39-46.

60. Shah, M., et al., *Longitudinal analysis of QuantiFERON-TB Gold In-Tube in children with adult household tuberculosis contact in South Africa: a prospective cohort study.* PloS one, 2011. **6**(10): p. e26787.

61. Oni, T., et al., *Smoking, BCG and employment and the risk of tuberculosis infection in HIV-infected persons in South Africa.* 2012.

62. Cauthen, G., A. Pio, and H. Ten Dam, *Annual risk of tuberculous infection.* Bulletin of the World Health Organization, 2002. **80**(6): p. 503-511.

63. Murray, C.J., K. Styblo, and A. Rouillon, *Tuberculosis in developing countries: burden, intervention and cost.* Bulletin of International Union against Tuberculosis and Lung Disease, 1990. **65**(1): p. 6-24.

64. Rodrigues, L.C. and P.G. Smith, *Tuberculosis in developing countries and methods for its control.* Transactions of the Royal Society of Tropical Medicine and Hygiene, 1990. **84**(5): p. 739-744.

65. Wood, R., et al., *Changing prevalence of TB infection with increasing age in high TB 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): p. 406.

66. Pai, M., S. Kalantri, and K. Dheda, *New tools and emerging technologies for the diagnosis of tuberculosis: Part I. Latent tuberculosis.* Expert Review of Molecular Diagnostics, 2006. **6**(3): p. 413-422.

67. Pai, M. and R. O'Brien. *New diagnostics for latent and active tuberculosis: state of the art and future prospects*. in *Seminars in respiratory and critical care medicine*. 2008.

68. Kaplan, J.E., et al., *Guidelines for prevention and treatment of opportunistic infections in HIV-infected adults and adolescents.* MMWR Recomm Rep, 2009. **58**(RR-4): p. 1-207.

69. Watkins, R.E., R. Brennan, and A.J. Plant, *Tuberculin reactivity and the risk of tuberculosis: a review.* Int J Tuberc Lung Dis, 2000. **4**(10): p. 895-903.

70. Akolo, C., et al., *Treatment of latent tuberculosis infection in HIV infected persons.* Cochrane Database Syst Rev, 2010(1): p. Cd000171.

71. Nelson, K., *Tuberculin testing to detect latent tuberculosis in developing countries.* Epidemiology, 2007. **18**(3): p. 348-9.

72. Nelson, K., *Tuberculin testing to detect latent tuberculosis in developing countries.* Epidemiology, 2007. **18**(3): p. 348-349.

73. WHO, S.T., *Partnership: The Global Plan to Stop TB 2011–2015.* Geneva: World Health Organization, 2011.

74. Lienhardt, C., et al., *Investigation of environmental and host-related risk factors for tuberculosis in Africa. I. Methodological aspects of a combined design.* American journal of epidemiology, 2002. **155**(11): p. 1066-1073.

75. Lienhardt, C., et al., *Risk factors for tuberculosis infection in sub-Saharan Africa: a contact study in The Gambia.* American journal of respiratory and critical care medicine, 2003. **168**(4): p. 448-455.

76. Hargreaves, J.R., et al., *The social determinants of tuberculosis: from evidence to action.* Am J Public Health, 2011. **101**(4): p. 654-62.

77. Nava-Aguilera, E., et al., *Risk factors associated with recent transmission of tuberculosis: systematic review and meta-analysis [Review article].* The International Journal of Tuberculosis and Lung Disease, 2009. **13**(1): p. 17-26.

78. Morrison, J., M. Pai, and P.C. Hopewell, *Tuberculosis and latent tuberculosis infection in close contacts of people with pulmonary tuberculosis in low-income and middle-income countries: a systematic review and meta-analysis.* The Lancet Infectious Diseases, 2008. **8**(6): p. 359-368.

79. Nguyen, T.H., et al., *Risk of latent tuberculosis infection in children living in households with tuberculosis patients: a cross sectional survey in remote northern Lao People's Democratic Republic.* BMC infectious diseases, 2009. **9**(1): p. 96.

80. Peltola, H., et al., *Risk of infection with Mycobacterium tuberculosis among children and mothers in Somalia.* Clinical infectious diseases, 1994. **18**(1): p. 106-111.

81. Sinfield, R., et al., *Risk factors for TB infection and disease in young childhood contacts in Malawi.* Annals of Tropical Paediatrics: International Child Health, 2006. **26**(3): p. 205-213.

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

83. Singh, V. and S. Patra, *A relook at preventive therapy for tuberculosis in children.* Indian J Pediatr, 2011. **78**(2): p. 205-10.

84. Mutsvangwa, J., et al., *Identifying recent Mycobacterium tuberculosis transmission in the setting of high HIV and TB burden.* Thorax, 2010. **65**(4): p. 315-320.

85. Shanaube, K., et al., *Risk factors associated with positive QuantiFERON-TB Gold In-Tube and tuberculin skin tests results in Zambia and South Africa.* PLoS One, 2011. **6**(4): p. e18206.

86. Hill, P.C., et al., *Large-scale evaluation of enzyme-linked immunospot assay and skin test for diagnosis of Mycobacterium tuberculosis infection against a gradient of exposure in The Gambia.* Clinical Infectious Diseases, 2004. **38**(7): p. 966-973.

87. Mahomed, H., 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): p. 331-336.

88. Verver, S., et al., *Proportion of tuberculosis transmission that takes place in households in a high-incidence area.* The Lancet, 2004. **363**(9404): p. 212-214.

89. Marais, B., et al., *The natural history of childhood intra-thoracic tuberculosis: a critical review of literature from the pre-chemotherapy era [State of the Art].* The International Journal of Tuberculosis and Lung Disease, 2004. **8**(4): p. 392-402.

90. Marais, B.J., et al., *Childhood pulmonary tuberculosis: old wisdom and new challenges.* American journal of respiratory and critical care medicine, 2006. **173**(10): p. 1078-1090.

91. Salazar-Vergara, R., et al., *Tuberculosis infection and disease in children living in households of Filipino patients with tuberculosis: a preliminary report.* The International Journal of Tuberculosis and Lung Disease, 2003. **7**(Supplement 3): p. S494-S500.

92. Dagnew, A.F., et al., *Diagnosis of latent tuberculosis infection in healthy young adults in a country with high tuberculosis burden and BCG vaccination at birth.* BMC research notes, 2012. **5**(1): p. 415.

93. Salaniponi, F., et al., *Risk of infection with Mycobacterium tuberculosis in Malawi: national tuberculin survey 1994.* The International Journal of Tuberculosis and Lung Disease, 2004. **8**(6): p. 718-723.

94. Akhtar, S. and S.K. Rathi, *Multilevel modeling of household contextual determinants of tuberculin skin test positivity among contacts of infectious tuberculosis patients, Umerkot, Pakistan.* The American journal of tropical medicine and hygiene, 2009. **80**(3): p. 351-358.

95. Middelkoop, K., et al., *Force of tuberculosis infection among adolescents in a high HIV and TB prevalence community: a cross-sectional observation study.* BMC infectious diseases, 2011. **11**(1): p. 156.

96. Pawlowski, A., et al., *Tuberculosis and HIV co-infection.* PLoS pathogens, 2012. **8**(2): p. e1002464.

97. Sheriff, F.G., et al., *Latent tuberculosis among pregnant mothers in a resource poor setting in Northern Tanzania: a cross-sectional study.* BMC infectious diseases, 2010. **10**(1): p. 52.

98. Holmes, C., H. Hausler, and P. Nunn, *A review of sex differences in the epidemiology of tuberculosis.* The International Journal of Tuberculosis and Lung Disease, 1998. **2**(2): p. 96-104.

99. Thorson, A. and V.K. Diwan, *Gender inequalities in tuberculosis: aspects of infection, notification rates, and compliance.* Current opinion in pulmonary medicine, 2001. **7**(3): p. 165-169.

100. Nahid, P., et al., *Racial differences in tuberculosis infection in United States communities: the coronary artery risk development in young adults study.* Clinical infectious diseases, 2011. **53**(3): p. 291-294.

101. Muniyandi, M. and R. Ramachandran, *Socioeconomic inequalities of tuberculosis in India.* Expert Opin. Pharmacother, 2008. **9**(10): p. 1623-1628.

102. Boccia, D., et al., *Tuberculosis infection in Zambia: the association with relative wealth.* The American journal of tropical medicine and hygiene, 2009. **80**(6): p. 1004-1011.

103. Van Leth, F., et al., *Measuring socio-economic data in tuberculosis prevalence surveys.* The International Journal of Tuberculosis and Lung Disease, 2011. **15**(Supplement 2): p. S58-S63.

104. Boccia, D., et al., *The measurement of household socio-economic position in tuberculosis prevalence surveys: a sensitivity analysis.* The International Journal of Tuberculosis and Lung Disease, 2013. **17**(1): p. 39-45.

105. Dye, C., *Health and urban living.* Science, 2008. **319**(5864): p. 766-769.

106. Eisenberg, J.N., et al., *Environmental determinants of infectious disease: a framework for tracking causal links and guiding public health research.* Environmental Health Perspectives, 2007: p. 1216-1223.

107. Kjellstrom, T., et al., *Our cities our health our future. Acting on social determinants for health equity in urban settings. Report to the WHO Commission on Social Determinants of Health from the Knowledge Network on Urban Settings.* 2007.

108. Dye, C. and B.G. Williams, *The population dynamics and control of tuberculosis.* Science, 2010. **328**(5980): p. 856-861.

109. Brockerhoff, M. and E. Brennan, *The poverty of cities in developing regions.* Population and development review, 1998: p. 75-114.

110. Chadha, V., et al., *Trends in the annual risk of tuberculous infection in India.* The International journal of tuberculosis and lung disease, 2013. **17**(3): p. 312-319.

111. Jensen, A.V., et al., *The prevalence of latent Mycobacterium tuberculosis infection based on an interferon-γ release assay: a cross-sectional survey among urban adults in Mwanza, Tanzania.* 2013.

112. Redman, C.L. and N.S. Jones, *The environmental, social, and health dimensions of urban expansion.* Population & Environment, 2005. **26**(6): p. 505-520.

113. Tornee, S., et al., *Risk factors for tuberculosis infection among household contacts in Bangkok, Thailand.* 2004.

114. Cohn, D.L., et al., *Targeted tuberculin testing and treatment of latent tuberculosis infection.* MMWR Morb Mortal Wkly Rep, 2000. **49**(6): p. 1-54.

115. Abu-Raddad, L.J., et al., *Epidemiological benefits of more-effective tuberculosis vaccines, drugs, and diagnostics.* Proceedings of the National Academy of Sciences, 2009. **106**(33): p. 13980-13985.

116. Dowdy, D.W., et al., *Heterogeneity in tuberculosis transmission and the role of geographic hotspots in propagating epidemics.* Proc Natl Acad Sci U S A, 2012. **109**(24): p. 9557-62.

117. Diel, R., et al., *Old ideas to innovate tuberculosis control: preventive treatment to achieve elimination.* Eur Respir J, 2013. **42**(3): p. 785-801.

118. Pfeiffer, D., et al., *Spatial analysis in epidemiology.* 2008.

119. Golub, A., W.L. Gorr, and P.R. Gould, *Spatial diffusion of the HIV/AIDS epidemic: modeling implications and case study of AIDS incidence in Ohio.* Geographical analysis, 1993. **25**(2): p. 85-100.

120. Hightower, A.W., et al., *A geographic information system applied to a malaria field study in western Kenya.* The American journal of tropical medicine and hygiene, 1998. **58**(3): p. 266-272.

121. Kleinschmidt, I., et al., *A spatial statistical approach to malaria mapping.* International Journal of Epidemiology, 2000. **29**(2): p. 355-361.

122. Schellenberg, J.A., et al., *An analysis of the geographical distribution of severe malaria in children in Kilifi District, Kenya.* International Journal of Epidemiology, 1998. **27**(2): p. 323-329.

123. Tanser, F., et al., *HIV heterogeneity and proximity of homestead to roads in rural South Africa: an exploration using a geographical information system.* Tropical Medicine & International Health, 2000. **5**(1): p. 40-46.

124. Beyers, N., et al., *The use of a geographical information system (GIS) to evaluate the distribution of tuberculosis in a high-incidence community.* 1996.

125. Chan-Yeung, M., et al., *Socio-demographic and geographic indicators and distribution of tuberculosis in Hong Kong: a spatial analysis.* The International Journal of Tuberculosis and Lung Disease, 2005. **9**(12): p. 1320-1326.

126. Kistemann, T., A. Munzinger, and F. Dangendorf, *Spatial patterns of tuberculosis incidence in Cologne (Germany).* Social science & medicine, 2002. **55**(1): p. 7-19.

127. Moonan, P.K., et al., *Using genotyping and geospatial scanning to estimate recent Mycobacterium tuberculosis transmission, United States.* Emerg Infect Dis, 2012. **18**(3): p. 458-465.

128. Munch, Z., et al., *Tuberculosis transmission patterns in a high-incidence area: a spatial analysis.* The international journal of tuberculosis and lung disease, 2003. **7**(3): p. 271-277.

129. Randremanana, R.V., et al., *Spatial clustering of pulmonary tuberculosis and impact of the care factors in Antananarivo City.* Trop Med Int Health, 2009. **14**(4): p. 429-37.

130. Tiwari, N., et al., *Investigation of geo-spatial hotspots for the occurrence of tuberculosis in Almora district, India, using GIS and spatial scan statistic.* International Journal of Health Geographics, 2006. **5**(1): p. 33.

131. Jia, Z.-W., et al., *Spatial analysis of tuberculosis cases in migrants and permanent residents, Beijing, 2000–2006.* Emerging infectious diseases, 2008. **14**(9): p. 1413.

132. Liu, Y., et al., *Investigation of space-time clusters and geospatial hot spots for the occurrence of tuberculosis in Beijing.* The International Journal of Tuberculosis and Lung Disease, 2012. **16**(4): p. 486-491.

133. Wang, T., et al., *The spatial epidemiology of tuberculosis in Linyi City, China, 2005–2010.* BMC public health, 2012. **12**(1): p. 885.

134. Touray, K., et al., *Spatial analysis of tuberculosis in an urban west African setting: is there evidence of clustering?* Tropical Medicine & International Health, 2010. **15**(6): p. 664-672.

135. Feske, M.L., et al., *Including the third dimension: a spatial analysis of TB cases in Houston Harris County.* Tuberculosis, 2011. **91**: p. S24-S33.

136. Lai, P.C., et al., *Risk of tuberculosis in high-rise and high density dwellings: an exploratory spatial analysis.* Environ Pollut, 2013. **183**: p. 40-5.

137. Li, T., et al., *Impact of new migrant populations on the spatial distribution of tuberculosis in Beijing.* Int J Tuberc Lung Dis, 2011. **15**(2): p. 163-8, i-iii.

138. Low, C.-T., et al., *Exploring tuberculosis by types of housing development.* Social Science & Medicine, 2013. **87**: p. 77-83.

139. Maciel, E.L., et al., *Spatial patterns of pulmonary tuberculosis incidence and their relationship to socio-economic status in Vitoria, Brazil.* Int J Tuberc Lung Dis, 2010. **14**(11): p. 1395-402.

140. de Queiroga, R.P., et al., *Spatial distribution of tuberculosis and relationship with living conditions in an urban area of Campina Grande--2004 to 2007.* Rev Bras Epidemiol, 2012. **15**(1): p. 222-32.

141. Houlihan, C.F., et al., *The tuberculosis challenge in a rural South African HIV programme.* BMC Infect Dis, 2010. **10**: p. 23.

142. Kandala, N.-B., M.A. Magadi, and N.J. Madise, *An investigation of district spatial variations of childhood diarrhoea and fever morbidity in Malawi.* Social Science & Medicine, 2006. **62**(5): p. 1138-1152.

143. Cegielski, J.P., et al., *Eliminating tuberculosis one neighborhood at a time.* Am J Public Health, 2013. **103**(7): p. 1292-300.

144. Guwatudde, D., et al., *Tuberculosis in Household Contacts of Infectious Cases in Kampala, Uganda.* Am J Epidemiol, 2003. **158**(9): p. 887-98.

145. Murray, E., et al., *A multidisciplinary method to map potential tuberculosis transmission 'hot spots' in high-burden communities.* The International Journal of Tuberculosis and Lung Disease, 2009. **13**(6): p. 767-774.

146. Classen, C.N., et al., *Impact of social interactions in the community on the transmission of tuberculosis in a high incidence area.* Thorax, 1999. **54**(2): p. 136-40.

147. Feske, M.L., et al., *Giving TB wheels: Public transportation as a risk factor for tuberculosis transmission.* Tuberculosis (Edinb), 2011. **91 Suppl 1**: p. S16-23.

148. Addo, K.K., et al., *A tuberculin skin test survey among Ghanaian school children.* BMC Public Health, 2010. **10**: p. 35.

149. Alonso-Echanove, J., et al., *Occupational transmission of Mycobacterium tuberculosis to health care workers in a university hospital in Lima, Peru.* Clin Infect Dis, 2001. **33**(5): p. 589-96.

150. Drobniewski, F., et al., *Rates of latent tuberculosis in health care staff in Russia.* PLoS Med, 2007. **4**(2): p. e55.

151. Jong Lee, K., et al., *Screening for latent tuberculosis infection in South Korean healthcare workers using a tuberculin skin test and whole blood interferon-gamma assay.* Scand J Infect Dis, 2010. **42**(9): p. 672-8.

152. Menzies, D., R. Joshi, and M. Pai, *Risk of tuberculosis infection and disease associated with work in health care settings.* Int J Tuberc Lung Dis, 2007. **11**(6): p. 593-605.

153. Rafiza, S., K.G. Rampal, and A. Tahir, *Prevalence and risk factors of latent tuberculosis infection among health care workers in Malaysia.* BMC Infect Dis, 2011. **11**: p. 19.

154. Hossain, S., et al., *Tuberculin survey in Bangladesh, 2007-2009: prevalence of tuberculous infection and implications for TB control.* Int J Tuberc Lung Dis, 2013. **17**(10): p. 1267-72.

155. de Alencar Ximenes, R.A., et al., *Is it better to be rich in a poor area or poor in a rich area? A multilevel analysis of a case-control study of social determinants of tuberculosis.* Int J Epidemiol, 2009. **38**(5): p. 1285-96.

156. Harling, G., R. Ehrlich, and L. Myer, *The social epidemiology of tuberculosis in South Africa: a multilevel analysis.* Soc Sci Med, 2008. **66**(2): p. 492-505.

157. Oren, E., et al., *Area-level socioeconomic disadvantage and severe pulmonary tuberculosis: U.S., 2000-2008.* Public Health Rep, 2013. **128**(2): p. 99-109.

158. Middelkoop, K., et al., *Childhood tuberculosis infection and disease: a spatial and temporal transmission analysis in a South African township.* SAMJ: South African Medical Journal, 2009. **99**(10): p. 738-743.

159. Middelkoop, K., et al., *Decreasing household contribution to TB transmission with age: a retrospective geographic analysis of young people in a South African township.* BMC infectious diseases, 2014. **14**(1): p. 221.

160. Hargreaves, J.R., et al., *The social determinants of tuberculosis: from evidence to action.* American journal of public health, 2011. **101**(4): p. 654-662.

161. Kish, L., *A procedure for objective respondent selection within the household.* Journal of the American Statistical Association, 1949. **44**(247): p. 380-387.

162. Cook, J.D., et al., *The influence of high-altitude living on body iron.* Blood, 2005. **106**(4): p. 1441-1446.

163. Myer, L., R.I. Ehrlich, and E.S. Susser, *Social epidemiology in South Africa.* Epidemiologic Reviews, 2004. **26**(1): p. 112-123.

164. Vyas, S. and L. Kumaranayake, *Constructing socio-economic status indices: how to use principal components analysis.* Health policy and planning, 2006. **21**(6): p. 459-468.

165. Diez-Roux, A.V., *Multilevel analysis in public health research.* Annual review of public health, 2000. **21**(1): p. 171-192.

166. Merlo, J., et al., *A brief conceptual tutorial of multilevel analysis in social epidemiology: linking the statistical concept of clustering to the idea of contextual phenomenon.* Journal of Epidemiology and Community Health, 2005. **59**(6): p. 443-449.

167. Merlo, J., et al., *A brief conceptual tutorial on multilevel analysis in social epidemiology: interpreting neighbourhood differences and the effect of neighbourhood characteristics on individual health.* Journal of Epidemiology and Community Health, 2005. **59**(12): p. 1022-1029.

168. Rabe-Hesketh, S., A. Skrondal, and A. Pickles, *GLLAMM manual.* University of California, Berkeley, 2004(UC Berkeley division of biostatistics working paper series).

169. Kulldorff, M., et al., *Cancer map patterns: are they random or not?* American journal of preventive medicine, 2006. **30**(2): p. S37-S49.

170. Aamodt, G., S. Samuelsen, and A. Skrondal, *A simulation study of three methods for detecting disease clusters.* International journal of health geographics, 2005. **5**: p. 15-15.

171. O'brien, R.M., *A caution regarding rules of thumb for variance inflation factors.* Quality & Quantity, 2007. **41**(5): p. 673-690.

172. Fotheringham, A.S., C. Brunsdon, and M. Charlton, *Geographically weighted regression: the analysis of spatially varying relationships*. 2003: Wiley. com.

173. Newson, R. and A.S. Team, *Multiple-test procedures and smile plots.* Stata J, 2003. **3**: p. 109-132.

174. Anselin, L., I. Syabri, and Y. Kho, *GeoDa: an introduction to spatial data analysis.* Geographical analysis, 2006. **38**(1): p. 5-22.

175. WHO., *Global tuberculosis report 2013.* 2013.

176. Statistics South Africa, *Mortality and causes of death in South Africa, 2013: Findings from death notification.* Statistical release P, 2014. **P0309.3**

177. Wood, R., et al., *Tuberculosis control has failed in South Africa: time to reappraise strategy.* SAMJ: South African Medical Journal, 2011. **101**: p. 111-114.

178. WHO., *Guidelines on the management of latent tuberculosis infection.* 2015.

179. Prevention., C.f.D.C., *Latent tuberculosis infection: a guide for primary health care providers*. 2013.

180. Dye, C., et al., *Prospects for tuberculosis elimination.* Annual review of public health, 2013. **34**: p. 271-286.

181. Van Rie, A., et al., *Prevalence, risk factors and risk perception of tuberculosis infection among medical students and healthcare workers in Johannesburg, South Africa.* SAMJ: South African Medical Journal, 2013. **103**(11): p. 853-857.

182. Shah, M., et al., *Longitudinal analysis of QuantiFERON-TB Gold In-Tube in children with adult household tuberculosis contact in South Africa: a prospective cohort study.* PloS one, 2011. **6**(10): p. e26787.

183. Mahomed, H., 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): p. 331-336.

184. Statistics South Africa, *Census 2011; Interactive data in SuperCROSS.* Pretoria 2011.

185. De Wet, T., et al., *Johannesburg poverty and livelihoods study.* Johannesburg: Centre for Social Development in Africa, University of Johannesburg, 2008.

186. Health, S.A.D.o., *National Tuberculosis Management Guidelines 2014*. 2014: Department of Health.

187. Booysen, F.R., *Using demographic and health surveys to measure poverty–an application to South Africa.* Journal for studies in Economics and Econometrics, 2002. **26**(3): p. 53-70.

188. Rieder, H., *Annual risk of infection with Mycobacterium tuberculosis.* European Respiratory Journal, 2005. **25**(1): p. 181-185.

189. Rieder, H., *Methodological issues in the estimation of the tuberculosis problem from tuberculin surveys.* Tubercle and Lung Disease, 1995. **76**(2): p. 114-121.

190. Bell, B.A., J.M. Ferron, and J.D. Kromrey, *Cluster size in multilevel models: the impact of sparse data structures on point and interval estimates in two-level models.* JSM Proceedings, Section on Survey Research Methods, 2008: p. 1122-1129.

191. Martinez, L., 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): p. 507-515.

192. Kizza, F.N., et al., *Prevalence of latent tuberculosis infection and associated risk factors in an urban African setting.* BMC Infectious Diseases, 2015. **15**(1): p. 165.

193. Shanaube K, H.J., Fielding K, Schaap A, Lawrence K-A, et al. , *Risk Factors Associated with Positive QuantiFERON-TB Gold In-Tube and Tuberculin Skin Tests Results in Zambia and South Africa.* PLoS ONE, 2011. **6**(4).

194. Kritzinger, F.E., 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): p. 136-142.

195. Zelner, J.L., et al., *Age-specific risks of tuberculosis infection from household and community exposures and opportunities for interventions in a high-burden setting.* Am J Epidemiol, 2014. **180**(8): p. 853-61.

196. Wood, R., et al., *Indoor social networks in a South African township: potential contribution of location to tuberculosis transmission.* PLoS One, 2012. **7**(6): p. e39246.

197. Johnstone-Robertson, S.P., et al., *Social mixing patterns within a South African township community: implications for respiratory disease transmission and control.* American journal of epidemiology, 2011: p. kwr251.

198. Wood, R., et al., *Tuberculosis transmission to young children in a South African community: modeling household and community infection risks.* Clinical infectious diseases, 2010. **51**(4): p. 401-408.

199. Andrews, J.R., et al., *Integrating social contact and environmental data in evaluating tuberculosis transmission in a South African township.* Journal of Infectious Diseases, 2014. **210**(4): p. 597-603.

200. den Boon, S., et al., *Association between smoking and tuberculosis infection: a population survey in a high tuberculosis incidence area.* Thorax, 2005. **60**(7): p. 555-7.

201. Uys, P., et al., *Transmission Elasticity in Communities Hyperendemic for Tuberculosis.* Clinical Infectious Diseases, 2011. **52**(12): p. 1399-1404.

202. Lindsay, R.P., et al., *The Association between active and passive smoking and latent tuberculosis infection in adults and children in the united states: results from NHANES.* PLoS One, 2014. **9**(3): p. e93137.

203. Patra, J., et al., *Exposure to second-hand smoke and the risk of tuberculosis in children and adults: a systematic review and meta-analysis of 18 observational studies.* PLoS Med, 2015. **12**(6): p. e1001835; discussion e1001835.

204. Narasimhan, P., et al., *Risk Factors for Tuberculosis.* Pulmonary Medicine, 2013. **2013**: p. 11.

205. Gao L, L.W., Bai L, Wang X, Xu J, Catanzaro A, Cárdenas V, Li X, Yang Y, JDu J, et al., *Latent tuberculosis infection in rural China: baseline results of a population-based, multicentre, prospective cohort study.* Lancet Infect Dis, 2015. **15**: p. 310-19.

206. Feske, M.L., et al., *Giving TB wheels: Public transportation as a risk factor for tuberculosis transmission.* Tuberculosis, 2011. **91, Supplement 1**(0): p. S16-S23.

207. Zamudio, C., et al., *Public transportation and tuberculosis transmission in a high incidence setting.* PLoS One, 2015. **10**(2): p. e0115230.

208. Farhat, M., et al., *False-positive tuberculin skin tests: what is the absolute effect of BCG and non-tuberculous mycobacteria?[Review Article].* The International Journal of Tuberculosis and Lung Disease, 2006. **10**(11): p. 1192-1204.

209. Tissot, F., et al., *Influence of bacille Calmette-Guérin vaccination on size of tuberculin skin test reaction: to what size?* Clinical infectious diseases, 2005. **40**(2): p. 211-217.

210. Denholm, J.T. and E.S. McBryde, *The use of anti-tuberculosis therapy for latent TB infection.* Infection and drug resistance, 2010. **3**: p. 63.

211. WHO., *Policy statement on preventive therapy against tuberculosis in people living with HIV. .* World Health Organization. Global Tuberculosis Programme and UNAIDS, 1998.

212. WHO., *Guidelines for intensified tuberculosis case-finding and isoniazid preventive therapy for people living with HIV in resource-constrained settings.* Geneva : World Health Organization, 2011.

213. Stop TB Partnership, *The Global Plan to Stop TB 2016-2020, Inital Working Draft.* World Health Organization, 2015.

214. Bhunu, C.P., W. Garira, and Z. Mukandavire, *Modeling HIV/AIDS and tuberculosis coinfection.* Bull Math Biol, 2009. **71**(7): p. 1745-80.

215. Cohen, T., et al., *Beneficial and perverse effects of isoniazid preventive therapy for latent tuberculosis infection in HIV-tuberculosis coinfected populations.* Proc Natl Acad Sci U S A, 2006. **103**(18): p. 7042-7.

216. Long, E.F., N.K. Vaidya, and M.L. Brandeau, *Controlling Co-Epidemics: Analysis of HIV and Tuberculosis Infection Dynamics.* Oper Res, 2008. **56**(6): p. 1366-1381.

217. Mills, H.L., T. Cohen, and C. Colijn, *Modelling the performance of isoniazid preventive therapy for reducing tuberculosis in HIV endemic settings: the effects of network structure.* J R Soc Interface, 2011. **8**(63): p. 1510-20.

218. Dye, C., et al., *Prospects for worldwide tuberculosis control under the WHO DOTS strategy.* The Lancet, 1998. **352**(9144): p. 1886-1891.

219. Esmail, H., et al., *The ongoing challenge of latent tuberculosis.* Philosophical Transactions of the Royal Society B: Biological Sciences, 2014. **369**(1645): p. 20130437.

220. WHO., *Guidelines on the management of latent tuberculosis infection.* 2014.

221. Grant, A., *Latent TB infection beyond HIV positives: why is it important?* , in *Session 29 at World Conference of Lung Health in Barcelona 2014*. 2014.

222. Churchyard, G.J., et al., *A Trial of Mass Isoniazid Preventive Therapy for Tuberculosis Control.* New England Journal of Medicine, 2014. **370**(4): p. 301-310.

223. Chan-yeung, M., et al., *Socio-demographic and geographic indicators and distribution of tuberculosis in Hong Kong: a spatial analysis.* Int J Tuberc Lung Dis, 2005. **9**(12): p. 1320-6.

224. Nunes, C., *Tuberculosis incidence in Portugal: spatiotemporal clustering.* International Journal of Health Geographics, 2007. **6**(1): p. 30.

225. Onozuka, D. and A. Hagihara, *Geographic prediction of tuberculosis clusters in Fukuoka, Japan, using the space-time scan statistic.* BMC infectious diseases, 2007. **7**(1): p. 26.

226. Randremanana, R.V., et al., *Spatial clustering of pulmonary tuberculosis and impact of the care factors in Antananarivo City.* Tropical Medicine & International Health, 2009. **14**(4): p. 429-437.

227. Tadesse, T., et al., *The clustering of smear-positive tuberculosis in Dabat, Ethiopia: a population based cross sectional study.* PloS one, 2013. **8**(5): p. e65022.

228. Nana Yakam, A., et al., *Spatial analysis of tuberculosis in Douala, Cameroon: clustering and links with socio-economic status.* Int J Tuberc Lung Dis, 2014. **18**(3): p. 292-7.

229. Nemeth, R. *Respondent selection within the household-A modification of the Kish grid*. in *Meeting of Young Statisticians*. 2002.

230. Pisati, M., *SPMAP: Stata module to visualize spatial data.* Statistical Software Components, 2008.

231. Kulldorff, M., *SaTScan v9.1.* http://www.satscan.org/, 2013.

232. Oni, T., et al., *Smoking, BCG and employment and the risk of tuberculosis infection in HIV-infected persons in South Africa.* PloS one, 2012. **7**(10): p. e47072.

233. Van Rie, A., et al., *Childhood tuberculosis in an urban population in South Africa: burden and risk factor.* Archives of disease in childhood, 1999. **80**(5): p. 433-437.

234. Souza, W.V., et al., *Tuberculosis in intra-urban settings: a Bayesian approach.* Tropical Medicine & International Health, 2007. **12**(3): p. 323-330.

235. Alvarez-Hernandez, G., et al., *An analysis of spatial and socio-economic determinants of tuberculosis in Hermosillo, Mexico, 2000–2006.* The International Journal of Tuberculosis and Lung Disease, 2010. **14**(6): p. 708-713.

236. Pelly, T., et al., *Tuberculosis skin testing, anergy and protein malnutrition in Peru.* The international journal of tuberculosis and lung disease: the official journal of the International Union against Tuberculosis and Lung Disease, 2005. **9**(9): p. 977.

237. Shavers, V.L., *Measurement of socioeconomic status in health disparities research.* Journal of the national medical association, 2007. **99**(9): p. 1013.

238. Braveman, P.A., et al., *Socioeconomic status in health research: one size does not fit all.* Jama, 2005. **294**(22): p. 2879-88.

239. Bärnighausen, T., et al., *The socioeconomic determinants of HIV incidence: evidence from a longitudinal, population-based study in rural South Africa.* AIDS (London, England), 2007. **21**(Suppl 7): p. S29-S38.