Towards a framework for improved targeting of tuberculosis interventions: a spatial analysis of patient notification and prevalence survey data from urban Blantyre, Malawi

McEwen Joseph Khundi

Thesis submitted in accordance with the requirements for the degree of

Doctor of Philosophy (PhD)
of the
University of London

March 2023

Department of Medical Statistics

Faculty of Epidemiology and Population Health

LONDON SCHOOL OF HYGIENE & TROPICAL MEDICINE

Funded by Wellcome Trust, UK

Supervisor: Professor James Carpenter
Professor of Medical Statistics
Department of Medical Statistics
Faculty of Epidemiology and Population Health
London School of Hygiene & Tropical Medicine

Co-supervisor: Professor Peter MacPherson
Wellcome Trust Fellow & Reader in Population Health
Department of Clinical Research
Liverpool School of Tropical Medicine &
Malawi-Liverpool-Wellcome Programme (MLW)

Other advisory committee members:
Professor Liz Corbett
Professor of Tropical Epidemiology
Department of Clinical Research
Faculty of Infectious and Tropical Diseases
London School of Hygiene & Tropical Medicine

Collaborating institutions:

1. Malawi Liverpool Wellcome Programme (MLW), Malawi
2. London School of Hygiene and Tropical Medicine (LSHTM), UK
3. Blantyre District Health Office, Malawi

Statement of own work

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT

T: +44 (0)20 7299 4646
F: +44 (0)20 7299 4656
www.lshtm.ac.uk

DECLARATION OF OWN WORK

All students are required to complete the following declaration when submitting their thesis.

Please note: Assessment misconduct includes any activity that compromises the integrity of your research or assessment of is will be considered under the Assessment Irregularity Policy. This includes plagiarism, cheating and failure to follow correct progression and examination procedures.

Please see the following documents for further guidance:

- Research Degrees Handbook
- Assessment Irregularities Policy

Supervisors should be consulted if there are any doubts about what is permissible.

1. STUDENT DETAILS

Student ID Number	lsh1805281	**Title**	Mr
First Name(s)	McEwen Joseph		
Surname/Family Name	Khundi		
Programme of Study	PhD		
LSHTM Email (if this is no longer active, please provide an alternative)	mcewen.khundi@lshtm.ac.uk		

2. TITLE OF THESIS

Title of Thesis	Towards a framework for improved targeting of tuberculosis interventions: a spatial analysis of patient notification and prevalence survey data from urban Blantyre, Malawi

3. DECLARATION

I have read and understood the LSHTM's definition of plagiarism and cheating. I declare that this thesis is my own work, and that I have acknowledged all results and quotations from the published or unpublished work of other people.

I have read and understood the LSHTM's definition and policy on the use of third parties (either paid or unpaid) who have contributed to the preparation of this thesis by providing copy editing and, or, proof reading services. I declare that no changes to the intellectual content or substance of this thesis were made as a result of this advice, and, that I have fully acknowledged all such contributions.

I have exercised reasonable care to ensure that the work is original and does not to the best of my knowledge break any UK law or infringe any third party's copyright or other intellectual property right.

Student Signature	[redacted]

Date	20 July 2022

Acknowledgements

I want to thank Professor Liz Corbett for mentoring me to get to the position where I could undertake this PhD. I joined her team in 2012 when I only had four months of working experience. I started as an assistant data manager, got promoted to the post of data manager, and was later promoted to research epidemiologist after completing my distance-based master's in epidemiology with the school (LSHTM). Her senior Wellcome Trust fellowship funded this PhD. The data that have been used for this thesis are also from the studies that were funded by the same fellowship.

I want to thank Professor James Carpenter and Professor Peter MacPherson for supervising and mentoring me from the start to the completion of this PhD. They assisted in reviewing my project objectives from when they were only ideas in a point format list, drafts of the manuscripts, and the PhD thesis. They were both gentle in correcting me when I would get things wrong and were quick to encourage me if I was getting discouraged by the PhD workload.

The MLW public health group's team deserves my gratitude. Those still in the group and those who have moved on to other projects. The Enhanced Monitoring Surveillance system, which provided the data for my thesis, was established in 2011 and has benefited from numerous significant contributions from different individuals over time. I want to thank Doctor Augustine Choko, the project's first data manager. Additionally, I'd like to express my gratitude to Deus Thindwa, the second data manager to work on the project. I would also like to thank Mr Wisdom Vinjeru Shonga for being the person who managed spatial-related data for MLW Public Health Group projects. He trained the field workers in collecting GPS data and was involved in curating the spatial data. Peter, Augustine, and Deus developed

the prototype system for identifying notified TB cases' neighbourhoods. Rebecca Harris, with the assistance of the Public Health Group at MLW, led the development of the electronic application (ePAL) for capturing GPS coordinates of notified TB cases.

I want to acknowledge my co-authors. They assisted in shaping the direction of the manuscripts and are also the individuals currently running the MLW Public Health Group trials.

I want to thank my family for being there for me. My parents have always encouraged me in my efforts to further my education. My top supporters are my two sisters, Muriel and Prudencia Khundi.

Thesis abstract

In 2020, Tuberculosis (TB) was the second-leading cause of death by a single infectious agent, trailing only COVID-19. Improved strategies for case finding and prompt treatment initiation are therefore a public health priority. However, routine TB case finding through health facilities depends on an individual recognising that they have TB symptoms and taking the initiative to seek care, and is not sufficient to control TB transmission. This is because a considerable number of people with TB symptoms delay and sometimes never attend healthcare facilities, meaning that they remain infectious in the community, and are at high risk of severe illness and death. By contrast, while community-based TB case-finding interventions are complementary to facility-based services, they are substantially more resource intensive, logistically challenging, and are generally only justifiable when targeted at groups of people at high risk of TB disease. Because the prevalence of microbiologically-confirmed pulmonary TB is difficult to measure and only rarely exceeds 1% even in the highest burden settings, identifying population groups likely to have a high burden of undiagnosed TB disease who could benefit from community-based case finding interventions can be challenging. Therefore, the aim of this thesis was to investigate how statistical spatial modelling of epidemiological patterns of spatial heterogeneity in urban TB epidemiology in Blantyre could be used to improve the targeting of community-based active case-finding (ACF) interventions for TB, and so increase the efficiency and effectiveness of the delivery of these interventions.

First, a systematic review of the effectiveness of spatially-targeted community ACF interventions demonstrated that spatially-targeted interventions are feasible and that they have potential as alternative design strategies for community ACF, but that more rigorous

approaches to design and analysis are required (Chapter 3, Khundi et al. 2021). The systematic review identified ten studies of spatially targeted interventions from six countries between 1 January 1993 and 22 March 2021: three directed against TB, three against leprosy, three against malaria, and one against HIV. Although data were limited, and understanding of effectiveness was limited by high risk of bias (particularly in classification of hotspots and ascertainment of outcomes), this demonstrated that spatially-targeted interventions have real potential to identify communities with a higher yield of identified cases, communities with a high prevalence of cases, and hence potential for accelerating reductions in the TB case notifications rates.

Second, multi-level spatial regression modelling of TB case fatality rates across the city of Blantyre (Chapter 4, Khundi et. al., 2021) – in which we evaluated 4397 newly-diagnosed TB cases, 10.9% (479) of whom died – found strong evidence that, while undergoing TB treatment, age, being HIV positive, and distance to TB treatment clinic were associated with an increased odd of death. Distance to TB treatment clinic is a proxy of ease of access to care: individuals that have to travel longer distances to get health care are at an increased risk of adverse health. In our study population, distance increased the odds of death only for patients that were registered at the referral clinic but not a primary health care clinic. This is consistent with the hypothesis that high quality facility-based TB screening and care is complementary to community-based interventions in reducing TB mortality.

Third, using data from a citywide TB prevalence survey and Malawi Liverpool Wellcome's Blantyre enhanced TB surveillance system (Chapter 5, Khundi et al., 2022), we developed a novel Bayesian statistical spatial modelling approach to enable identification of TB hotspot neighbourhoods, based on ranking of 72 neighbourhood prevalence-to-notification ratios. In

2019, the prevalence of microbiologically-confirmed TB in Blantyre was 215 per 100,000 population. The model derived mean neighbourhood prevalence-to-notification ratio was 4.49 (95% credible interval [CrI]: 0.98–11.91, range: 1.70–10.40, standard deviation: 1.79). This indicates that, overall, there remains a substantial burden of undiagnosed TB. Our model should support researchers and health workers in other settings with similar characteristics to urban Blantyre in identifying potential hotspot neighbourhoods without the need for conducting a full prevalence survey.

Looking forward, the approaches developed in this thesis, need to be validated, and then further developed and applied in other similar urban settings to prioritise neighbourhoods by burden of undiagnosed TB, and hence efficiently direct community based ACF interventions.

Table of contents

Abbreviations and acronyms

AIDS acquired immunodeficiency syndrome

ACF active case-finding

ART antiretroviral treatment

BCG bacilli Calmette-Guerin

CHW community health worker

CI confidence interval or credible interval

CNR Case notification rate

COVID-19 Coronavirus Disease 2019

CrI credible interval

CRP C-reactive protein

DAG Directed Acyclic Graph

DOTS directly observed treatment , short-course

ePAL electronic participant locator

GPS global positioning system

HIV human immunodeficiency viruse

IGRA interferon gamma release assay

IPT isoniazid preventive treatment

IRS indoor residual spraying

LAM lateral flow lipoarabinomannan

LLINs Long lasting insecticidal nets

LMIC Low and middle income countries

LTBI Latent Tuberculosis Infection

KM Kilometre

MGIT Mycobacteria Growth Indicator Tube

MLW Malawi Liverpool Wellcome Trust

MTB Mycobacterium Bovis

NTP National Tuberculosis Programme

NSO National Statistical Office

OR odds ratio

PCF passive case finding

P:N ratio prevalence to notification ratio

PROSPERO The International Prospective Register of Systematic Reviews

SD standard deviation

SDG sustainable development goals

QECH Queen Elizabeth Central Hospital

TB Tuberculosis

TST tuberculin skin test

WHO World Health Organisation

ZAMSTAR Zambia South Africa Tuberculosis and AIDS reduction

Chapter 1: Introduction

1.1 Burden of TB globally, and in Africa

Tuberculosis (TB) is among the top ten leading causes of death in the world and is the second leading cause of death from a single infectious agent, after COVID-19 (1). In 2020, 10.0 million people were estimated to have developed active TB disease. In addition, TB caused an estimated 1.3 million deaths among HIV negative individuals and a further 214,000 deaths among people living with HIV (1). The highest burden was in adult men (56% of deaths) followed by adult women (33%) and children (11%). The COVID-19 pandemic, which started in 2019, severely affected essential TB services, with an estimated 7.1% increase in TB deaths in 2020 compared to 2019 (1). In addition, the gap between incident cases and notified cases widened in 2020, rising to 41% from 29% in 2019 (1,2).

Geographically in 2020, most TB cases were in the WHO regions of South-East Asia (43%), Africa (25%), and the Western Pacific (18%), with smaller shares in the Eastern Mediterranean (8.2%), the Americas (2.9%) and Europe (2.3%) (1). The global reduction in estimated TB incidence between 2015 and 2019 was 11%, only halfway towards the 2020 EndTB milestone (20%) (1). The African region narrowly missed the target, achieving a reduction of 19% (1). The global reduction in TB deaths between 2015 and 2019 was 9% against a target of 35%. However, the African continent made relatively good progress, achieving a reduction of 18% (1).

The first systematic global estimate of TB incidence was done in 1990 by the WHO, which estimated that there were about 8 million incident TB cases and 2.6 to 2.9 million deaths (3). Globally, between 2006 and 2016, there was an annual reduction in age standardised incidence of TB among HIV negative individuals of -1.3% (95% confidence interval (CI): -1.5%

to -1.2%), which was a slower rate compared to the period from 1990 to 2006 (-1.6%, 95% CI: -1.8 to -1.3). However, the reduction in rate of deaths in incident TB HIV negative individuals was (-4.5%, 95% CI: -5.0 to -4.1) from 2006 to 2016 which was greater than the period between 1990 and 2006 (-3.2%, 95% CI, -3.7 to -2.9) (4).

People living with HIV are more likely to develop active TB when infected with TB and are also at a substantially increased risk of adverse outcomes, including death (1). Incidence and mortality from HIV/AIDS in Africa increased rapidly from 1990 to 2006, and began to slowly decline afterwards (5). The decrease was due to the massive scale up of antiretroviral therapy (ART) coverage and the prevention of mother-to-child transmission (PMTCT) (6). Global scale up of ART was the primary factor in reducing AIDS related deaths from the highest peak of 1.9 million (95% CI: 1.7 million to 2.2 million) in 2005 to 1 million (95% CI: 830 000 to 1.2 million) in 2016 (7). The decline in AIDS deaths was highest in eastern and southern Africa, the region most severely affected by the HIV pandemic. The percentage of AIDS related deaths reduced by 62% from the peak in 2004, 1.1 million (95% CI: 950, 000 to 1.2 million) to 420 000 (95% CI: 350, 000 to 510, 000) in 2016.

In 2020, the highest burden of HIV-associated TB was in the African WHO region, which included 23 of the WHO's list of 30 countries with the highest burden of HIV associated TB (1), and the majority of HIV-associated TB cases and death occurred in Africa (75%) (8). Apart from HIV infection, other risk factors for development of TB disease include poverty, undernutrition, smoking and diabetes (9). Most of these risk factors are inter-related. For instance, individuals that are from less affluent families are also likely to have fewer options for nutritious meals and hence might be undernourished (9). In addition, living in congested settings, unplanned housing and slums is also associated with increased risk of acquiring and

developing TB, because of crowding, increased social contact mixing and poor air quality (9). Most African cities have a large population that live under these conditions, due to the rapid pace of urbanisation, and also failure of governments to provide basic services and facilities (9,10). Hence strategies to address the burden of TB are required to be multisectoral to address the various determinants of TB disease.

Accelerated efforts to reduce TB incidence and deaths are required to achieve the 2035 End TB goals, which are defined as a reduction of 90% of TB incidence and 95% of TB deaths in 2035 compared to 2015 and having no TB-affected families facing catastrophic costs due to TB (1,11). To achieve this, continuing improvements in TB diagnosis and care must be made to accelerate progress toward removing barriers to access to care to attain universal health coverage. Additionally, action on determinants that contribute to TB transmission and disease must be taken (1).

The goal of this Chapter is to provide background and motivation for the PhD thesis. Section 1.2 describes the natural history of TB and its implication for public health approaches to diagnosis, care and prevention of TB. Section 1.3 discusses the current public health practices in screening, diagnosis, treatment, and prevention of TB. Section 1.4 discusses the current approaches to case-finding interventions. Section 1.5 describes the current approaches implemented for TB surveillance burden estimation. Section 1.6 outlines the PhD thesis research aims and objectives and concludes the Chapter.

1.2 Tuberculosis: its natural history and implications for public health

Understanding the natural history of TB is important because it allows greater insights into the clinical pathways experienced by people with TB, formalised in the TB cascade of care; it supports rational approaches to improving diagnosis and treatment; informs effective public health interventions for TB control (1).

TB is caused by *Mycobacterium tuberculosis (MTB),* with a small fraction of clinical disease attributed to *Mycobacterium Bovis* and other mycobacterial species (8). TB primarily causes disease in the lungs (85% of clinical presentations, known as pulmonary TB) followed by sites outside of the lung parenchyma (known as extrapulmonary TB), including the pleura, lymph glands, abdomen, meninges and other sites (1,12).

In immunocompromised individuals (e.g. people living with HIV, people at the extremes of age, and people taking immunosuppressive medication) TB can spread to multiple organs; this is known as disseminated TB (8). Symptoms of TB depend on which part of the body is affected. Common symptoms of TB include fever, fatigue, night sweats, unexplained weight loss, and in pulmonary TB, chronic cough and haemoptysis (coughing sputum mixed with blood) (1,8). Without treatment, about 50% of HIV negative individuals with active TB disease die and nearly all HIV-positive individuals die, with often a fulminant disease course (1,8).

TB is spread when an infected individual expels droplets with *MTB* into the air, for example by coughing, sneezing, shouting, or singing (8). Depending on the environment, these droplets can remain in the air for several hours (13). *MTB* is transmitted when the inhaled droplets containing *MTB* reach the alveolar air sacs of the lungs. In the early stages of the infection, the *MTB* is contained in the lungs. Later however, the bacteria can spread through

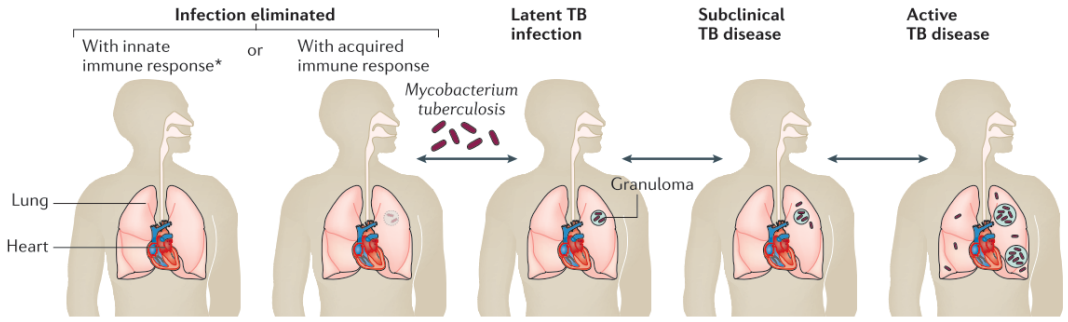
the lymphatic system or the bloodstream to other organs (8). Within 2 to 8 weeks, lung macrophages ingest and surround the bacilli to form granulomas (Figure 01) (1,8). These cells form a barrier shell which keeps the bacilli encapsulated (Often known as the Ghon focus). At this point, individuals will have no symptoms of TB, but will frequently demonstrate immunoreactivity with evidence of exposure to TB, known as latent TB infection (LTBI). In the granuloma, bacilli reproduce slowly, avoiding the host cell's immune system defences. If replication is uncontrolled however, eventually TB bacilli will break out from the granuloma, with bacilli entering the small (and occasionally large) airways; at this stage the individual becomes infectious and symptomatic (active TB disease) (8).

When individuals have active TB disease, the immune system's efforts to control the multiplication of bacilli in the granulomas causes the human host to destroy its own lung tissue (8). If the replication is not stopped, further replication leads to the creation of more cavities, and the lung parenchyma is destroyed progressively in a process that might take months or years. This lung damage can often be seen on a chest X-ray, and this is a good indication that someone has active pulmonary TB. Approximately 10% of people infected with TB progress to active TB disease over their lifetime (8). HIV positive individuals are 26-times more likely to develop active TB disease compared to HIV negative individuals (8).

Progression to active disease depends on the immune response of the host and other reasons that have not been well characterised (8). Depending on the immune response a spectrum of infection stages can arise; latent or subclinical states, ranging from the early containment of infection by an innate or an acquired immune response, through the containment of live but non-replicating bacilli and bacterial replication at low levels partially contained by immunity (Figure 01) (8).

Exposure to, and infection with TB depends on a wide range of factors some of which are environmental, others depend on the TB pathogen, and some upon social dynamics. Environmental and social factors that are important in determining TB exposure and infection include humidity, sunlight, temperature, ventilation, air quality, poverty and crowding (14). For the human host, HIV status, age and sex are the most important risk factors for development of active TB, with others including diabetes, immunosuppressive medications, alcohol and tobacco use (14). The TB disease cycle offers three points for possible interventions for avoiding active TB disease: before infection, during the course of an established infection (during latency) and after progression to active TB disease (8) (Figure 01).

Figure 01: schematic for the natural history of TB



TST	Negative	Positive	Positive	Positive	Usually positive
IGRA	Negative	Positive	Positive	Positive	Usually positive
Culture	Negative	Negative	Negative	Intermittently positive	Positive
Sputum smear	Negative	Negative	Negative	Usually negative	Positive or negative
Infectious	No	No	No	Sporadically	Yes
Symptoms	None	None	None	Mild or none	Mild to severe
Preferred treatment	None	None	Preventive therapy	Multidrug therapy	Multidrug therapy

Source: Figure extracted from Pai et al., Nature (8)

1.2.1 Implications of TB history on public health

People can transmit TB when they have active TB disease, subclinical TB, or asymptomatic TB (Figure 01) (8,15). Historically it was assumed that only symptomatic TB cases were able to transmit TB, but this has changed with a better understanding of the TB disease spectrum (Figure 01) (8,15). It is now recognised that people with subclinical stages of TB can also transmit infection (16,17). Therefore, interventions that are intended to reduce and eliminate the spread of TB need to involve the identification of people with both asymptomatic and symptomatic TB (8,15). In routine TB care, individuals self-present at a health facility to request TB investigation when they experience symptoms. Since, by definition, asymptomatic TB cases do not have symptoms, they do not go to health facilities to seek care (18). This has important implications for the ongoing transmission of TB at the population level, since asymptomatic TB cases are more likely to be missed by routine TB care system (19,20). Hence both symptomatic and symptomatic TB cases need to be

targeted for diagnosis and treatment to reduce TB transmission at the population level (8,15,16).

1.3 Current public health practice, screening, diagnosis, treatment, and prevention

1.3.1 Diagnosis and screening

The choice of diagnostic tool for TB depends on the stage of the TB infection and the characteristics of the individuals to be tested. Two main tests are available for latent TB infection, the tuberculin skin test (TST) and the interferon-gamma release assay (IGRA) (21), which both measure immunological response to previous exposure to TB. Newer tests for latent TB infection (c-TB, Diaskintest) are becoming available, but are not in wide use currently (22).

Testing for latent TB infection and treatment is only recommend for groups that are at an increased risk of progression to active TB disease, for instance people living with HIV, household contacts of TB patients and other at risk HIV-negative groups such as patients who are receiving dialysis or solid organ transplant (21). Treatment of latent TB is only likely to be beneficial for a small percentage of people, predominately those with evidence of recent infection, those in high-risk groups, and those with documented evidence of exposure or latent infection (21). Before initiation of LTBI diagnosis or treatment, active TB disease must be ruled out to avoid treatment of active disease with inappropriate single anti-TB agents. This is commonly affirmed through the absence of any TB symptoms on a WHO-recommended TB symptom screen, and occasionally by the absence of abnormal chest radiograph findings. Where suspicion of active TB remains, a microbiological confirmatory test of TB is carried out, as discussed further in the paragraphs below (21).

By contrast, detection of active TB disease utilises two main approaches: screening, and diagnostic testing. Two main methods are available for either screening or triaging individuals (19,23). Screening tests are done in people who would not have sought care under usual circumstances, and who don't recognise themselves to be unwell (19). Triaging tests (a form of intensified screening to support clinical decision making) are administered to people with TB symptoms and/or high-risk factors for TB. Both screening and triage tests aid in determining who should undergo confirmatory testing (23). Until very recently the main approaches for screening or triaging individuals were TB symptom screening and chest X-ray radiography (24). More recently, a cluster-randomised trial has investigated a TB screening approach based on universal sputum testing with a rapid PCR-based molecular platform, and in 2021, WHO additionally made recommendation that testing for the presence of a raised non-specific inflammatory maker (C-reactive protein [CRP]) could be used for triage testing (1,25). To confirm the results of a positive symptom screen, chest X-ray or CRP, a microbiological diagnosis is usually required (19,24), although treatment may be started on the basis of clinical suspicion alone, especially when patients are too unwell to produce a sputum sample. Confirmatory testing for pulmonary TB is based on smear microscopy, molecular tests and/or culture-based methods. The sputum smear test is the most widely used confirmatory test in low-income and middle-income countries despite it having low accuracy, with sensitivity of 32%-94% and specificity of 50%-99% (8). WHO now recommends the Xpert MTB/Rif and TRUENAT molecular platforms as diagnostic tests because they have greater sensitivity than smear microscopy and make lower laboratory demands (24).

Initially, the focus of WHO recommendations for diagnosis and TB treatment was symptomatic sputum smear positive pulmonary TB cases until the year 1999 (26), because

they were thought to be the only group that was infectious and rapid reductions in TB transmission could be achieved by only focusing on this group (8,18,27).

However, focusing only on symptomatic smear positive cases has had suboptimal impact on the epidemiology of TB since people with smear negative TB, asymptomatic and extra-pulmonary TB are ignored (28). This is because approximately 50% of detected active TB cases in TB prevalence surveys do not have TB symptoms, and 50% of culture confirmed TB cases are smear negative TB (15,17,29). This means that people with asymptomatic, smear negative TB and extra-pulmonary TB are less likely to be diagnosed with TB and initiate treatment, resulting in a high risk of morbidity and mortality (8,26). Neglecting these individuals in screening programmes meant that they were at high risk of experiencing prolonged diagnostic delays, during which they would continue to transmit TB in the community, thereby undermining TB control and eradication efforts (8,18). In the End-TB Strategy (2015-2035), there has been a refocusing of priorities to ensure that all people affected by TB – not just individuals with sputum smear positive active TB- are supported to achieve a rapid diagnosis, treatment, and supportive care (1).

WHO recommended diagnostics for extra-pulmonary TB and disseminated TB include: the Xpert-MTB/Rif and TRUENAT testing of extra-pulmonary samples; lateral flow lipoarabinomannan (LAM) detection in urine; microscopy of extra-pulmonary samples; histopathology; and clinical diagnosis, although these are not always widely available in resource-limited settings (1,24). First generation urine LAM testing is recommended for HIV-positive patients who are seriously ill with a high probability of disseminated TB disease (CD4 cell count of less than or equal to 100/mm^3) (30).

1.3.1.1 Limitations of TB diagnosis

Both currently-recommended LTBI tests (tuberculin skin test [TST] and interferon gamma release assay [IGRA] have a low predictive value for identifying individuals who are likely to progress to having active TB and hence benefit from LTBI treatment (8). Smear microscopy has been replaced with a more accurate test Xpert MTB/Rif, but this is relatively expensive, difficult to maintain in poor resource settings, needs to operate under a room temperature of 30 °C or lower and requires uninterrupted power supply (24). This means that many primary health clinics do not have the capacity to install and maintain an Xpert MTB/Rif, hence samples have to be referred to referral TB laboratories. This in turn means that individuals having TB tests at these clinics cannot receive same day results (31), a factor that is important to link patients into care and support treatment initiation (32). Children <15 years are also not well served by the currently-recommended Xpert MTB/Rif sputum test, since it is difficult to get sputum samples from children as they may be unwilling or might have difficulties expectorating (32). However, according to a WHO update in 2020, stool samples can now be utilised for Xpert MTB/Rif tests in children (33–35).

1.3.2 Treatment and prevention

Active pulmonary TB is treated with a WHO-recommended standardised short-course treatment: six-month combination of antibiotics treatment (1). The main intervention for reducing the risk of progression from LTBI infection to active TB disease is a six-month Isoniazid preventative therapy (IPT) (21). In children, the bacilli Calmette- Guérin (BCG) vaccine protects from severe forms of TB (1).

Both treatment and prevention can only be successful when a multifaceted approach to TB control is adopted. Costs incurred during seeking of diagnosis and treatment can cause

access to TB barriers and affect adherence to TB treatment. According to WHO, in 2020 47% of TB patients faced catastrophic costs against a target of 0%; these are expenditures on seeking TB care that account 10% or more of household income (1). People with TB should not lose their jobs because they have been diagnosed with TB; instead, they should be supported to return to work once their treatment is completed and they have healed sufficiently to work (36). Poverty is a risk factor of TB, and countries where economic conditions have improved have also experienced reduced incidence of TB (9), hence WHO has put strong focus on sustainable development goals (SDG) that are associated with TB (1). In essence, TB medical interventions must be combined with improvements in the social and economic determinants of TB (9).

1.3.2.1 TB treatment outcomes

Individuals diagnosed with active TB in the community via ACF or in health facilities are referred to start treatment at their local TB treatment clinic, where they are also entered into the national TB register for surveillance, and monitoring and evaluation purposes. At the end of TB treatment each patient is assigned an outcome in the register. The categories of TB treatment outcomes are outlined in Table 01 below; they are defined according to WHO guidelines on reporting of TB outcomes and they are mutually exclusive except one, treatment success (37).

Table 01: TB treatment outcomes

Outcome	Definition
Cured	A pulmonary TB patient with bacteriologically confirmed TB at the beginning of treatment who was smear- or culture-negative in the last month of treatment and on at least one previous occasion.
Treatment completed	A TB patient who completed treatment without evidence of failure but with no record to show that sputum smear or culture results in the last month of treatment and on at least one previous occasion were negative, either because tests were not done or because results are unavailable.
Treatment failed	A TB patient whose sputum smear or culture is positive at month 5 or later during treatment.
Died	A TB patient who dies for any reason before starting or during the course of treatment.
Lost to follow-up	Lost to follow-up A TB patient who did not start treatment or whose treatment was interrupted for 2 consecutive months or more.
Not evaluated	A TB patient for whom no treatment outcome is assigned. This includes cases “transferred out” to another treatment unit as well as cases for whom the treatment outcome is unknown to the reporting unit.
Treatment success	The sum of cured and treatment completed.

Source: Table extracted from WHO guideline (37)

In 2019, the overall global treatment success rate for people treated for TB with first-line drugs was 86% against a target of 90%. But the treatment success rates were lower in people living with HIV (77% globally in 2019) (1).

1.4 Current approaches to case-finding interventions

The discovery of TB treatment drugs marked a change in the approach to public health interventions of TB control (18). Before TB transmission control was done by detection and isolation of symptomatic TB cases into special care homes known as sanatoria (18). The discovery of TB treatment allowed TB patients to be safely treated in hospital, or at home with monitoring by health personnel or a designated guardian under the directly observed therapy (18). This allowed WHO and countries to set ambitious targets for detection and treatment of individuals with TB (19,38). Approaches for detecting TB cases are discussed below.

1.4.1 Passive case finding

Passive case finding (PCF) is where TB symptomatic individuals in the community decide to go and seek care at a health facility for their symptoms, on their own or after being prompted by family or friends (39). Once patients are diagnosed with TB either through a confirmed microbiological test or through a clinical decision (where a clinician decides based on patients' symptoms regardless of a negative microbiological TB test), they are started on TB treatment (1). Even when TB symptoms are recognised, approximately 42 percent (31) of people delay reporting to healthcare facilities to be investigated for TB; the estimated average delay is approximately 81 days (40) from the time an individual notices the first symptoms of TB to the time they make their first visit to a health facility that provides TB care. This might be due to a number of factors that include the social economic status of the patient's household, since individuals from poorer households might not have enough resources to get access to health facilities even when the services offered are free (19).

In Malawi and other African countries, qualitative work has also shown that delays in accessing TB care might also be driven by harmful gender-based norms, where men have a culture of only accessing help when the situation is severe because early care seeking is perceived as a sign of weakness (41). Perceptions of poor service delivery may also discourage individuals from self-presenting to health facilities in a timely manner (42). Such perceptions might push individuals to seek care from other care providers that are outside the formal health facilities such as traditional healers, or self-medicating with off the-shelf medicine bought from local grocery stores (42).

Since self-presentation to health facilities starts with an individual having symptoms, this means that asymptomatic active TB cases cannot self-present to health facilities for TB investigations (40). Individuals with undiagnosed active TB disease in the community continue to transmit infection hence putting the community at risk and also increasing their own risk of morbidity and mortality, since delays in seeking health care are associated with poorer outcomes (43). Hence reliance on PCF alone, even when this is followed by prompt diagnosis and initiation of treatment, is not sufficient to meaningfully impact TB transmission and reduce incidence, prevalence, and mortality (44).

1.4.2 Active case finding

Active case finding (ACF) campaigns were first carried out in the industrialized countries between 1930s and 1960s, prompted by high TB mortality rates (18). These were screening campaigns that would target whole populations, and were credited for rapidly reducing TB mortality rates (18). WHO, over time has been updating its guidelines on ACF to tackle the current state of the TB epidemic based on evidence from research studies (19,38). Unlike PCF, ACF is health service provider initiated. ACF is the systematic investigation of

individuals for active TB in pre-specified groups of the population. It uses investigation procedures that can be rapidly applied (19). When ACF is done within a health facility, it is referred to as intensified case finding, in contrast to PCF, health service providers request health facility attendees to be investigated for TB, if the attendees have TB symptoms or have risk factors for TB, even when their reason for that clinic visit was for another condition. ACF can also be done in congregate settings such as prisons, refugee camps and in other priority communities (45,46). The screening procedure should be sensitive enough to identify individuals who are likely to have active TB and exclude the rest, while confirmatory testing should have high specificity to avoid false positives (19,45).

The primary objective of ACF is to ensure that active TB is detected early to reduce the risk of transmission, poor disease outcomes and the adverse social and economic consequences of the disease (17,19,45). ACF typically targets people in the community who do not seek care because either they do not have symptoms or do not recognise the symptoms that they have or they face barriers to accessing facility-based health services (19).

1.4. 2.1 Community-based active case-finding

Interventions that rely on individuals attending a health facility – PCF or facility based ACF approaches – fail when individuals do not present at facilities (45). Unlike PCF and health facility-based ACF, the goal of community ACF is to identify and diagnose individuals with active TB in the community (19,45). Community ACF diagnoses individuals with TB at an earlier stage of their disease compared to PCF, therefore potentially reducing the length of time which an individual is infectious to others in the community (17,19). Therefore, a successful ACF intervention is expected to reduce TB disease prevalence and incidence of TB

in the community (19). Below I summarise the evidence for the effectiveness of community ACF interventions (45).

A recent systematic review on the effectiveness of community ACF that informed the 2021 WHO TB Screening Guidelines identified 28 studies, and found that when community interventions are done with a high coverage and with repeated rounds, they can reduce the prevalence of undiagnosed TB cases in the community (45). The studies covered a range of target populations: five in the general population, seven in high-density urban areas with high poverty rates, four rural populations, two camps of internally displaced people, four in incarcerated populations, one in mines, and two in homeless people (45). The studies implemented different types of ACF interventions: door-to-door screening, sputum collection by community health workers or volunteers, community engagement, and mobile TB screening clinics (45).

In the above systematic review, three randomised trials provided varying evidence in support of community ACF. First, the ZAMSTAR study (44) was a cluster-randomised trial in 24 communities in Zambia and South Africa in a general population with a high TB prevalence. The ACF intervention was a combination of community mobilisation, improving awareness of TB symptoms in schools and setting up sputum collection points in the health-care facilities and the community. TB diagnosis in the intervention arm was based on smear microscopy, and the study's goal was to investigate if there was a difference in the prevalence of active TB cases between the intervention and control arms 4 years after the start of the intervention. In a post-intervention prevalence survey, the intervention was evaluated by comparing the prevalence of tuberculosis between the two arms. Overall, 1277 per 100,000 people had culture-positive TB, compared to 1485 per 100,000 in the ACF

arm. The adjusted TB prevalence ratio comparing the control arm versus the intervention arm was 1.09 (95% CI: 0.89-1.40). There was no difference in the prevalence of active case TB between the intervention and control groups.

In the ACT3 study in Vietnam (25), all people aged 15 years and older, regardless of symptoms, were screened annually in their homes using sputum Xpert MTB/Rif tests for 3 years, in 120 communities in the general population. A further 120 communities were randomised to the control arm. The effect of the intervention was assessed by a prevalence survey in the fourth year since the start of the intervention, where a prevalent TB case was defined as an individual that had at least one sputum sample with a confirmed TB diagnosis using Xpert test. The active TB prevalence in the control arm was 226 per 100000 and 126 per 100000 in the intervention arm. The adjusted prevalence ratio between the control arm versus the intervention arm was 0.56 (95% CI: 0.40-0.78). In other words, the intervention arm had a statistically significant 44% lower prevalence of active tuberculosis than the control arm.

Third, the DETECTB study in Zimbabwe (47) was a cluster randomised trial in the urban general population in Harare. The outcome was the prevalence of culture positive TB among a random sample of 12% of households in each of 46 clusters compared before and after six rounds of the intervention. The 23 clusters were allocated to mobile van ACF and the other 23 to door-to-door screening, adults (16 years) with a cough of more than two weeks were asked to provide two sputum samples for smear microscopy test. The adjusted risk ratio for active TB disease after ACF versus before ACF was 0.59 (95% CI: 0.40-0.89). In other words, there was about a 40% reduction in the prevalence of culture positive TB in the intervention arm compared to the control arm.

Based on the evidence of the systematic review, WHO has updated its TB community systematic screening guidelines (19). According to the latest WHO community ACF guidelines, indiscriminate community ACF is only recommended in areas with a prevalence of undetected adult pulmonary TB of 0.5% or more (19). Community ACF is also recommended among people with structural risk factors for TB (19). These include populations that are urban poor, homeless, live in remote or isolated areas, indigenous communities, migrants, refugees, internally displaced persons, and marginalised groups with limited access to health care (19). The other groups that are recommended for systematic screening are people living with HIV, close contacts of individuals with TB disease, incarcerated populations, miners, and other people exposed to silica dust (19).

1.4.3 Shortcomings of current community case-finding recommendations

ACF programmes can be extremely resource-intensive and expensive (18). TB cases are usually concentrated in subgroups of the population (48,49). Because of this, indiscriminate generalised community ACF is only recommended for communities with 0.5% prevalence and in marginalised communities with structural risk factors for TB (19). It is also important to realise that even in these recommended communities, TB incidence might not be uniformly distributed (50). Since most African cities have a lot of communities that can be categorised as marginalised (for instance, informal communities) prioritisation is important for effective allocation of community ACF interventions and health resources (50). The identification of these communities needs to be carefully conducted in order to identify areas that have a higher burden of TB and also poor access to TB care at health facilities (48,49). Unfortunately, the current WHO recommendations do not provide detailed guidance on how this can be achieved (19).

1.4.3.1 Spatial, hotspot targeted TB community interventions

Spatially targeted interventions are infectious disease community interventions targeted at areas with elevated disease burden or high transmission or risk factors (48,49). Such areas are referred to as hotspots (49). Spatially targeted interventions aim to reach people in neighbourhoods that have higher barriers to access of care at health facilities (49). Therefore, given the current recommendations for selection of communities for community ACF, hotspot targeted TB interventions could be another alternative approach for carrying out community ACF (49,51).

TB cases are known to cluster into hotspots geographically (52). These TB hotspots are usually characterised by poverty, informal urban settlements, displaced people, a higher proportion of male residents (as men are known to have a greater risk of TB infection and disease compared to women) (53), and poor access to health services (54–56). To identify true TB hotspots for effective targeting of community TB interventions, it is important that the spatial heterogeneity of TB and its drivers are identified. These factors are likely to vary by area, so it is important that the analyses are contextualised to each setting (50).

TB hotspots are usually based on geolocation of TB case individuals' residences that are aggregated at the area level (57). The area level aggregation might be district level or neighbourhood level depending on the resolution of the available data. Common spatial clustering methods that are used to identify the spatial hotspots include Kulldorff's spatial scan statistic, the local Moran test, and Getis and Ord's local Gi statistic (48). Care has to be taken when defining hotspots that are based on notified cases, due to the possibility that the identified hotspots can be areas that have better access to TB diagnosis and notification and not hotspots of undiagnosed TB (48).

Also, the usefulness of the identified disease hotspot patterns depends on the spatial scale chosen. If the area aggregate units are too big, they do not give information that can be used for targeted interventions at neighbourhood level (48,49). It is also important that the area units that are used are those that are related to the disease that is being investigated (51). For instance, community health worker (CHW) catchment areas would be preferable to area units such as electoral areas, since health programmes are already familiar working with CHW catchment areas (58).

The time window of analysis is also an important dimension that can influence the spatial pattern of an infectious disease (48). The time period to consider depends on the specific disease; for instance, those that have a long incubation period could be analysed at a wide window period compared to acute infectious diseases (48). The risk of progression to active TB disease after being infected with TB is highest in the first 2 years after infection (8,51). Hence typically, TB data are analysed at 1 year or greater periods of time (8,51).

Since TB is a relatively rare disease even in high-burden settings, analysis of TB data commonly faces challenges of sparsity, where some areas might have fewer cases or no cases (59). Statistical smoothing techniques can mitigate the problems of stochastically unstable rates from areas with small populations, or low prevalence, by smoothly incorporating information from nearby locations (59). This enables predictions to be made in areas with sparse data (59).

1.5 TB surveillance and burden estimation

TB incidence is the number of new TB cases arising each year in a given population (1). Monitoring trends in TB incidence allows WHO and TB programmes to assess progress towards meeting the End TB goals and appropriately respond to the pandemic (1).

Unfortunately, TB incidence cannot be directly estimated because the design of such a survey would be impossible in most populations (1). Instead, TB burden is estimated using TB prevalence and TB notifications (1), and occasionally TB incidence is modelled from these data.

1.5.1 Prevalence of pulmonary TB surveys

The prevalence of pulmonary TB disease is the number of pulmonary TB cases that exist in the population at a given point in time (60). It is reported as the number of TB cases per given population — usually per 100,000 population or as a percentage of the population (60). Cross sectional survey designs are used for TB prevalence surveys (60). The prevalence of TB is only measured among adults as children <15 years generally cannot produce sputum samples; prevalence surveys also ignore extrapulmonary TB as it is less common than pulmonary TB and it can also be difficult to diagnose under survey conditions, as the symptoms may be less specific and the disease may not be detectable through chest X-ray or sputum sample. According to WHO prevalence survey guidelines, survey participants can be screened using either or both a symptom enquiry and a chest X-ray (60). Sputum samples are then taken from all those with abnormal chest X-ray and/or symptoms suggestive of pulmonary TB (60). Each eligible study participant normally has two sputum samples taken, one immediately and one later in the day. However, both sputum samples may be provided at the same time in some cases (29). The sputum samples are tested in laboratories to identify which individuals have bacteriologically-positive pulmonary TB using Xpert MTB/Rif and or culture.

TB prevalence surveys help to inform estimates of incidence, and TB burden trends, when repeat surveys are available (60). The insights from prevalence surveys inform policy choices

on action to reduce the TB disease burden. WHO prioritises national prevalence surveys for countries that have an estimated TB burden of about 150 incident TB cases per 100 000 population per year or more, and do not have high coverage of routine notification systems and cause of death registries to enable direct estimates of TB incidence and TB death (29). A total of 33 national prevalence surveys have been done in 30 countries between 2007 and 2020 with WHO support (29). This included 16 countries in Asia and 17 in Africa.Repeat surveys during this period were only done in three countries Myanmar, Philippines and Vietnam (29,33). Malawi last had a national prevalence survey in 2013: the prevalence of sputum smear positive TB was 220 (95% CI: 142 – 297) and the prevalence of sputum smear positive or culture positive TB was 452 (95% CI: 312 – 593) per 100, 000 adult population.

TB prevalence surveys give information that helps to understand the TB epidemic, in addition to providing estimates of the disease burden (60). Surveys can assist in understanding why those who have been diagnosed with TB were not diagnosed before the survey in countries where access to diagnosis is difficult. They also detail the extent to which patients diagnosed with TB in the private sector are not reported to the national TB programmes (NTP) (60). This information is used to identify strategies that can increase the proportion of cases that are diagnosed and reported to the national routine TB surveillance programme. The data can also be used to describe the spatial distribution of the burden of disease in the community, although at a basic level of rural/urban strata (48,49).

1.5.2 Notifications

TB notifications refer to the number of TB cases that have been detected and reported to the national surveillance system (1). When compared to mortality, prevalence, and incidence data, TB notifications offer a measure of TB burden that is readily available to

national TB programmes (49). In most countries there is usually a gap between the number of people that are notified with TB and the actual number that are estimated to have developed incident active TB (1). This gap is due to a combination of underreporting of people diagnosed with TB (individuals diagnosed with TB but not reported to the national TB programme) and underdiagnosis (i.e. when people with TB cannot access health care or are not diagnosed when they do) (1).

WHO estimates incidence based on TB notifications only for countries where there is better quantification of under-reporting and over-reporting and universal access to quality health-care. Only 147 countries fulfilled this criterion in 2020, representing about 23% of the estimated global number of incident TB cases in 2020 (1). Alternatively, TB prevalence surveys were used to estimate TB incidence in 29 countries, these countries accounted for 66% of the global TB incidence in 2020 (61). Finally, case notification combined with expert opinion about case detection gaps was used in 39 countries, which accounted for 11 % of the estimated global number of incident cases (61).

The COVID-19 epidemic that started in 2019, affected TB notifications in most countries; globally, there was an 18% drop in notifications of people diagnosed with TB in 2020 compared to 2019 (1). In addition, in 2019 and 2020 the gap between notified TB cases and the number of people estimated to have developed TB grew wider (1,2). The number of notified TB cases increased between 2009 to 2019 from 5.7 to 7.1 million. But due to the COVID-19 pandemic, the number of notified TB cases in 2020 fell to 5.8 million, a number that is similar to the number of notified TB cases that were diagnosed and notified in 2012 (1).

1.5.3 Data collection practices

In most countries, notified TB cases from health facilities are reported at the district/province health offices where they are aggregated and sent to the National TB programme (1). NTP's aggregate all the data from all the districts across the country and send the data to WHO (1).

For example, in Malawi, the routine data collected when notifying TB patients are age, sex, physical address, duration of cough, previous treatment, HIV test results, antiretroviral status, smear microscopy result, TB classification (pulmonary or extra-pulmonary) and TB treatment outcome (62). Prevalence surveillance studies provide an opportunity for collecting data that is more representative of the study population (29). This data includes social demographic data, TB risk factors and health care-seeking behaviour data (29).

In Africa, collection of TB patients' or prevalence survey participants' global positioning satellites (GPS) coordinates could provide additional epidemiological insights that enable identification of TB hotspots at granular area level (63). The majority of people that are disproportionately affected by the disease live in informal urban settlements that do not have postal codes or zip codes (56), meaning that it is currently challenging to spatially resolve the TB case's location of residence (64).

The data that were used for this PhD thesis came from Blantyre, Malawi. Blantyre provided a good study setting because it has had an ongoing enhanced TB notification system since 2015 (65,66). This enhanced system was set up to improve data collection practices among NTP staff based at health facilities as well as provide TB microbiologically confirmed lab tests (65,66). This has allowed collection of high-quality TB notification data: apart from collecting the routine data, other data such as GPS coordinates of TB patients' households are also

collected (65,66). The other source of data is going to come from a TB prevalence survey that was carried out in urban Blantyre in 2019.

So far in this introduction, we have argued that TB is a major ongoing public health challenge globally, in Africa and in Malawi (1). Passive Case Finding (PCF) – waiting for sick individuals to self-present at health facilities for TB investigations – cannot eradicate the TB epidemic (15,19). Even under optimal circumstances, PCF would miss asymptomatic cases which make up about 50% of all prevalent cases (15). One potential solution is to complement PCF with community-based ACF, but the current recommendations for ACF are unlikely to be cost-effective for resource constrained settings like Africa and Malawi (18,19). Exploiting the risk factors that lead to the concentration of TB cases into hotspots could aid in the identification of most at need communities that could be targeted with ACF (51). The aims of my PhD in the section below, are towards addressing these gaps in WHO community ACF recommendations.

1.6 Research aim:

The **overall aim of the thesis** is to investigate how spatial heterogeneity in TB epidemiology can be used to improve the delivery of community-based case-finding interventions for TB.

Hypothesis: Can we find epidemiologically important spatial variability in TB epidemiology in urban Malawi, which may be used to guide better-targeted and perhaps more successful TB care and prevention activities, based on recent research and TB notification and prevalence data?

Objectives:

1. To systematically review the available literature on the effectiveness of spatially targeted interventions for HIV, TB, leprosy, and malaria. Note that, since spatially targeted interventions of TB are still relatively new, I included the other infectious diseases to broaden the scope of our understanding of spatially targeted interventions.
2. To identify spatial and non-spatial risk factors for case fatality rates of notified TB cases in urban Blantyre Malawi.
3. To estimate neighbourhood-level TB case prevalence to notification rate ratios (P:N ratios) using data from systematic surveillance of notified TB cases and a citywide TB prevalence survey form Urban Blantyre, Malawi. To help identify neighbourhoods that have a higher burden of undiagnosed TB that may be being missed by the routine TB notification surveillance system.
4. To discuss the findings from objectives 1 to 3 to inform the design of effective community ACF interventions.

1.7 Structure of the thesis

The plan for this thesis is as follows: Chapter 2 introduces the study setting and describes the data sources for this PhD. In addition, I will describe the preliminary modelling that was carried out at the start of the project to investigate the feasibility of the proposed approach. Chapter 3 presents a published systematic review of the effectiveness of spatially targeted interventions of HIV, TB, leprosy and malaria, whose findings provide formal justification for our approach. Chapter 4 presents published TB case fatality analysis for the Blantyre urban area, 2015-19. Building on this, Chapter 5 presents published results of modelling of neighbourhood prevalence to notification ratios for identification of TB hotspots. We conclude with a comprehensive discussion of the findings and their implications in Chapter 6.

1.8 References

1. World Health Organization (WHO). Global tuberculosis report 2021 [Internet]. Geneva; 2021. Available from: https://www.who.int/publications/i/item/9789240037021

2. Soko RN, Burke RM, Feasey HRA, Sibande W, Nliwasa M, Henrion MYR, et al. Effects of Coronavirus Disease Pandemic on Tuberculosis Notifications, Malawi. Emerg Infect Dis [Internet]. 2021 Jul [cited 2021 Jun 23];27(7):1831–9. Available from: https://wwwnc.cdc.gov/eid/article/27/7/21-0557_article.htm

3. World Health Oganisation. Update Le point Tuberculosis: a global overview of the situation today 1992. Geneva; 1992.

4. Hmwe Kyu H, Maddison ER, Henry NJ, Ledesma JR, Wiens KE, Reiner Jr R, et al. Global, regional, and national burden of tuberculosis, 1990–2016: results from the Global Burden of Diseases, Injuries, and Risk Factors 2016 Study. Lancet Infect Dis [Internet]. 2018 [cited 2022 Mar 1];18:1329–49. Available from: www.thelancet.com/infection

5. Ortblad KF, Lozano R, Murray CJL. The burden of HIV: Insights from the global burden of disease study 2010. AIDS [Internet]. 2013 Aug 24 [cited 2022 Mar 2];27(13):2003–17. Available from: https://journals.lww.com/aidsonline/Fulltext/2013/08240/The_burden_of_HIV__insights_from_the_Global_Burden.1.aspx

6. Group IMH working. Prevention of mother-to-child transmission of HIV Internal document 2020 edition. 2020.

7. Joint United Nations Programme on HIV/AIDS (UNAIDS). Progress towards the 90-90-90 targets Ending AIDS GLOBAL AIDS UPDATE | 2017 [Internet]. 2017 [cited 2022 Mar 2]. Available from: https://www.unaids.org/sites/default/files/media_asset/Global_AIDS_update_2017_en.pdf

8. Pai M, Behr MA, Dowdy D, Dheda K, Divangahi M, Boehme CC, et al. Tuberculosis. Nat Rev Dis Prim 2016 21 [Internet]. 2016 Oct 27 [cited 2021 Sep 28];2(1):1–23. Available from: https://www.nature.com/articles/nrdp201676

9. Lönnroth K, Jaramillo E, Williams BG, Dye C, Raviglione M. Drivers of tuberculosis epidemics: The role of risk factors and social determinants. Soc Sci Med. 2009;68(12):2240–6.

10. United Nations Human Settlements Programme. SLUM ALMANAC 2015 2016 Tracking Improvement in the Lives of Slum Dweller [Internet]. Nairobi; 2015. Available from: https://unhabitat.org/slum-almanac-2015-2016/

11. Partinership ST. THE PARADIGM SHIFT 2016-2020 Global Plan to End TB. 2015.

12. Peto HM, Pratt RH, Harrington TA, LoBue PA, Armstrong LR. Epidemiology of extrapulmonary tuberculosis in the United States, 1993-2006. Clin Infect Dis [Internet]. 2009 Nov [cited 2021 Nov 23];49(9):1350–7. Available from: https://pubmed.ncbi.nlm.nih.gov/19793000/

13. Wells WF. On air-borne infection: Study II. Droplets and droplet nuclei. Am J Epidemiol [Internet]. 1934 Nov 1 [cited 2022 Feb 28];20(3):611–8. Available from: https://academic.oup.com/aje/article/20/3/611/280025

14. Narasimhan P, Wood J, Macintyre CR, Mathai D. Risk Factors for Tuberculosis. Pulm Med [Internet]. 2013 [cited 2021 Nov 16];2013:11. Available from: /pmc/articles/PMC3583136/

15. Frascella B, Richards AS, Sossen B, Emery JC, Odone A, Law I, et al. Subclinical Tuberculosis

Disease—A Review and Analysis of Prevalence Surveys to Inform Definitions, Burden, Associations, and Screening Methodology. Clin Infect Dis [Internet]. 2021 Aug 2 [cited 2022 May 12];73(3):e830–41. Available from: https://academic.oup.com/cid/article/73/3/e830/5906549

16. Ryckman TS, Dowdy DW, Kendall EA. Infectious and Clinical Tuberculosis Trajectories: Bayesian modeling with case finding implications. medRxiv [Internet]. 2022 Jun 29 [cited 2022 Jul 14];2022.06.27.22276965. Available from: https://www.medrxiv.org/content/10.1101/2022.06.27.22276965v2

17. Ku C-C, MacPherson P, Khundi M, Nzawa Soko RH, Feasey HRA, Nliwasa M, et al. Durations of asymptomatic, symptomatic, and care-seeking phases of tuberculosis disease with a Bayesian analysis of prevalence survey and notification data. BMC Med 2021 191 [Internet]. 2021 Nov 10 [cited 2021 Nov 16];19(1):1–13. Available from: https://bmcmedicine.biomedcentral.com/articles/10.1186/s12916-021-02128-9

18. Golub JE, Mohan CI, Comstock GW, Chaisson RE. Active case finding of tuberculosis: historical perspective and future prospects. Int J Tuberc Lung Dis [Internet]. 2005 Nov [cited 2022 May 12];9(11):1183. Available from: /pmc/articles/PMC4472641/

19. World Health Organisation. Module 2: Screening WHO operational handbook on tuberculosis Systematic screening for tuberculosis disease [Internet]. Geneva; 2021 [cited 2021 Apr 6]. Available from: https://reliefweb.int/report/world/who-consolidated-guidelines-tuberculosis-module-2-screening-systematic-screening

20. Drain PK, Bajema KL, Dowdy D, Dheda K, Naidoo K, Schumacher SG, et al. Incipient and subclinical tuberculosis: A clinical review of early stages and progression of infection. Clin Microbiol Rev. 2018;31(4).

21. World Health Organization. Latent tuberculosis infection Updated and consolidated guidelines for programmatic management. Geneva; 2018.

22. Zellweger J-P, Sotgiu G, Corradi M, Durando P. The diagnosis of latent tuberculosis infection (LTBI): currently available tests, future developments, and perspectives to eliminate tuberculosis (TB). Med Lav [Internet]. 2020 [cited 2022 Jul 14];111:170–83. Available from: www.lamedicinadellavoro.it

23. Nathavitharana RR, Yoon C, Macpherson P, Dowdy DW, Cattamanchi A, Somoskovi A, et al. Guidance for Studies Evaluating the Accuracy of Tuberculosis Triage Tests. J Infect Dis [Internet]. 2019 Oct 8 [cited 2022 May 13];220(Supplement_3):S116–25. Available from: https://academic.oup.com/jid/article/220/Supplement_3/S116/5583863

24. World Health Organization (WHO). WHO consolidated guidelines on tuberculosis rapid diagnostics for tuberculosis detection; Module 3: Diagnosis. [Internet]. Geneva; [cited 2022 Mar 1]. Available from: https://www.who.int/publications/i/item/9789240029415

25. Marks GB, Nguyen N V., Nguyen PTB, Nguyen T-A, Nguyen HB, Tran KH, et al. Community-wide Screening for Tuberculosis in a High-Prevalence Setting. N Engl J Med [Internet]. 2019 Oct 3 [cited 2021 Jun 16];381(14):1347–57. Available from: https://www.nejm.org/doi/full/10.1056/NEJMoa1902129

26. Behr MA, Warren SA, Salamon H, Hopewell PC, Ponce De Leon A, Daley CL, et al. Transmission of Mycobacterium tuberculosis from patients smear-negative for acid-fast bacilli. Lancet (London, England) [Internet]. 1999 Feb 6 [cited 2022 May 19];353(9151):444–9. Available from: https://pubmed.ncbi.nlm.nih.gov/9989714/

27. Dye C, Garnett GP, Sleeman K, Williams BG. Prospects for worldwide tuberculosis control under the WHO DOTS strategy. Lancet. 1998;352(9144):1886–91.

28. Dowdy DW, Basu S, Andrews JR. Is passive diagnosis enough? The impact of subclinical disease on diagnostic strategies for tuberculosis. Am J Respir Crit Care Med [Internet]. 2013 Mar 1 [cited 2019 Mar 7];187(5):543–51. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23262515

29. World Health Organization (WHO). National tuberculosis prevalence surveys 2007-2016 [Internet]. Geneva; 2021 [cited 2022 Mar 1]. Available from: https://www.who.int/publications/i/item/9789240022430

30. Peter JG, Zijenah LS, Chanda D, Clowes P, Lesosky M, Gina P, et al. Effect on mortality of point-of-care, urine-based lipoarabinomannan testing to guide tuberculosis treatment initiation in HIV-positive hospital inpatients: a pragmatic, parallel-group, multicountry, open-label, randomised controlled trial. Lancet [Internet]. 2016 Mar 19 [cited 2022 Mar 1];387(10024):1187–97. Available from: http://www.thelancet.com/article/S0140673615010922/fulltext

31. Getnet F, Demissie M, Assefa N, Mengistie B, Worku A. Delay in diagnosis of pulmonary tuberculosis in low-and middle-income settings: Systematic review and meta-analysis. BMC Pulm Med [Internet]. 2017 Dec 13 [cited 2020 Oct 2];17(1). Available from: https://pubmed.ncbi.nlm.nih.gov/29237451/

32. Treatment Action Group (TAG). An activist's guide to Tuberculosis Diagnostic Tools [Internet]. Treatment Action Group. 2017. Available from: https://www.treatmentactiongroup.org/wp-content/uploads/2017/04/TB-Diagnostics-Guide.pdf

33. World Health Organisation. Global tuberculosis report 2020 [Internet]. Geneva; 2020 [cited 2021 May 31]. Available from: https://www.who.int/publications/i/item/9789240013131

34. Chipinduro M, Mateveke K, Makamure B, Ferrand RA, Gomo E. Stool Xpert® MTB/RIF test for the diagnosis of childhood pulmonary tuberculosis at primary clinics in Zimbabwe. Int J Tuberc Lung Dis [Internet]. 2017 Feb 1 [cited 2022 May 19];21(2):161. Available from: /pmc/articles/PMC5234433/

35. Walters E, Van Der Zalm MM, Palmer M, Bosch C, Demers AM, Draper H, et al. Xpert MTB/RIF on Stool is Useful for the Rapid Diagnosis of Tuberculosis in Young Children with Severe Pulmonary Disease. Pediatr Infect Dis J [Internet]. 2017 Sep 1 [cited 2022 May 19];36(9):837. Available from: /pmc/articles/PMC5558052/

36. Partinership ST. Declaration of the Rights of People Affected by Tuberculosis. 2019.

37. World Health Organization. Definitions and reporting framework for tuberculosis-2013 revision (updated December 2014 and January 2020) [Internet]. Geneva; 2020 [cited 2021 Dec 2]. Available from: https://apps.who.int/iris/bitstream/handle/10665/79199/9789241505345_eng.pdf

38. Lonnroth, Knut. Systematic screening for active tuberculosis Principles and recommendations [Internet]. [cited 2019 May 27]. Available from: https://www.who.int/tb/publications/Final_TB_Screening_guidelines.pdf

39. Keshavjee S, Dowdy D, Swaminathan S. Stopping the body count: a comprehensive approach to move towards zero tuberculosis deaths. Lancet [Internet]. 2015 Dec 5 [cited 2019 Feb 22];386(10010):e46–7. Available from: https://www.sciencedirect.com/science/article/pii/S0140673615003207?via%3Dihub

40. Bello S, Afolabi RF, Ajayi DT, Sharma T, Owoeye DO, Oduyoye O, et al. Empirical evidence of delays in diagnosis and treatment of pulmonary tuberculosis: Systematic review and meta-regression analysis. BMC Public Health [Internet]. 2019 Jun 25 [cited 2022 Mar 1];19(1):1–11. Available from: https://bmcpublichealth.biomedcentral.com/articles/10.1186/s12889-019-7026-4

41. Chikovore J, Pai M, Horton KC, Daftary A, Kumwenda MK, Hart G, et al. Missing men with tuberculosis: the need to address structural influences and implement targeted and multidimensional interventions. BMJ Glob Heal [Internet]. 2020 May 1 [cited 2022 May 10];5(5):e002255. Available from: https://gh.bmj.com/content/5/5/e002255

42. Kumwenda M, Desmond N, Hart G, Choko A, Chipungu GA, Nyirenda D, et al. Treatment-Seeking for Tuberculosis-Suggestive Symptoms: A Reflection on the Role of Human Agency in the Context of Universal Health Coverage in Malawi. PLoS One [Internet]. 2016 Apr 1 [cited 2022 May 10];11(4):e0154103. Available from: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0154103

43. Peters JS, Andrews JR, Hatherill M, Hermans S, Martinez L, Schurr E, et al. Advances in the understanding of Mycobacterium tuberculosis transmission in HIV-endemic settings. Lancet Infect Dis [Internet]. 2018 Dec 13 [cited 2019 Mar 3];19(3):e65–76. Available from: http://www.ncbi.nlm.nih.gov/pubmed/30554995

44. Ayles H, Muyoyeta M, Du Toit E, Schaap A, Floyd S, Simwinga M, et al. Effect of household and community interventions on the burden of tuberculosis in southern Africa: the ZAMSTAR community-randomised trial. Lancet [Internet]. 2013 Oct 5 [cited 2019 Feb 7];382(9899):1183–94. Available from: https://www.sciencedirect.com/science/article/pii/S0140673613611319?via%3Dihub

45. Burke RM, Nliwasa M, Feasey HRA, Chaisson LH, Golub JE, Naufal F, et al. Community-based active case-finding interventions for tuberculosis: a systematic review. Lancet Public Heal [Internet]. 2021;6(5):e283–99. Available from: http://dx.doi.org/10.1016/S2468-2667(21)00033-5

46. Department of HIV/AIDS, Stop TB Department, World Health Organization (WHO). Guidelines for intensified tuberculosis case-finding and isoniazid preventive therapy for people living with HIV in resource-constrained settings [Internet]. Geneva; 2011 [cited 2019 May 24]. Available from: https://apps.who.int/iris/bitstream/handle/10665/44472/9789241500708_eng.pdf;jsessionid=21FF5A6C6174558D73F38D25777B56F0?sequence=1

47. Corbett EL, Bandason T, Duong T, Dauya E, Makamure B, Churchyard GJ, et al. Comparison of two active case-finding strategies for community-based diagnosis of symptomatic smear-positive tuberculosis and control of infectious tuberculosis in Harare, Zimbabwe (DETECTB): a cluster-randomised trial. Lancet [Internet]. 2010 Oct 9 [cited 2019 Feb 7];376(9748):1244–53. Available from: https://www.sciencedirect.com/science/article/pii/S0140673610614250?via%3Dihub

48. Shaweno D, Karmakar M, Alene KA, Ragonnet R, Clements AC, Trauer JM, et al. Methods used in the spatial analysis of tuberculosis epidemiology: a systematic review. BMC Med [Internet]. 2018 [cited 2019 Feb 20];16(1):193. Available from: https://bmcmedicine.biomedcentral.com/articles/10.1186/s12916-018-1178-4

49. Khundi M, Carpenter JR, Nliwasa M, Cohen T, Corbett EL, MacPherson P. Effectiveness of spatially targeted interventions for control of HIV, tuberculosis, leprosy and malaria: a systematic review. BMJ Open [Internet]. 2021 Jul 1 [cited 2021 Jul 20];11(7):e044715.

Available from: https://bmjopen.bmj.com/content/11/7/e044715

50. Dowdy DW, Golub JE, Chaisson RE, Saraceni V. Heterogeneity in tuberculosis transmission and the role of geographic hotspots in propagating epidemics. Proc Natl Acad Sci U S A [Internet]. 2012 Jun 12 [cited 2021 Jun 14];109(24):9557–62. Available from: /pmc/articles/PMC3386125/

51. Shaweno D, Karmakar M, Alene KA, Ragonnet R, Clements AC, Trauer JM, et al. Methods used in the spatial analysis of tuberculosis epidemiology: a systematic review. BMC Med. 2018;16(1):193.

52. Shaweno D, Shaweno T, Trauer JM, Denholm JT, McBryde ES. Heterogeneity of distribution of tuberculosis in Sheka Zone, Ethiopia: drivers and temporal trends. Int J Tuberc Lung Dis [Internet]. 2017;21(1):79–85. Available from: http://www.ingentaconnect.com/content/10.5588/ijtld.16.0325

53. Horton KC, MacPherson P, Houben RMGJ, White RG, Corbett EL. Sex Differences in Tuberculosis Burden and Notifications in Low- and Middle-Income Countries: A Systematic Review and Meta-analysis. Metcalfe JZ, editor. PLOS Med [Internet]. 2016 Sep 6 [cited 2019 May 24];13(9):e1002119. Available from: https://dx.plos.org/10.1371/journal.pmed.1002119

54. Wubuli A, Xue F, Jiang D, Yao X, Upur H, Wushouer Q. Socio-Demographic Predictors and Distribution of Pulmonary Tuberculosis (TB) in Xinjiang, China: A Spatial Analysis. PLoS One [Internet]. 2015 [cited 2019 Feb 5];10(12):e0144010. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26641642

55. Dangisso MH, Datiko DG, Lindtjorn B, Lindtjørn B. Spatio-Temporal Analysis of Smear-Positive Tuberculosis in the Sidama Zone, Southern Ethiopia. PLoS One [Internet]. 2015 [cited 2019 Feb 5];10(6):e0126369. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26030162

56. Corbett EL, Bandason T, Cheung Y-B, Makamure B, Dauya E, Munyati SS, et al. Prevalent infectious tuberculosis in Harare, Zimbabwe: burden, risk factors and implications for control. Int J Tuberc Lung Dis [Internet]. 2009 Oct [cited 2019 Feb 6];13(10):1231–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19793427

57. Glynn JR, Guerra-Assunção JA, Houben RMGJ, Sichali L, Mzembe T, Mwaungulu LK, et al. Whole Genome Sequencing Shows a Low Proportion of Tuberculosis Disease Is Attributable to Known Close Contacts in Rural Malawi. PLoS One [Internet]. 2015 Jul 16 [cited 2021 Dec 15];10(7):e0132840. Available from: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0132840

58. Parenteau MP, Sawada MC. The modifiable areal unit problem (MAUP) in the relationship between exposure to NO 2and respiratory health. Int J Health Geogr [Internet]. 2011 Oct 31 [cited 2022 Feb 23];10(1):1–15. Available from: https://ij-healthgeographics.biomedcentral.com/articles/10.1186/1476-072X-10-58

59. Shaweno D, Trauer JM, Denholm JT, McBryde ES. A novel Bayesian geospatial method for estimating tuberculosis incidence reveals many missed TB cases in Ethiopia. BMC Infect Dis [Internet]. 2017 Dec 2 [cited 2019 Feb 11];17(1):662. Available from: http://www.ncbi.nlm.nih.gov/pubmed/28969585

60. World Health Organization. Tuberculosis PREVALENCE SURVEYS: a handbook [Internet]. Geneva; 2011 [cited 2021 Jun 23]. Available from: www.who.int/tb/advisory_bodies/impact_measurement_taskforce/resources_documents/thelimebook/en/index.html

61. Philippe G, Charalambos S, Carel P, Katherine F. Methods used by WHO to estimate the Global burden of TB disease. 2016;(October). Available from: http://arxiv.org/abs/1603.00278

62. Malawi Ministry of Health. Malawi National Tuberculosis Control Programme 7th edition [Internet]. Lilongwe; 2012 [cited 2019 May 26]. Available from: https://pdf.usaid.gov/pdf_docs/PA00KGDD.pdf

63. Theron G, Jenkins HE, Cobelens F, Abubakar I, Khan AJ, Cohen T, et al. Series How to eliminate tuberculosis 1 Data for action: collection and use of local data to end tuberculosis. Lancet [Internet]. 2015 [cited 2019 Feb 18];386(10010):2324–33. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med11&NEWS=N&AN=26515676

64. Theron G, Jenkins HE, Cobelens F, Abubakar I, Khan AJ, Cohen T, et al. Data for action: collection and use of local data to end tuberculosis. Lancet (London, England) [Internet]. 2015;386(10010):2324–33. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med11&NEWS=N&AN=26515676

65. MacPherson P, Khundi M, Nliwasa M, Choko AT, Phiri VK, Webb EL, et al. Disparities in access to diagnosis and care in Blantyre, Malawi, identified through enhanced tuberculosis surveillance and spatial analysis. BMC Med. 2019;17(1):21.

66. Khundi M, MacPherson P, Feasey HRA, Soko RN, Nliwasa M, Corbett EL, et al. Clinical, health systems and neighbourhood determinants of tuberculosis case fatality in urban Blantyre, Malawi: a multilevel epidemiological analysis of enhanced surveillance data. Epidemiol Infect [Internet]. 2021 [cited 2021 Nov 19];149. Available from: https://www.cambridge.org/core/journals/epidemiology-and-infection/article/clinical-health-systems-and-neighbourhood-determinants-of-tuberculosis-case-fatality-in-urban-blantyre-malawi-a-multilevel-epidemiological-analysis-of-enhanced-surveillance-data/F1

Chapter 2

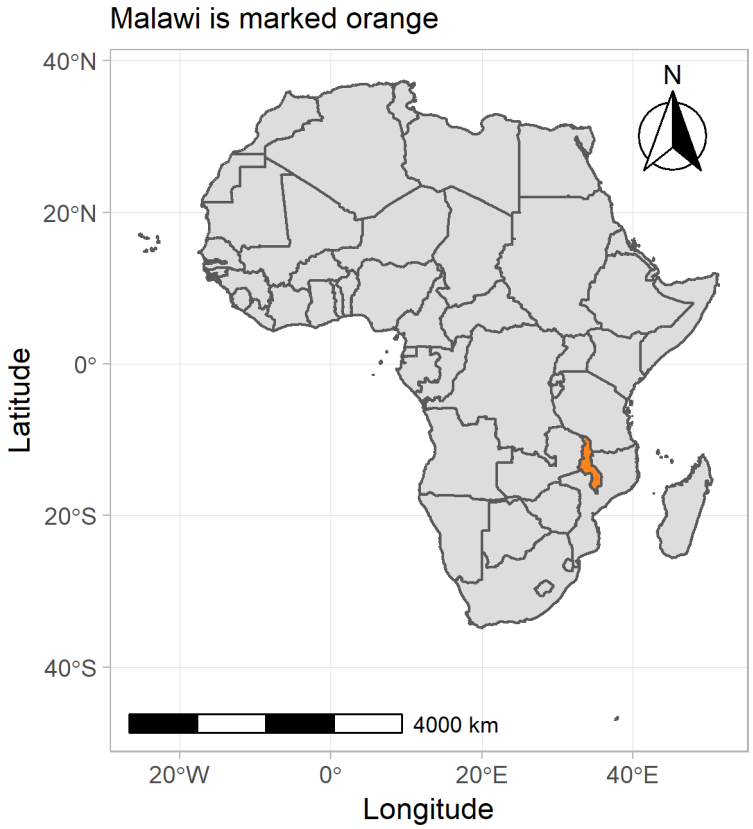
2.1 Introduction

This chapter describes the study area in Section 2.2, and Section 2.3 gives a brief description of the sources of data that were used for this PhD. Section 2.4 describes data simulation and modelling work that was done to demonstrate the feasibility of the proposed approach for achieving objective three of the thesis – specifically the feasibility of using the 2015-19 notification data and the 2019 prevalence survey to estimate TB Prevalence to notification ratios for the 72 neighbourhoods across urban Blantyre, Malawi. The simulation and modelling work was part of my PhD proposal.

2.2 Description of the study area

Malawi is in the Sub-Saharan African region (Figure 2.1). It shares borders with Mozambique on the east, south and southwest, Zambia to the northwest and Tanzania to the northeast (1). It has three climatic seasons: a warm wet season (November to April), a cool dry winter season (May to August) and a hot dry season (September to October) (2). It is classified among the least developed countries in the world by the United Nations (3). Malawi's gross domestic product per capita was estimated to be 424 USD in 2020, and 68% of the population was estimated to live below the poverty line in 2019 (less than $1.90 per day) (4,5).

Figure 2.1: Map of Africa

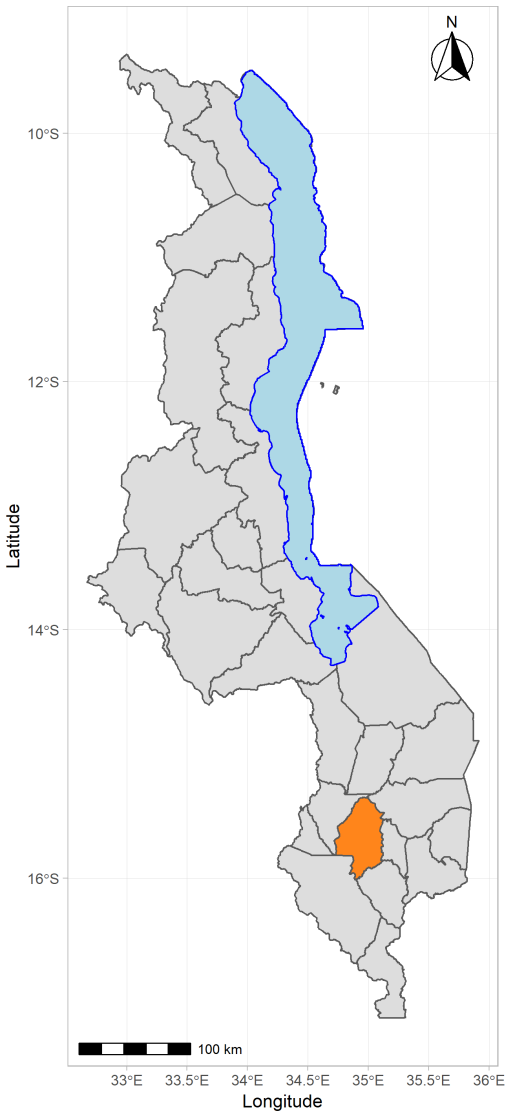


Shapefiles for the map of Africa were obtained from the afrilearndata R package, code to produce the map can be accessed at https://github.com/mcewenkhundi/phd_exploratory_chapter.

The study site was urban Blantyre, Malawi (Figure 2.2). Blantyre is the second largest district in Malawi after the capital city Lilongwe, and the country's commercial centre (6). Blantyre is located in the southern region of Malawi, and according to the 2018 national census, urban Blantyre had a population of 800,264 and a mean household size of 4 (6). The city has a high population growth rate of 2% per annum according to the 2018 population census (6). A national TB prevalence survey that was done in 2014 estimated that the city had an estimated TB prevalence of about 1% (7) and an HIV prevalence of 18% in 2018 (8). Malawi

as a country is categorised as an area of high HIV and TB prevalence burden by the World Health Organization (9).

Figure 2.2: Map of administrative districts of Malawi, Blantyre district marked in orange



The shapefiles for the Malawi admirative regions were obtained from Malawi National Statistical Office, code to produce map can be accessed at https://github.com/mcewenkhundi/phd_exploratory_chapter

2.3 Description of the sources of data for the PhD

In this section, I describe the sources of data for this PhD, which are summarised in Table 2.1.

Table 2.1: Data sources summary table

Data name	Census data	TB Notifications	TB treatment outcomes	Prevalence survey data
Description of study	A study area enumeration exercise collected household-level data for all the neighbourhoods in the study area.	Data is collected on all notified TB cases. In addition to the national TB registration data, other study-specific data are also collected.	TB treatment outcomes of notified TB cases are collected from all the TB registers after the end of the six-month TB treatment.	Individuals with a cough or abnormal chest X-ray or both were asked to submit sputum for smear and Xpert microscopy and MIGT culture.
Time period of data collection	10 Oct 2015 and 30 Dec 2015.	From 1 Jan 2015 to 30 Dec2019.	From 1 Jan 2015 to 30 Dec 2019.	From May 2019 to October 2019.
Collection method for GPS data	Using a GPS device, a circumferential walk was done to mark the boundary of each neighbourhood.	Using ePAL, a software application that has inbuilt maps. TB officers were assisted by patients to capture household coordinates	All TB Treatment outcome data are linked to TB notifications data	GPS devices were used to get household GPS coordinates
Total records	The total population from the census was 753,489 individuals	10,306	10,306	15000 adults

2.3.1 Study census

In 2015, the MLW study team conducted a population census in all of the city's community health worker (CHW) areas (10). Each CHW area has a ministry of health employee that is linked to its closest primary health clinic. These CHW officers are responsible for providing community engagement activities to the community members in the CHW area on behalf of the health facility that they report to. Our study team data officers have previously collected the GPS boundaries of these CHW areas, this was accomplished by doing a circumferential

walk around the CHW areas accompanied by the CHW officers, since these are the ones who knew the physical boundaries of their areas. These GPS vector shapefile boundaries were imported into Qgis (a geographic information system software application) and were further processed to create smooth boundaries and remove accidental gaps between neighbouring CHW areas. Each area was also assigned a numeric code and a name for easy identification.

During the census CHW's accompanied by our study field workers would move around the areas that they were responsible for and enumerate the household members of all the households within the boundaries of that CHW area. Aggregated information of household members was collected using a paper-based questionnaire. According to the census, household members were defined as people who live together and eat meals together. At each household, a household member was interviewed; preference was given to the household head but they were represented by another member that was available at the time of the interview. The household level data that was collected includes the total number of members, male members aged >15yrs, female members aged >15yrs, children aged 5 to 14yrs, toddlers aged 1 to 4yrs and infants <1yr. The paper-based questionnaires were entered into a data base using Access forms. 5% of records entered each week were reviewed for quality checks of the entry process.

2.3.2 Enhanced TB notification surveillance and TB outcomes

Patients diagnosed with TB in Malawi are registered by the National TB Programme. TB registration clinics in Blantyre include a government referral hospital, free public clinics and a small number of private health facilities (11). A collaboration between Malawi-Liverpool-Wellcome Trust Clinical Research Programme (MLW), the Blantyre District Health Office, and the Malawi National TB Control Programme (NTP) has set up an enhanced TB

notification surveillance system in Blantyre. At registration, all TB patients are offered provider-initiated HIV testing and are initiated on antiretroviral therapy if newly diagnosed with HIV (11).

Since 2015, an electronic data capture application (ePAL) has been used to collect clinical, sociodemographic, and household level data, this data includes the data that is collected routinely into the TB registers and study specific data. As part of the same enhanced system, a sputum sample is collected from the patient and sent to the study lab at Malawi College of Medicine Lab for sputum smear microscopy and culture (10,12). ePAL supports the capture of global positioning satellites (GPS) coordinates of TB patient's households. This is made possible because ePAL has preinstalled offline high resolution satellite maps which are geotagged with reference locations ("points of interest") within each neighbourhood of the city. Using the name of place of residence and the closest point of interest that is to the patient's household the TB officer is able to narrow the satellite map image of ePAL to a resolution that the TB patient is able to locate their household. The GPS coordinates are automatically collected by ePAL after pressing at the household location with a finger on the touch screen of the device. The ePAL application has previously been validated and described elsewhere (10,13).

The city was subdivided into 72 neighbourhoods; these neighbourhoods were formed by combining several CHW areas to have approximately 4000 adults in a neighbourhood. Only CHW areas that belonged to the same community health worker were combined. Neighbourhood denominators were derived from the respective CHW areas that made up the neighbourhood from the census data collected in Section 2.3.2

2.3.3 TB treatment outcomes

At the end of TB treatment, treatment outcomes of TB patients are recorded into the national TB registers at the TB treatment health facilities where they were getting their treatment. The TB outcomes are defined according to WHO guidelines on reporting of TB outcomes as outlined in in Section 1.4.4 and Table 1.1 (14). A study team data officer goes around the TB clinics and collects the TB treatment outcomes using an electronic data capture device. Each record of the TB outcome was linked to a TB notification registration of that patient in the database.

2.3.4 TB prevalence

In 2019, a TB prevalence survey was carried out in Blantyre City by the MLW study team at the start of a planned cluster-randomised trial of community-based TB screening interventions, subsequently interrupted by COVID-19 (ISRCTN11400592). 72 neighbourhoods were defined, each comprised of several community health worker (CHW) areas as described in Section 2.3.1. Using Google Earth, a geographical information specialist captured the GPS coordinates of all the houses in the 72 neighbourhoods to be used as the prevalence survey sampling frame. In each neighbourhood, 115 households were selected at random for participation into the prevalence survey with the aim of recruiting 215 adults (≥18 years old) per neighbourhood.

Adults from the randomly selected households were visited and invited to attend a study tent located at a central point within the neighbourhood for TB and HIV investigations. TB screening was provided at the tent using a digital chest radiograph that was immediately read by an experienced radiographer trained in TB prevalence surveys, and with interpretation supported by computer assisted diagnostic software (Qure.ai version 2.0).

Participants who had an abnormal chest X-ray or reported cough of any duration were asked to provide two spot sputum samples for Xpert, smear microscopy, and MGIT culture. Patients with positive TB results were asked for confirmatory specimens and were assisted to register for TB treatment at the nearest TB registration centre. The study team would stay in a neighbourhood for a period of one week before moving to the next neighbourhood.

2.5 Simulation and modelling of neighbourhood prevalence to notification ratios

Data for the third objective of my PhD came from a TB prevalence survey and the TB notification surveillance of Urban Blantyre. Because the prevalence survey was incomplete at the time of presenting my PhD proposal to the PhD upgrade committee, I undertook the following simulation exercise to inform the modelling strategy and help establish the feasibility of the proposed approach for objective 3.

I undertook the simulation exercise of the prevalence cases guided by the preliminary estimates from the ongoing prevalence survey and plausible values that would be expected according to our understanding of the study setting. Regression modelling based on the observed TB notifications and the simulated prevalent data were fitted to the data to check the feasibility of the proposed modelling approach for the calculation of prevalence to notification ratio.

Neighbourhood prevalence simulation approach:

1. Prevalence study neighbourhoods were coded as $j = 1,2,3 \ldots 72$. We fixed the mean number of participants for each neighbourhood to $p_j = 215$. The number of participants were fixed to 215 since that was the number that was aimed for in the prevalence survey referred to Section 2.3.4. We simulated the TB prevalence rate λ_j following the Equation below:

$$\lambda_j = \frac{1}{D}\{0.01(D - d_j) + 0.0025d_j\} + \emptyset_{CAR,j} \text{ (Equation 2.1)}$$

$$where\ \emptyset_{CAR,j}\ are\ conditional\ auto - regressive\ random\ effects$$

$where\ D\ was\ the\ Max\ distance\ from\ the\ center\ of\ Blantyre$ to the centroid of the furthest neighbourhood.

$d_j\ was\ the\ distance\ from\ neighbourhood\ centroid\ of\ neighbourhood\ j\ to\ the\ center\ of\ Blantyre\ city$

The λ_j were therefore allowed to vary between 0.01 to 0.0025. The centre of Blantyre was chosen as the location of the Queen Elizabeth Hospital, which is a referral hospital located approximately at the centre of the city.

The rationale for choosing Equation 2.1 was based on preliminary data and previous studies (15,16). We assumed that TB prevalence will be highest in neighbourhoods towards the densely-populated centre of the city where more people live (15), and that the prevalence will reduce in more distal neighbourhoods. Therefore, we simulated neighbourhood prevalence as a function of distance from the centre of urban Blantyre to each neighbourhood centre (Equation 2.1).

2. The conditional autoregressive (CAR) random effects for each neighbourhood were calculated as follows. We computed an adjacency matrix W that was $J \times J$, where row 1 had 1's for neighbourhoods that were adjacent (with touching boundaries or being nearest to each other) to neighbourhood 1, and 0 otherwise, similarly for rows 2, 3, …72.

   Then we drew $\emptyset_j \sim N\left(0, \sigma_{\emptyset}^2 = 0.002\right), j = 1,2,3 \ldots 72$ and this was put into a vector $\widehat{\emptyset}$ $(1 \times J)$. We further calculated a vector of the number of neighbours for each neighbourhood j, $n = W \times 1$.

   To calculate random effects for each neighbourhood j , $\widehat{\emptyset}_{CAR} = W\widehat{\emptyset}$ and then divided $\widehat{\emptyset}_{CAR}$ by the vector of neighbours n to get the mean random effects from the adjacent neighbours.

3. The prevalent TB cases for each neighbourhood j $were\ simulated\ as\ Y_j \sim Pois(\mu_j = \lambda_j p_j)$ (Equation 2.2)

4. A spatial Bayesian Poisson regression was used with weakly informative prior $N(0,8)$ for intercept and coefficients and $\mathrm{Cauchy}(0,1)$ for the standard deviation of spatial random effects. The models were computed using rstan via the brms R package (17). Ten thousand samples were drawn from each model posterior distribution using Markov Chain Monte Carlo with a burn of 1000 samples. Analysis was done using R version 3.5.2 (R Foundation for Statistical Computing, Vienna).

5. The model fitted to simulated data was specified as below:

$$log\{Y_j, p\} = log(p_j) + \beta_0 + \beta_1 d_j + \cdots + \emptyset_j \ (equation\ 2.3)$$

$$Where\ \emptyset_j \sim N\left(\mu_j, \sigma_j^2\right)$$

$$\mu_j = \frac{1}{n_j} \sum_{k \in N_j} \emptyset_k$$

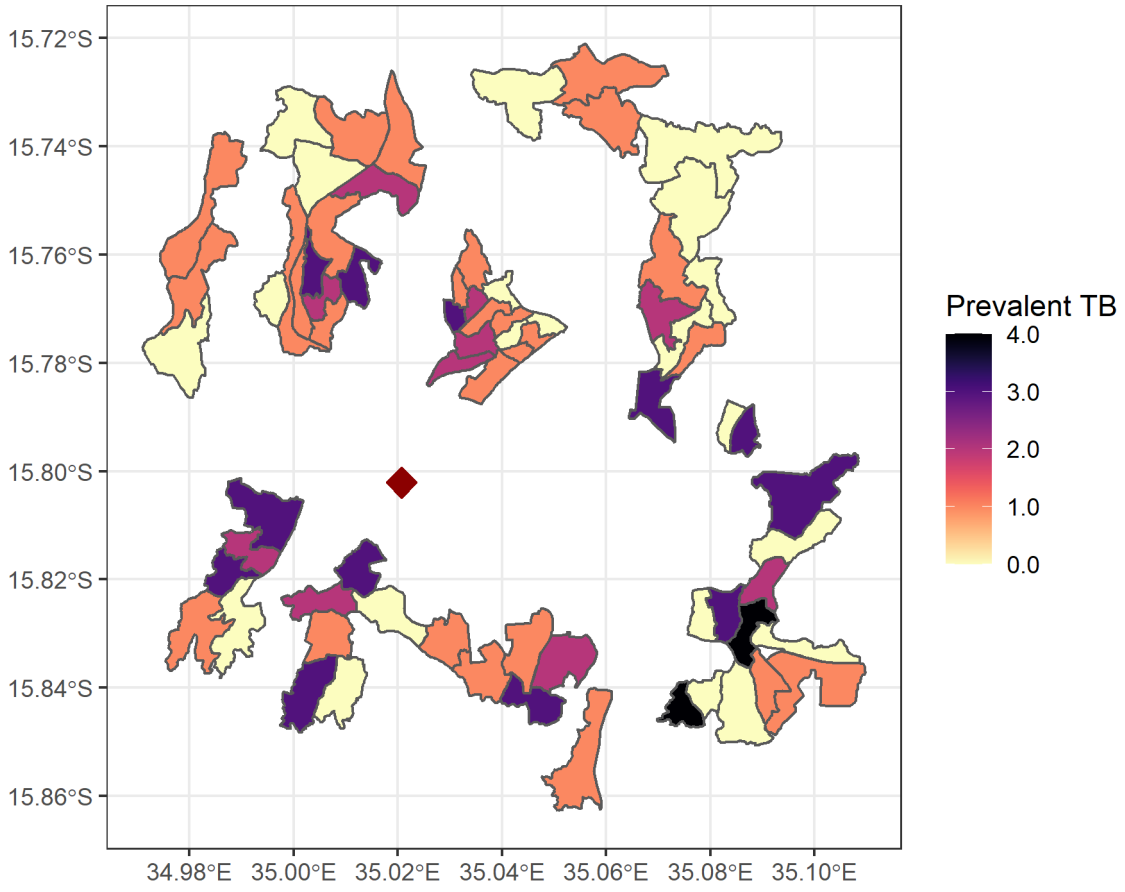
$$\sigma_j^2 = {\sigma_\emptyset^2}/{n_j}$$

j	For neighbourhood j, $j = 1,2,3 \ldots 72$
$Y_{j,p}$	Number of prevalent TB cases in neighbourhood j as simulated by Equation 2.2
$log(p_j)$	The log offset. p_j is the number screened for TB in cluster j. Which was fixed to be 215.
β_0	Intercept
$\beta_1 d_j$	Coefficient for distance for centre of the city to neighbourhood centroid
$\emptyset_j$	Spatial random effect of each area that are modelled as $\emptyset_j \sim N(\mu_j, \sigma_j^2)$
	Where: $\mu_j = \frac{1}{n_j} \sum_{k \epsilon N_j} \emptyset_k$, ${\sigma_j^2} = {\sigma_\emptyset^2}/{n_j}$ $N_j \; are \; neighbours \; of \; size \; n_j$

2.5.2 Results of the simulation modelling:

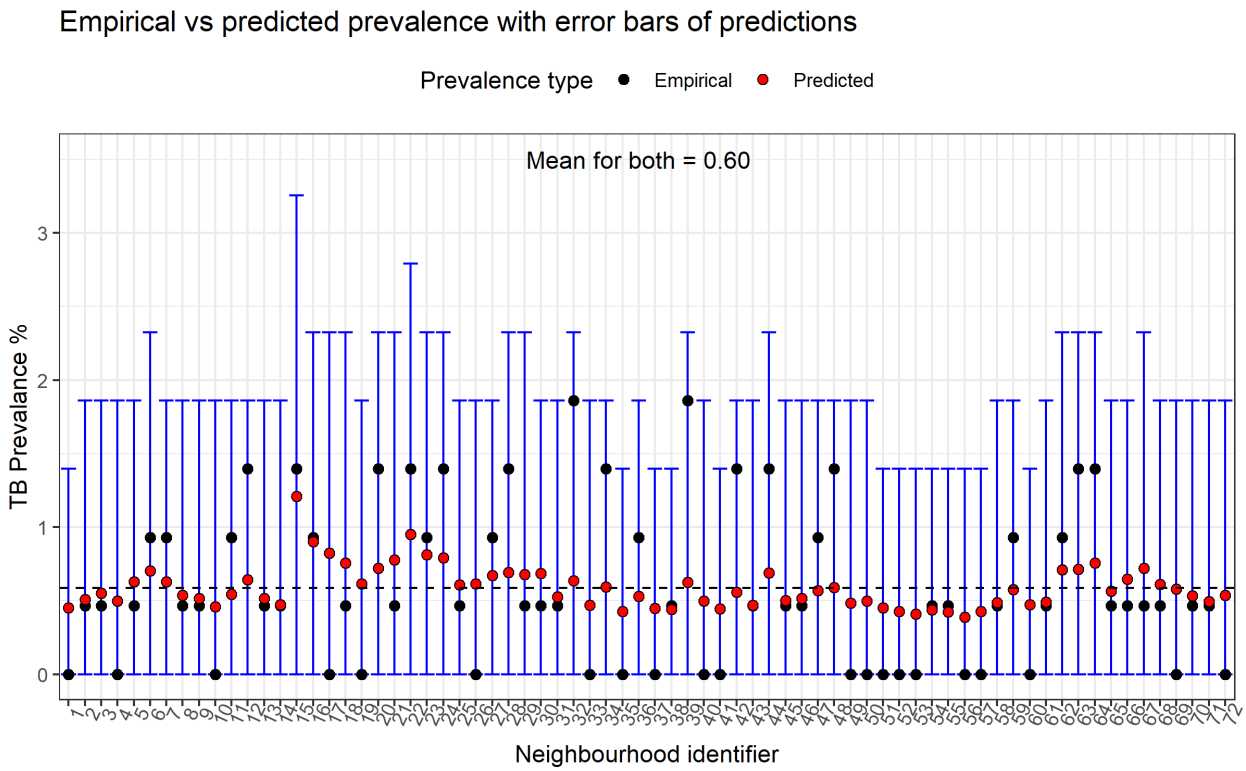
The model computation was good as evidenced by trace plots and Rhat statistics that showed model convergence and sufficient effective samples.

Figure 2.3: The simulated TB prevalence counts per neighbourhood that were estimated from Equation 2.2 below:



The total simulated prevalent TB cases were 92, with a minimum of 0, maximum of 4, and a mean of 1.3. The estimated mean prevalence was 0.60%. The simulation successfully generated the spatial correlation that was likely plausible given our understanding of the local epidemiology of TB in the study setting (15,16).

Figure 2.4: A graph of the simulated TB prevalence (from Equation 2.2) versus the model predicted prevalence (Equation 2.3)



The mean predicted prevalence was similar to that which was simulated 0.6%. The predicted neighborhood-level prevalence's showed shrinkage towards the mean prevalence.

2.5.3 Simulation and modelling of prevalence to notification ratios

We used the prevalent cases (Y_j) that were simulated in Equation 2.2. The data for notification cases ($Y_{j,n}$) came from 2018 notified TB cases in Urban Blantyre, captured by our citywide enhanced TB surveillance system (15). The adult population for each neighbourhood ($p_{j,n}$) came from population census of Urban Blantyre that was done in 2015.

The steps for this analysis were:

1. We fitted Equation 2.1 without covariates to obtain the CAR smoothed prevalence rates for each area via $\lambda_{j,p} = \exp(\beta_0 + \emptyset_j)$ (Equation 2.4)

2. For each area j (j=1,2,3….72) we have $p_{j,n}$ person-years at risk for the notified cases $Y_{j,n}$.
3. A spatial Bayesian Poisson regression was used with weakly informative prior $N(0,8)$ for intercepts and model coefficients and $\text{Cauchy}(0,1)$ for the standard deviation of spatial random effects. The models were computed using rstan via brms R package (17). The spatial random effects were modelled with an intrinsic conditional autoregressive (ICAR) prior. Ten thousand samples were drawn from each model posterior distribution using Markov Chain Monte Carlo with a burn of 1000 samples. Analysis was done using R version 3.5.2 (R Foundation for Statistical Computing, Vienna).
4. The model to predict notified TB cases was specified as below:

$$log\{Y_{j,n}\} = log(p_{j,n} + \lambda_{j,p}) + \beta_0 + \beta_1 d_j + \cdots + \emptyset_j \ (equation\ 2.5)$$

$$Where\ \ \emptyset_j \sim N(\mu_j, \sigma_j^2)$$

$$\mu_j = \frac{1}{n_j} \sum_{k\,\epsilon\, j} \emptyset_k$$

$$\sigma_j^2 = {\sigma_\emptyset^2}/{n_j}$$

j	For cluster j, $j = 1,2,3 \ldots 72$
$Y_{j,n}$	Number of notified TB cases in neighbourhood j
$log(p_{j,n} + \lambda_{j,p}) +$	The log offset. $p_{j,n}$ is the number of TB cases in cluster j in 2018. $\lambda_{j,p}$ is the CAR smoothed prevalence rate for cluster j as obtained from step 1.
β_0	Intercept
$\beta_1 d_j$	Coefficient for distance for centre of the city to cluster centroid
$\emptyset_j$	Spatial random effect of each area that are modelled as $\emptyset_j \sim N(\mu_j, \sigma_j^2)$
	Where: $\mu_j = \frac{1}{n_j} \sum_{k\,\epsilon\, N_j} \emptyset_k$, $\sigma_j^2 = {\sigma_\emptyset^2}/{n_j}$

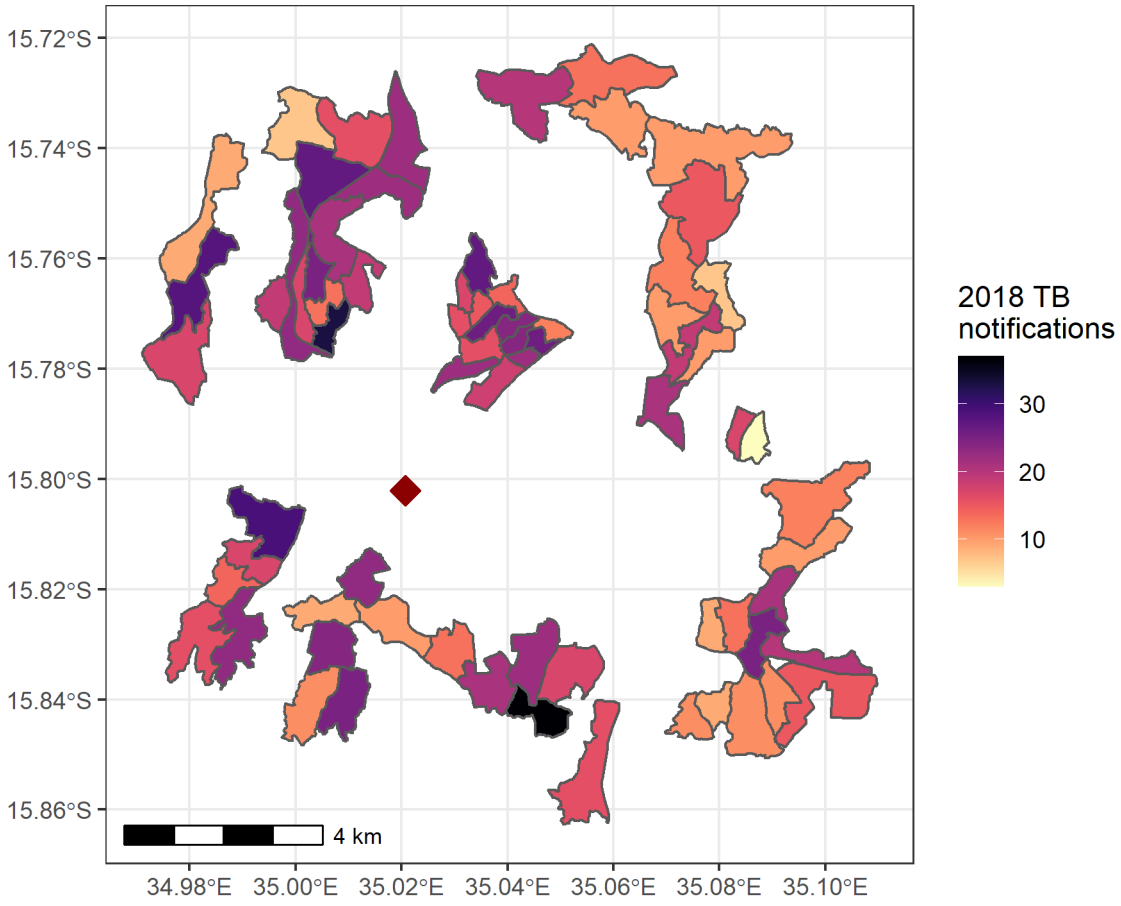
	N_j are $neighbours$ of $size$ n_j

5. The model was then used to predict notified cases for each cluster which were later used to calculate the prevalence to notification ratios (prevalent TB cases per 100,000/Notified TB cases per 100,000).

2.5.5 Results of the models based on the TB notifications and simulated prevalence data

The empirical TB neighbourhood notification distribution in 2018 was minimum 3, maximum 37 and mean 18 (15).

Figure 2.5: 2018 empirical TB notifications



The following graph shows the neighbourhood prevalence to notification ratios (P:N ratio). The prevalence was simulated from Equation 2.1. The P:N ratio was calculated by dividing the posterior prevalence rate (from Equation 2.3) by the posterior notification rate (from Equation 2.5) for each of the 72 neighbourhoods. The median and 95% credible interval of the posterior P:N ratio samples were also calculated for each neighbourhood.

Figure 2.6: Neighbourhood prevalence to notification ratios (green dashed line y-axis line at 1)

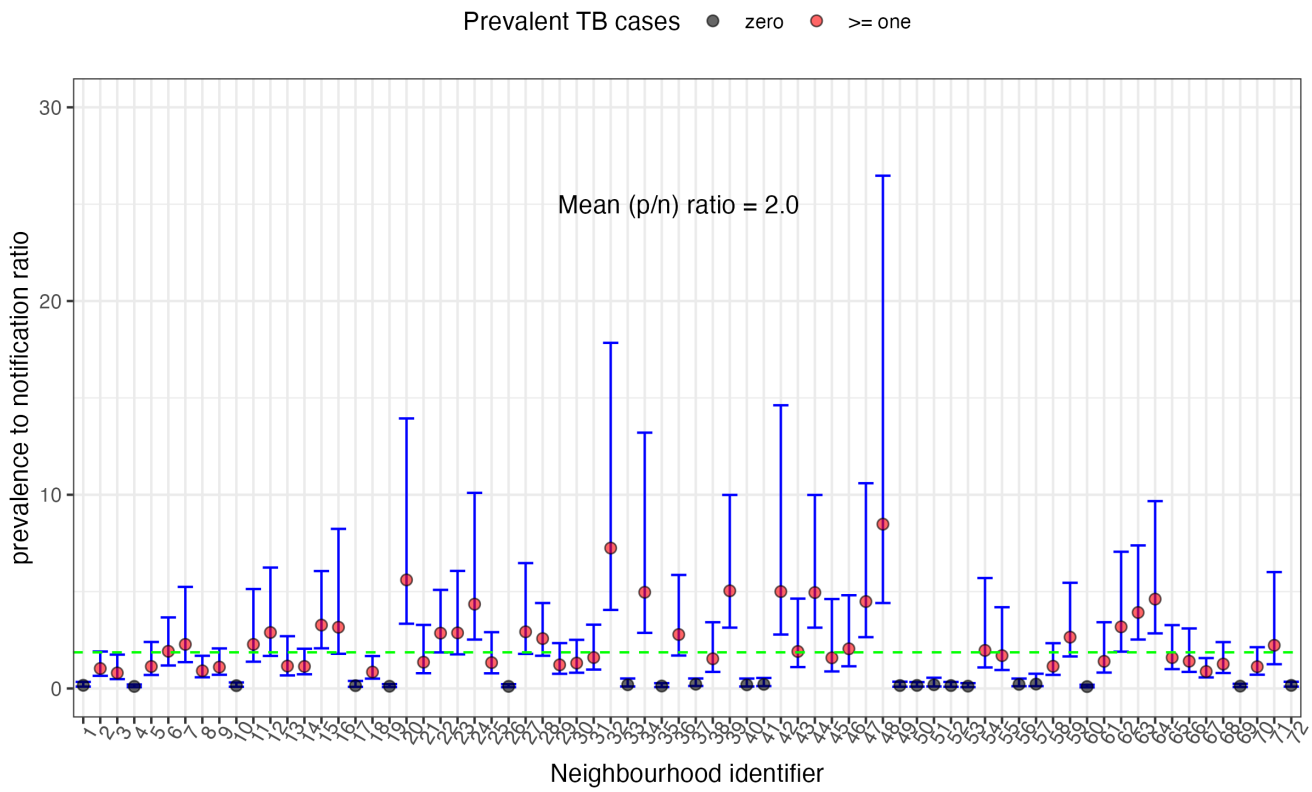
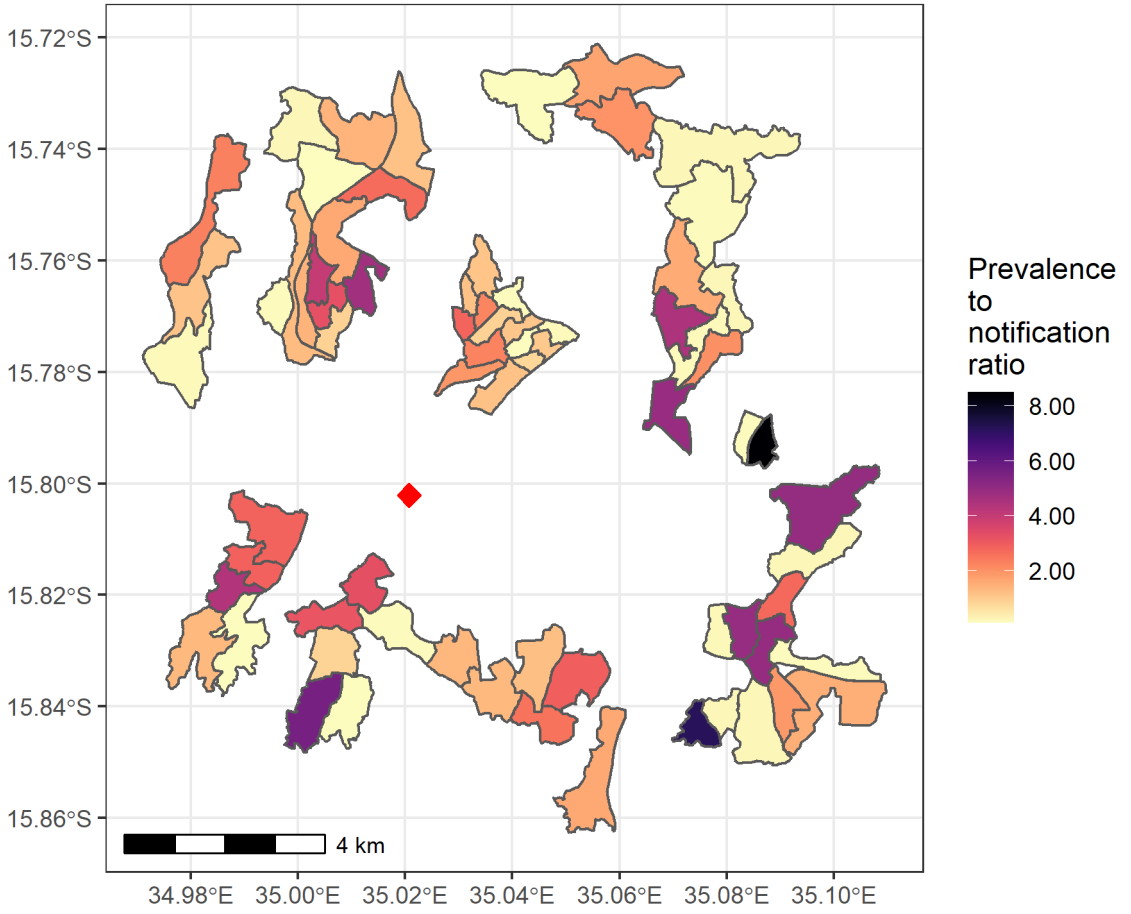


Figure 2.7: A map of the prevalence to notification ratios (red diamond is the location of QECH)



The overall mean neighbourhood P:N ratio from this simulation exercise was 1.87 with a 95% credible interval of 0.11 to 5.96 and a range of 0.10 to 8.50. The areas with the highest P:N ratios were assumed to have a higher likelihood of undiagnosed TB cases based on this scenario of simulated prevalent TB cases. This analysis helped in proving the feasibility of the methods that were proposed for the PhD thesis.

2.5.6 Results of the models based on the TB notifications and simulated prevalence data

This exercise demonstrated that even when empirical TB prevalence is low (as was observed in the ongoing prevalence survey), we were confident that we would be able to fit multi-level spatial models with acceptable levels of uncertainty. In addition, this exercise assisted in coming up with a code template that was used to run the final models.

2.6 References

1. U.S. Agency for International Development. MALAWI-PROPERTY RIGHTS AND RESOURCE GOVERNANCE PROFILE 1 [Internet]. [cited 2019 May 22]. Available from: http://usaidlandtenure.net/malawi
2. Mcsweeney C, New M, Lizcano G. UNDP Climate Change Country Profiles Malawi General Climate [Internet]. [cited 2019 May 17]. Available from: http://country-profiles.geog.ox.ac.uk
3. United Nations committee for development policy. List of Least Developed Countries (as of December 2018) [Internet]. New York City; [cited 2019 May 15]. Available from: https://www.un.org/development/desa/dpad/wp-content/uploads/sites/45/publication/ldc_list.pdf
4. World Bank. Malawi recent developments [Internet]. Washington DC; [cited 2019 May 17]. Available from: http://pubdocs.worldbank.org/en/395451492188166005/mpo-mwi.pdf
5. Malawi Systematic Country Diagnostic: Breaking the Cycle of Low Growth and Slow Poverty Reduction [Internet]. 2018 [cited 2019 May 17]. Available from: http://documents.worldbank.org/curated/en/723781545072859945/pdf/malawi-scd-final-board-12-7-2018-12122018-636804216425880639.pdf
6. National Statistical Office of Malawi (NSO). Malawi population census preliminary report [Internet]. Zomba; 2018 [cited 2019 May 17]. Available from: www.nsomalawi.mw
7. Banda R. Results from the national TB prevalence survey of Malawi. In: 45th World Conference on Lung Health of the International Union Against Tuberculosis and Lung Disease. Barcelona; 2014.
8. Malawi M of H of. Malawi Population-based HIV Impact Assessment 2015-2016. 2018.
9. World Health Organisation. Global Tuberculosis Control Report 2018. Geneva; 2018.
10. MacPherson P, Khundi M, Nliwasa M, Choko AT, Phiri VK, Webb EL, et al. Disparities in access to diagnosis and care in Blantyre, Malawi, identified through enhanced tuberculosis surveillance and spatial analysis. BMC Med [Internet]. 2019 Dec 29 [cited 2019 Feb 6];17(1):21. Available from: https://bmcmedicine.biomedcentral.com/articles/10.1186/s12916-019-1260-6
11. Malawi Ministry of Health. Malawi National Tuberculosis Control Programme 7th edition [Internet]. Lilongwe; 2012 [cited 2019 May 26]. Available from: https://pdf.usaid.gov/pdf_docs/PA00KGDD.pdf
12. Soko RN, Burke RM, Feasey HRA, Sibande W, Nliwasa M, Henrion MYR, et al. Effects of Coronavirus Disease Pandemic on Tuberculosis Notifications, Malawi. Emerg Infect Dis [Internet]. 2021 Jul [cited 2021 Jun 23];27(7):1831–9. Available from: https://wwwnc.cdc.gov/eid/article/27/7/21-0557_article.htm

13. Macpherson P, Choko AT, Webb EL, Thindwa D, Squire SB, Sambakunsi R, et al. Development and validation of a global positioning system-based “map book” system for categorizing cluster residency status of community members living in high-density urban slums in Blantyre, Malawi. Am J Epidemiol [Internet]. 2013 May 15 [cited 2019 Feb 15];177(10):1143–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23589586

14. World Health Organization. Definitions and reporting framework for tuberculosis-2013 revision (updated December 2014 and January 2020) [Internet]. Geneva; 2020 [cited 2021 Dec 2]. Available from: https://apps.who.int/iris/bitstream/handle/10665/79199/9789241505345_eng.pdf

15. MacPherson P, Khundi M, Nliwasa M, Choko AT, Phiri VK, Webb EL, et al. Disparities in access to diagnosis and care in Blantyre, Malawi, identified through enhanced tuberculosis surveillance and spatial analysis. BMC Med. 2019;17(1):21.

16. World Health Organisation. WHO | Global tuberculosis report 2018. Geneva; 2018.

17. Bürkner P-C. Advanced Bayesian Multilevel Modeling with the R Package brms. R J [Internet]. 2018 [cited 2019 Oct 15];10(1):395. Available from: https://journal.r-project.org/archive/2018/RJ-2018-017/index.html

3 Systematic Review

3.1 Introduction

This chapter is a systematic literature review that was undertaken to examine the existing evidence regarding the effectiveness of spatially targeted interventions for the control of HIV, TB, leprosy, and malaria. I was the lead author of the protocol which was submitted to Prospero and I also led the writing of this manuscript, including coordinating with co-authors so that they were able to complete their assigned tasks.

In communities, HIV, TB, leprosy and malaria cluster geographically into hotspots, and community interventions can be targeted at these hotspots (spatially targeted interventions). The other diseases were included because spatially targeted interventions are relatively new, and we wanted to also learn the progress that has been made in controlling these other epidemics. This review was submitted on 16 September 2020 to BMJ Open Journal and was published on 15 June 2021 and has been reproduced below.

In summary, we concluded that, although data was limited and understanding of effectiveness inferred by high risk of bias (particularly in classification of hotspots and ascertainment of outcomes), the spatially-targeted interventions demonstrated potential to identify communities with a higher yield of identified cases, communities with a high prevalence of cases, and hence potentially for accelerating reductions in the TB case notifications rates.

3.2 Systematic review manuscript

LONDON SCHOOL of HYGIENE & TROPICAL MEDICINE

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT

T: +44 (0)20 7299 4646
F: +44 (0)20 7299 4656
www.lshtm.ac.uk

RESEARCH PAPER COVER SHEET

Please note that a cover sheet must be completed for each research paper included within a thesis.

SECTION A – Student Details

Student ID Number	lsh1805281	**Title**	Mr
First Name(s)	McEwen Joseph		
Surname/Family Name	Khundi		
Thesis Title	Towards a framework for improved targeting of tuberculosis interventions: a spatial analysis of patient notification and prevalence survey data from Blantyre, Malawi		
Primary Supervisor	James Carpenter		

If the Research Paper has previously been published please complete Section B, if not please move to Section C.

SECTION B – Paper already published

Where was the work published?	BMJ Open		
When was the work published?	15 June 2021		
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion			
Have you retained the copyright for the work?*	**No**	Was the work subject to academic peer review?	**Yes**

*If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be published?	
Please list the paper's authors in the intended authorship order:	

Stage of publication	Choose an item.

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	I developed the research question, wrote the systematic review protocol and submitted it to prospero, led the analysis of the systematic review, led the writing of the manuscript, and submitted it for publication. I am the first author.

SECTION E

Student Signature	[redacted]
Date	7 July 2022

Supervisor Signature	[redacted]
Date	21.VII.2022

EFFECTIVENESS OF SPATIALLY-TARGETED INTERVENTIONS FOR CONTROL OF HIV, TUBERCULOSIS, LEPROSY AND MALARIA: A SYSTEMATIC REVIEW

Corresponding author:
McEwen Khundi
Malawi Liverpool Welcome Trust
P.0 Box 30096
Malawi,
mcewenkhundi@gmail.com
Tel: +265 1811628
Fax: +265 1875774

James Carpenter
Department of medical statistics
London School of Hygiene and Tropical Medicine
London
United Kingdom

Marriot Nliwasa
Department of Pathology
College of Medicine, Malawi
Blantyre
Malawi

Ted Cohen
School of Public Health
Yale University
Connectut, New Haven
USA

Elizabeth L Corbett
Department of Clinical Research
London School of Hygiene and Tropical Medicine
London
United Kingdom

Peter MacPherson
Department of Clinical Sciences
Liverpool School of Tropical Medicine
Liverpool
United Kingdom

Keywords: infectious diseases, community interventions, spatial hotspot, spatially targeted, systematic review.
Word Count:
Abstract: 287
Main text: 4658

Abstract (287 words)

Background: As infectious diseases approach global elimination targets, spatial targeting is increasingly important to identify community hotspots of transmission and effectively target interventions. We aimed to synthesise relevant evidence to define best practice approaches and identify policy and research gaps.

Objective: To systematically-appraise evidence for the effectiveness of spatially-targeted community public health interventions for HIV, Tuberculosis, Leprosy and Malaria.

Design: Systematic review.

Data sources: We searched Medline, Embase, Global Health, Web of Science and Cochrane Database of Systematic Reviews between 01 January 1993 and 22 March 2021.

Study selection: The studies had to include HIV or TB or leprosy or malaria and spatial hotspot definition, and community interventions.

Data extraction and synthesis: A data extraction tool was used. For each study, we summarised approaches to identifying hotpots, intervention design and effectiveness of the intervention.

Results: Ten studies, including one cluster randomised trials and nine with alternative designs (before-after, comparator area), satisfied our inclusion criteria. Spatially targeted interventions for HIV (one United States of America (USA) study), TB (three USA) and leprosy (two Brazil, one Federated States of Micronesia) each used household location and disease density to define hotspots followed by community-based screening. Malaria studies (one each from India, Indonesia and Kenya) used household location and disease density for hotspot identification followed by complex interventions typically combining community

screening, larviciding of stagnant water bodies, indoor residual spraying and mass drug administration. Evidence of effect was mixed.

Conclusions: Studies investigating spatially targeted interventions were few in number, and mostly underpowered or otherwise limited methodologically, affecting interpretation of intervention impact. Applying advanced epidemiological methodologies supporting more robust hotspot identification and larger or more intensive interventions would strengthen the evidence-base for this increasingly important approach.

PROSPERO registration: CRD42019130133

Strengths and limitations of the study

- This was a thoroughly conducted systematic review which only included published literature.
- We developed the search strategy with input from infectious disease experts, statisticians and librarians, and we only included studies that met our objective assessment criteria.
- We acknowledge that some studies might still have been missed even with the systematic approach that we used.
- The effect of publication bias cannot be ruled out as it is possible that studies that had negative results might not have been published; thereby, they would not have been captured by our search strategy.
- The studies had different interventions and outcomes as such a meta-analysis was not done contrary to the initial plans.

Introduction

The world has made tremendous strides in controlling human immunodeficiency virus (HIV), tuberculosis (TB), leprosy and malaria epidemics. Despite impressive achievements in reductions of new cases, considerable efforts are required to meet and maintain elimination targets (1–4). Much of the current progress in disease control is attributable to the combination of facility based routine health services, supplemented by community-based interventions (1–4).

The success of disease control strategies provided through facility-based services relies on prompt recognition of symptoms by the patient, early health seeking by the patient, and correct recognition and management by health providers to provide early diagnosis and effective medical care (1–4). However, health-seeking delays can be prolonged, and this strategy will also have no effect on transmission during subclinical illness (2,5–7). In low-income settings, lack of faith in the quality of services, high opportunity and indirect costs, and non-adherence to syndromic management protocols can combine to cause substantial health-seeking and diagnostic delays, undermining the effectiveness of facility-based strategies based on early diagnosis and treatment (2,4,5,8–12). Community interventions can be complementary to facility-based services but are more resource intensive, logistically challenging and are generally only justifiable when targeted at groups of people at risk or in a geographically defined area with high prevalence or incidence of disease (1–4).

In communities, HIV, TB, leprosy and malaria cluster geographically into hotspots (13–19). Hotspots are defined as areas of high incidence or prevalence compared to neighbouring geographical areas (20). For malaria, which is vector-borne, environmental characteristics conducive to replication of anopheles mosquitos are a key consideration (18). For HIV, TB

and leprosy, diseases that are transmitted exclusively person-to-person, hotspots tend to be characterised by poverty, poor access to health services, overcrowding, concentrations of migrant populations, and poor housing (13–18). Since hotspots are likely to be areas of relatively high transmission, targeting interventions at hotspots may prevent many more infections and cases of disease than similar efforts in low-transmission settings or untargeted efforts (**Figure 3.1**) (21). Public health interventions amenable to spatially-defined hotspot targeting include: screening, case-finding, prevention (including vaccination), and improvement in access to services for diagnosis and treatment (15,20).

Community health and prevention interventions frequently target people in slums or informal urban settlements (22). In low and middle income countries (LMIC), many millions of people live in informal urban settlements (23), and disease control programmes would benefit from more precise identification of high priority areas within these to allow interventions to be delivered to communities where the impact is likely to be greatest (24,25). At the moment targeting tends to use relatively crude epidemiological criteria such as age and gender (1–4,26). But since cases of the diseases in this review are known to cluster geographically, targeting of all people in a carefully identified geographically-defined hotspots offers an alternative criterion to define at-risk groups for interventions (13–19).

The main aim of this systematic review was to systematically-appraise evidence for spatially-targeted community public health interventions directed towards major infectious diseases that are transmitted in different ways: HIV is predominantly sexually transmitted; TB by respiratory droplet transmission; leprosy by direct contact and droplet transmission; and malaria by vector-borne transmission. We aim to summarise lessons learnt from evaluation of spatially targeted interventions against these diseases and to make recommendations for

researchers, policymakers and disease control programmes, as well as to inform the design and evaluation of future studies investigating spatially-targeted interventions.

Methods

Study design

Systematic literature review.

Public involvement statement

We developed the research question for this project after realising the evidence gap on the effectiveness of spatially targeted interventions despite the growing interest that this approach has. To develop the scope of the research question we engaged experts in public health interventions of HIV, tuberculosis, leprosy and malaria. The search strategy was also developed in consultation with the experts. The results of the research will be disseminated to researchers and policy makers at local and international conferences.

Search strategy

We systematically searched the literature using major subject headings and keywords to identify published studies meeting the inclusion criteria following our published protocol (PROSPERO ID: CRD42019130133). Databases searched included: Medline, Embase, Global Health, Web of Science and the Cochrane Database of Systematic Reviews. The three central concepts included in our search strategy (**Table 3.1**) were the disease condition (HIV, TB, leprosy and malaria); space (techniques used to identify hotspots); and community interventions.

Eligibility criteria

We included studies published between the period from 01 January 1993 and 22 July 2019 and then updated to 22 March 2021. The start of the search period was chosen as the year that TB was declared as a "global emergency" by the World Health Organization (WHO). We

focused on studies of spatially targeted interventions for HIV, TB, leprosy and malaria with the aim of identifying all the available literature on the effectiveness of spatially targeted intervention in the selected diseases, and then summarising the findings. We did not limit inclusion by age group of participants or geographic region. We included randomised controlled trials, non-randomised observational studies, before-and-after studies and time-series analysis studies. The following articles were excluded: editorials; narratives; systematic reviews; case studies; case reports; case-control studies; contact investigation studies; and non- spatially targeted community studies.

Selection of studies and data extraction

We imported studies into an Endnote (Thompsons Reuters) database, and duplicates were removed. McEwen Khundi (MK) and Marriot Nliwasa (MN) independently screened the title and abstracts against inclusion and exclusion criteria (**Table 3.2**), identified studies eligible for full-text review, and subsequently reviewed the full text of each selected study independently. Discrepancies between MK and MN were resolved with discussion, and where agreement could not be reached, two other reviewers (Peter MacPherson (PM) and James Carpenter (JC)) participated in a consensus review.

We developed a data extraction form that was piloted on a sample of selected manuscripts. This form was independently completed for each study selected for full-text review by MK and MN.

Assessment of study quality

Two reviewers (MK and MN) assessed the quality of each study. We used separate tools to assess the methodological quality of cluster randomised trials (Cochrane Collaboration risk

of bias tool) (27) and non-randomized studies (ROBINS-I tool for risk of bias assessment) (28).

Definitions

We defined spatially-targeted interventions as community interventions that targeted hotspots of disease. Hotspots were defined as sub-district geospatially-defined areas that had a high number of incident or prevalent cases of the infection or disease compared to surrounding areas, studies that based their hotspots around diagnosed index cases were not included. In practice, as definitions varied considerably between studies, we extracted and compared hotspot definitions between studies.

Statistical analysis

The main objective of the systematic review was to compare the impact of the intervention in the intervention hotspot areas compared to control areas. We report the outcomes defined by the included studies, with outcomes being prevalence, incidence, case notification rates, and number needed to screen to identify a positive case. We calculated measures of effect and uncertainty in effect estimates using data available in manuscripts where these were not provided by the authors using R 3.61 (R Core Team). Because of high anticipated heterogeneity between and within diseases and interventions investigated, we decided not to undertake meta-analysis.

Ethics statement

This review used published data, and ethical review was not required.

Data sharing statement

The search strategy and summary data tables for this systematic review have all been included in the main text and in the appendix section.

Results

A total of 3919 unique abstracts were identified by the search strategy, from which 3886 were excluded from the title and abstract review, leaving 28 studies that were reviewed as full-text articles. Overall, 10 studies met the inclusion criteria (**Figure 3.2**). The reasons for exclusion are found in **Table 3.2**.

One United States of America (USA) based study was identified for HIV. Three TB studies were found: all three studies were from the USA. Three leprosy studies were found: two from Brazil and one study from the Federated States of Micronesia. For malaria, we identified three studies: one each from India, Indonesia and Kenya. The most common study design was implementation demonstration studies (i.e. studies in which spatially targeted interventions were introduced without random allocation of hotspots to study arms). Nine studies had this design: one HIV, three TB, three leprosy and two malaria. We identified one cluster randomised trial (CRT), which evaluated interventions targeted against malaria (**Table 3.3)**.

Table 3.4 summarises the characteristics of the included studies, presenting outcomes, measures of intervention effects and study quality/risk of bias assessment results. **Table 3.5** synthesises methods used to geolocate cases and identify hotspots in the included studies. **Table 3.3** has details on why studies were excluded following full text review.

Spatially-targeted HIV interventions

In Goswami et al. (29), cases of TB (N=150), HIV (N=665) and syphilis (N=155) notified between 1 Jan 2005 and 31 Dec 2007 in Wake County, North Carolina USA, were geocoded to households. Kernel density maps of the three diseases were generated; maps identified

two areas with the densest number of cases (hotspots). Hotspots were defined as areas with more than 10 cases of either TB, HIV or syphilis notified per square mile during the three years. A map of streets and local businesses of Wake County was used to identify the locations of the hotspot areas.

Between 2 June 2009 and 3 Nov 2011, adult community screening for HIV was done in the defined hotspots at specific sites by community nurses and disease intervention specialists from the HIV, syphilis, and TB clinics at the county health department. Of 247 community participants screened by the study, 240 had valid HIV test results. Prevalence of HIV in hotspot areas was compared to that among patients presenting to a sexually transmitted disease (STD) clinic located outside of the hotspot areas. HIV prevalence was higher among community screened participants (8/240, 3%, 95% CI:1.4-6.5) compared to the Wake County STD clinic (64/15936, 0.4%, 95%CI: 0.3-0.5) with a risk ratio of 8.3 (95% CI: 4.0-17.1, $p<0.001$). Community HIV screening identified eight HIV positive cases, only one of whom was previously undiagnosed.

Spatially-targeted tuberculosis interventions

In Moonan et al. (24), notified TB cases from 1 Jan 1993 to 31 Dec 2000 (N=991) were geolocated to zip codes in Tarrant County, north-central Texas, USA. Three zip codes were found to have the highest TB notification rates per 100,000 of the population, in addition to having genotypically-clustered *Mycobacterium tuberculosis* isolates (24). The TB notification rates of the zip codes were 94, 55 and 32 cases per 100,000 while the case notification rate for the whole county for the same period was 5.9 cases per 100,000; 95 of 117 (81%) isolates were genotypically clustered in these three zip codes (30). Between 1 Sep 2002 and 31 Dec 2004 community-based organizations offered screening for TB and latent TB

infection (LTBI) to hotspot residents. Overall, 3,645 individuals were screened. 1.2% (N = 44) people were diagnosed and treated for active TB, and 18.6% (N=681) were diagnosed and treated for LTBI. This targeted screening identified one person with active TB for every 83 screened, and one person with LTBI for every five screened. The yield of the targeted approach was considered to be more than what would be expected in a county with an active TB notification rate of 5.7 per 100,000 population year.

Goswami et al. (29) (see above in HIV section) also offered screening for LTBI to the same individuals who were offered HIV testing. Both tests were performed from the same blood sample. Hotspot prevalence was compared to LTBI prevalence at the Wake County TB clinic, which was outside of the hotspot areas. Latent TB testing at the TB clinic was offered to high risk individuals who had close contact with a recently diagnosed active TB cases, refugees and those referred by primary care providers or employers. 234/247 had valid LTBI screening results. LTBI prevalence was higher among community screened participants (36/234, 15%, 95% CI:11.0-21.7) versus 6% (95%CI: 5.6-6.6) at the TB clinic with a risk ratio of 2.5 (95% CI: 1.9-3.5, $p<0.001$).

In Cegielski et al. (31), notified TB cases in Smith County, Texas USA between 1985 to 1995 (N=128) and all notified LTBI cases from 1993 to 1995 (N=311) were geocoded to their households. The geocoded cases were loaded into geographical information systems (GIS) software to produce a point map of both active TB and LTBI cases. The two densest neighbourhoods of TB and LTBI cases were identified visually from the map. In 1996, study field workers went door to door in these neighbourhoods and offered tuberculin skin testing (TST). 1236 /2258 had LTBI testing and received results, 229/1236 (18.5%) were TST positive and 147 received treatment. To assess the intervention, the notified TB cases from 1996 to

2006 in Smith County were mapped to do a before and after intervention comparison. The TB notification rates in the targeted hotspots declined from 39.6 per 100,0000 people per year (95% CI:30.4-48.8) from 1985 to 1995 to zero from 1996 to 2006 ($p<0.001$). While for the entire Smith County the TB notification rates reduced from 8.1 per 100,0000 people per year (95% CI: 5.2-11.0) from 1985 to 1995 to 3.7 per 100,000 people per year (95% CI: 1.2-6.1) from 1996 to 2006 ($p<0.001$)

Spatially-targeted leprosy interventions

In De Souza Dias et al. (32), leprosy cases notified between 1998 and 2002 (N=368) in the municipality of Mossoro, Rio Grande do Norte in Brazil were geocoded to households. The geocoded cases were used to create a density map with a radius of 100m, and four neighbourhoods with the highest concentration of cases were identified as hotspots. Four active case finding (ACF) campaigns were conducted in these hotspots between March and September 2005. Study team members went door to door to identify people with symptoms of leprosy and referred them to the nearest primary health clinic. 512 possible leprosy cases were referred, and 104 leprosy cases were diagnosed. The cases identified through hotspot ACF represented 50% of the total cases diagnosed in the city in 2005. In addition, the case notification rate in 2005 was higher than in 2004; 9.34 per 10,000 versus 5.16 per 10,000 respectively.

In Jim et al. (33), in the state of Pohnpei in the Federated States of Micronesia, notified leprosy cases from 2002 to 2006 (N=502) were geocoded to households, producing a point-density map of 1 mile radius; areas with the densest areas were identified for ACF. During 2007 to 2008, ACF teams undertook door to door visits and screened household members for leprosy. There was an eight-fold statewide decrease in the number of households that

were screened between 2007 to 2009, while the number of identified cases was similar to the pre-spatially targeted ACF campaign period of 2002 to 2006.

In Barreto et al. (25), notified leprosy cases from the two municipalities of Castanhal and Oriximinal in the state of Para, Brazil from January 2004 to February 2010 (N=633) were geocoded to households. The notified leprosy cases were aggregated by census tract to calculate census case notification rates, and a spatially empirical case detection rate was used to smooth notification rates. Census tracts were classified into four categories, with the highest category assigned to areas with >40 cases per 100,000 population. Kulldorff's spatial scan statistic was used to define groups of hotspot census tracts with statistically higher than average notification rates. Two schools were selected: one within a hotspot and one from a census tract with ≥40 notified leprosy cases per 100,000 populations. At these schools, 134 students with a mean age of 10 years were screened and 11/134, 8.2% (95% CI:3.5-13.0) were diagnosed with leprosy based on clinical signs and symptoms. The diagnostic yield in hotspot area schools was significantly higher than in students from randomly-selected schools from eight municipalities in the state of Para in a cross-sectional study between 2009 to 2011: 63/1592, 3.9% (95% CI:3.0-4.9) with a risk ratio of 2.1 (95% CI: 1.1-3.8).

Malaria

In Srivastava et al. (34), notified malaria cases between 2000 to 2005 obtained from the State Department of Health in Madhya Predesh were aggregated at block and district level. Malaria hotspots were defined based on the percentage of malaria cases that had *P. falciparum* malaria of all notified cases from 2000 to 2005. Unspecified targeted malaria interventions in 2007 were evaluated for effectiveness by comparing overall number of

confirmed cases notified in 2006 with those notified in 2007. Absolute numbers of notified cases decreased by 5.7% (95%CI:4.7-6.7) from 96,042 in 2006 ($p < 0.05$).

In Herdiana et al. (35), in Sabang island, Indonesia, documented malaria cases from 2007 to 2008 and self-reported malaria cases identified during a survey (N=319), were geocoded to households to produce point maps that classified 14 out of 18 villages as hotspots, based on absolute number of cases. From May 2010, home visits were conducted twice a month in hotspot villages, and only once in non-hotspot villages. Malaria blood smears were taken from anyone with current or recent history of fever, with smear-positive participants referred for treatment. Household contacts and neighbours within 500 meters of smear-positive participants were also screened. Interventions in hotspot areas included: improving malaria diagnostic labs, introduction of Artemisinin-based Combination Therapy (ACT) for malaria treatment, scale-up of indoor residual spraying (IRS) and distribution of long-lasting insecticide-treated nets (LLINs). The incidence of malaria in hotspot areas decreased by 30-fold from 3.18 to 0.13 per 1000 population from 2008 to 2011. (34) (35) (36)

In Bousema et al. (36), between June and July 2011 in Rachuonyo, western Kenya, 17,503 individuals were screened for a malaria prevalence survey, including *P. falciparum* antibodies. SaTScan software was used to define hotspots based on the prevalence of antibody-positivity and age-adjusted antibody density. In total 27 hotspots were identified. A randomised controlled trial randomly allocated ten hotspots 1:1 to either intervention or control. The intervention activities were weekly larviciding of stagnant water bodies, provision of LLINs, IRS and mass drug administration (MDA). MDA was only administered to households that had a confirmed malaria case; febrile individuals (temperature > 37.5^0C) and children aged 6 months – 15 years were offered malaria rapid diagnostic tests (RDT).

The following standard interventions were available to both arms, hotspots and evaluation zones (the area 500 metres around each hotspot): annual IRS, case management at health facilities and distribution of LLINs from antenatal clinics.

There was no statistically significant difference in parasite prevalence in evaluation zones (area around the hotspot) at 8 weeks and 16 weeks post-intervention time points. The first evaluation zone (1-249m) at the 8th week found 3.6% (95% CI: -2.6 to 9.7%, p=0.216) and the second evaluation zone (250-500m) at the 8th week 3.8% (95% CI: -2.4 to 10.0%, p=0.187). Neither was there any significant difference in the first evaluation zone (1-249m) at the 16th week: 1.0%, (95% CI: -7.0 to 9.1%, p=0.713) and the second evaluation zone (250-500m) at 16th week: 1.0% (95% CI: -8.3 to 10.4%, p=0.809).

Influence of study quality on results

The risk of bias assessment of the studies in this review focused on how the hotspots were selected and how the intervention were assessed in each of the studies that were included. The studies in this review were found to have issues that would make them susceptible to risk of bias. Hotspot selection was based mainly on notified cases (24,25,29,32–35,37,38). Relying on notified cases alone can introduce selection bias because notified cases can over-represent cases from areas that have good access to health systems (15).

Further, identified geographical hotspots need to be investigated for stability overtime to ensure that they display spatio-temporal consistency in notification trends (15,39). This was not accounted for in the studies included in this review (24,25,40,29,32–38).

The majority of the identified studies were implementation/pragmatic studies (24,25,29,31–35,40,41). These did not randomise hotspots to interventions and therefore the studies

might have compared outcomes of the study in groups of people that had baseline characteristics that were different.

Discussion

Main findings

The key finding of this review is that, despite increasing enthusiasm for the application of spatially targeted interventions in public health research addressing infectious diseases (42), very few studies have rigorously evaluated the effectiveness of such approaches against non-hotspot comparator areas. We identified only 10 studies conducted since 1993, nearly all of which had substantial limitations in how hotspots were defined and identified, how spatially-targeted interventions were evaluated, or how comparator areas were selected. With the limited evidence available, we found some suggestion that hotspot-targeted interventions for tuberculosis, malaria and leprosy may be efficient and effective approaches in increasing diagnostic yield, reducing unnecessary screening, and perhaps in improving disease epidemiology. However, almost all the studies were vulnerable to bias due to regression to the mean: hotspots identified by highest prevalence/incidence will typically see prevalence/incidence decline, even in the absence of an intervention. As such, spatially-targeted approaches hold promise, but require further evaluation in high-quality studies to guide policymakers considering implementing this approach.

Strengths and limitations

We found fewer studies than we anticipated, potentially due to the challenges of collecting spatial data in developing countries where most areas do not yet have municipal address systems or subdistrict postal code systems (15,43). Our search strategy was designed to be inclusive, and we did identify one previous systematic review of spatially targeted intervention for TB (44). However, for data synthesis we included only primary manuscript sources and ensured that studies included in the previous, more narrowly focused

systematic review were also included here. Nevertheless, in recent years technologies for obtaining spatial data have improved, become cheaper and hence more widely available; consequently, the inclusion of spatial data within epidemiological and surveillance studies as a tool to identify disease hotspots has become more common (42). In theory, this should enable disease control programmes to collect spatial data for identifying and targeting hotspots (42)(43). Thus, in future, prioritising complete mapping of low-income countries, with introduction of effective residential identification systems should be a major priority. This would allow some of the methods for geolocation of notified cases in the studies in this review, which used manual and labour intensive approaches that also had high potential for imprecision, to be replaced with more efficient approaches (33).

Investment in national geospatial and surveillance tools that allow rapid, accurate and scalable geolocation of cases from within health facilities in settings that lack postal codes would greatly facilitate spatial mapping of areas of interest (45). Such tools would enable the identification of hotspots at finer scale and limit the use of arbitrary blocks (34). One important concern about targeting interventions on the basis of case-notifications alone, however, would be the potential for ascertainment bias to further disadvantage underserved populations from which cases are already under-notified due to limited access to health services (46). For populations where this is likely, the ideal approach would be to investigate disease burden using prevalence surveys, or developing models that include data on geographical distribution of measures of poverty and health service access that can then allow routine notification data to be adjusted to account for likely under-ascertainment (46).

We report a wide variety of interventions in the studies included in this systematic review. For HIV (29), TB (24,29,31) and leprosy (25,32,33), studies tended to focus on community screening interventions in suspected geographic hotspots. Malaria studies (34–36), however, were more focused on complex interventions that combined community screening and mass drug administration, with environmental measures to address the vector, including larviciding of stagnant water bodies and indoor residual spraying.

Despite the above reservations, the studies that we have identified demonstrate the feasibility of the spatially targeted approach to direct resources towards community-based interventions. For case-finding interventions or community screening, spatially targeted interventions are a compromise between tracing known contacts of infectious cases and ACF interventions across the entire general populations or sub-population (i.e. whole urban slums) (47). One limitation of contact tracing is that it misses out cases that are unknown to the index case, and so will fail to identify cases arising from transmission to casual contacts outside of the household (48,49). Spatially targeted interventions become increasingly important for diseases that are approaching elimination, since these conditions predispose the disease cases to cluster into disease hotspots (13–19). Hence spatially targeted interventions can help the most successful control programmes to systematically identify and address residual disease hotspots (12,19,32,50). Among all of the studies reviewed here, only one (Bousema et al.) (36) was assessed as having low risk of bias. This was a malaria intervention that evaluated the impact of a spatially targeted intervention in pre-defined inner and outer regions of hotspots (evaluation zones). This study design allows the evaluation of how targeting the hotspots benefits the individuals that are in the hotspot as well as the individuals that are in the outer regions of the hotspot. This in turn allows assessment of the impact of the intervention on reducing transmission. For vector borne

diseases, including malaria, consideration needs to be given to the mobility of the vectors themselves when considering spatially targeted interventions (51) i.e., if mosquitoes are able to travel over moderate distances the effect of a hotspot targeted intervention might end up being diluted.

However, this study had several limitations (36). It was underpowered to detect small changes due to the hotspot targeted intervention. In addition, the intervention was also done in one transmission season which might not have been enough to interrupt the epidemic (36). The area of study might also not have been suitable for a hotspot targeted intervention because even though malaria was heterogeneous there might have been widespread transmission of malaria going on in the non-hotspot areas; this might have meant that a non-targeted approach would have been more appropriate for this setting (52).

Overall, our systematic review's limitation was that the studies we identified were heterogeneous, and we were unable to do a meta-analysis as planned. Also, the studies' methodological quality had challenges due to the lack of validation of disease hotspots and the lack of random allocation of clusters to interventions. We also only included manuscripts in English due to resource limitations, and only four diseases were included. Despite these limitations, we were able to show the potential that spatially targeted interventions have towards improving disease epidemiology.

Recommendations

In general spatially-targeted interventions are only justifiable in settings where the epidemic is heterogenous and secondly where the non-hotspot areas cannot sustain sufficient transmission of the disease (36). The influence of the hotspots on the epidemic of the

disease in the surrounding communities also depends on the effective contact rates between the residents of the hotspots and the surrounding areas; in the case of malaria this also depends on how far mosquitoes can travel between the hotspots and the surrounding communities. Careful attention needs also to be given to identifying hotspots to make sure that the hotspots capture the true burden of the disease by making sure that the hotspots are persistent in time and are adjusted for confounding (51–53).

Conclusion

The rapid increase in cheap, reliable, geolocation technology now provides such a significant tool in the fight to control and eradicate endemic and epidemic infectious diseases that support for rapid development of effective patient mapping systems should be considered essential for health systems in low-income countries. Ideally, mapping should be combined with statistical modelling to identify hotspots with the most pressing need for intervention. In this systematic review we provide some evidence in support of the effectiveness of spatially targeted intervention, but with the major conclusion being that the current body of evidence is weak. There is an urgent need to define optimal methodology, allowing recommendations to be made to support the design of more rigorous studies that allow clear evaluation of likely impact to be made for the major categories of infectious diseases, including vector-borne and those transmitted person-to-person. Recommendations should include strategies for timely identification of hotspots and the critical aspects needed to measure the effect of intervention strategies targeting those hotspots.

Figure 3.1: Research Overview

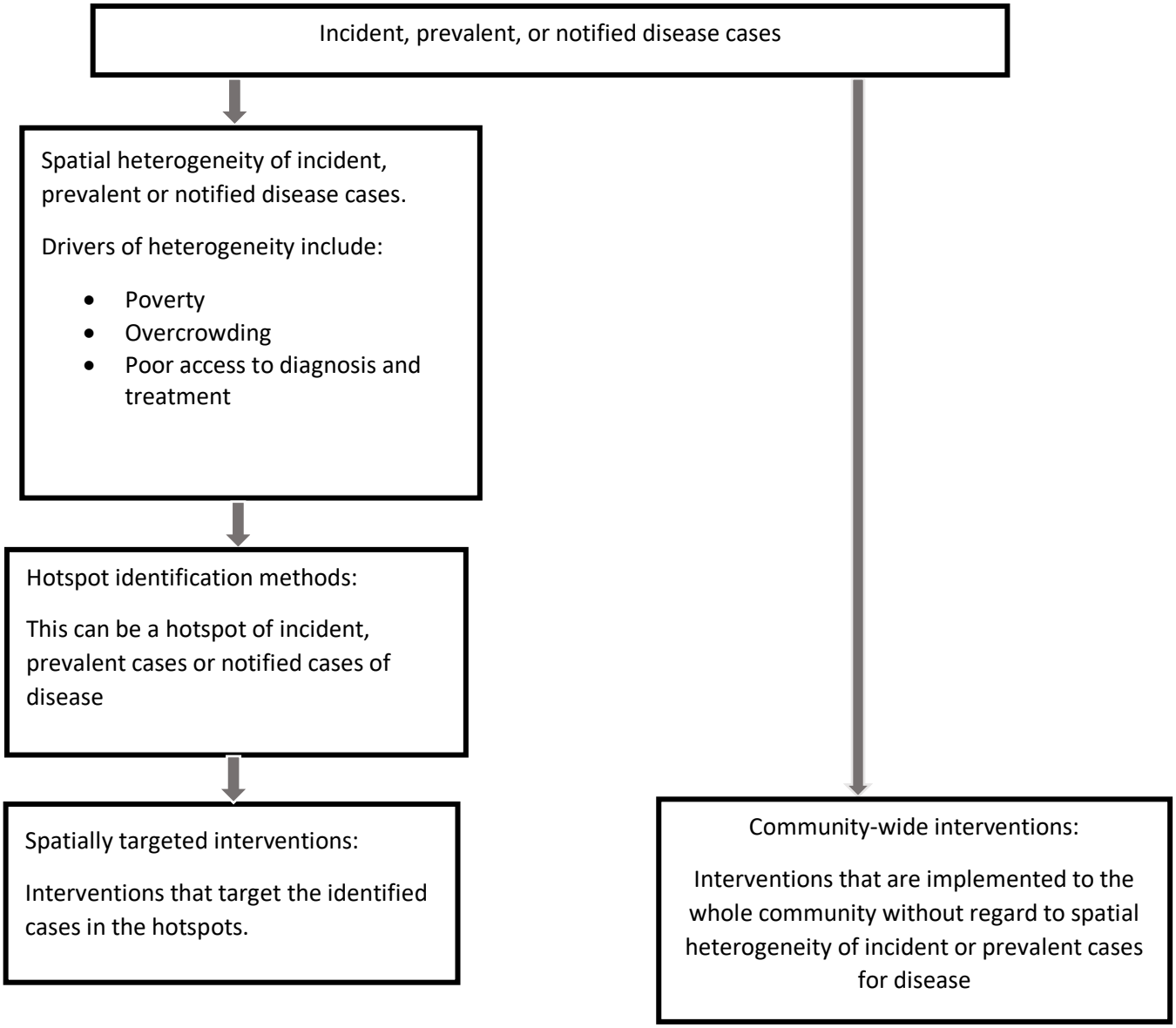


Figure 3.2: PRISMA flow diagram for the process of selecting included studies

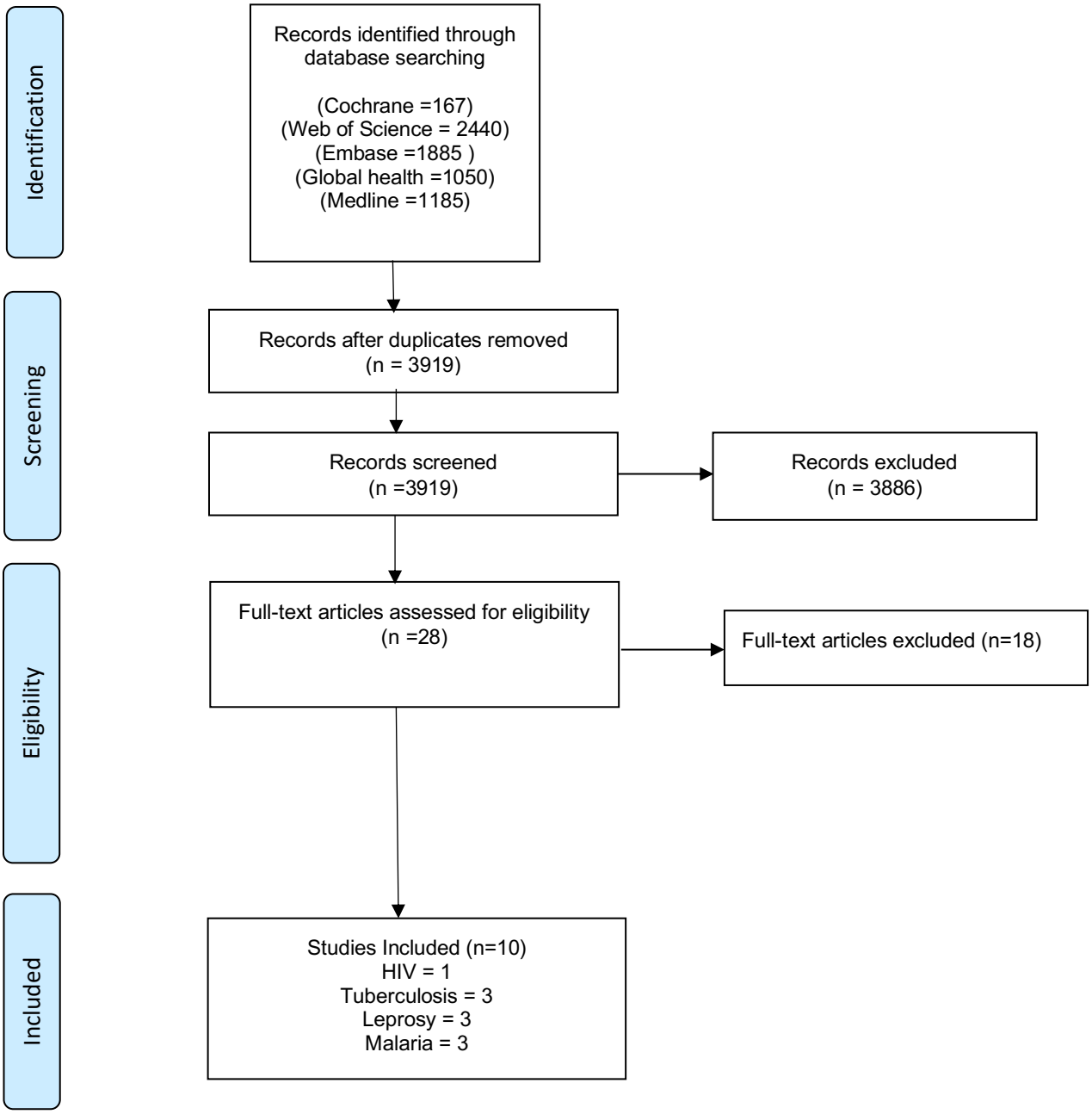


Table 3.1: Search Strategy

1	(HIV or human immunodeficiency virus or HIV-1 or HIV-2).mp.
2	exp HIV/
3	exp HIV Infections/
4	(aids or acquired immune deficiency syndrome).mp.
5	exp Acquired Immunodeficiency Syndrome/
6	(tuberculosis or TB).mp.
7	exp Tuberculosis/
8	exp Mycobacterium tuberculosis/
9	(malaria* or Plasmodium).mp.
10	exp Malaria/
11	exp Plasmodium/
12	(Leprosy or lepra or lepromatosis or hanseniase).mp.
13	exp Leprosy/
14	exp Mycobacterium leprae/
15	(spatial adj3 (analys?s or regression or temporal or autocorrelation* or auto-correlation* or statistics or epidemiology)).mp.
16	exp spatial analysis/
17	(geographic* adj3 (analys?s or regression or temporal or autocorrelation* or auto-correlation* or system* or epidemiology)).mp.
18	(GPS or global position* system* or GIS or global information system* or space?time or geospatial or hotspot* or hot-spot*).mp.
19	exp Geographic Information Systems/
20	(communit* or neighbo?rhood*).mp.
21	exp Residence Characteristics/
22	(intervention* or target*).mp.
23	((high-burden or highburden) adj3 (area* or region* or tract* or setting*)).mp.
24	(high-incidence adj3 (area* or region* or tract* or setting*)).mp.
25	(high-prevalence adj3 (area* or region* or tract* or setting*)).mp.
26	((hot-spot* or hotspot*) adj3 (area* or region* or tract*)).mp
27	1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14
28	15 or 16 or 17 or 18 or 19
29	20 or 21 or 22 or 23 or 24 or 25 or 26
30	27 and 28 and 29
31	limit 30 to english language
32	limit 31 to yr="1993 -Current"

Table 3.2: Reasons for exclusion of studies

Reference, year	Reason for excluding
HIV (three studies)	
Odek 2014 (1)	HIV hotspots not based on geolocated cases
Ikpeazu 2014 (2)	HIV hotspots not based on geolocated cases
Mburu 2017 (3)	HIV hotspots not based on geolocated cases
Tuberculosis (four studies)	
Clark 1999 (4)	Review
Wilkinson 1999 (5)	No hotspots based on geolocated TB cases.
Fatima 2016 (6)	Contact tracing study
Vo 2020 (7)	Hotspots were not based on TB cases
Malaria (eleven studies)	
Mnzava 2001 (8)	Cluster randomised trial with no standard of care arm.
Ghosh 2007 (9)	Review
Dongus 2007 (10)	Hotspot not based on geolocated malaria cases
Cook 2007 (11)	No spatially targeted intervention was evaluated
Shiff 2012 (12)	No spatially targeted intervention was evaluated
Marston 2014 (13)	Health economics study
Cook 2014 (14)	Hotspot not based on geolocated malaria cases
Hsiang 2019 (15)	Defining optimal strategies for contact tracing strategies
Ousmane 2019 (16)	No clear definition of hotspot and study evaluation not based on hotspots.
Ngo 2019 (17)	Feasibility study of acceptability of geographic information technology
Bhondoekhan 2020 (18)	Contact tracing study

Table 3.3: Full description of the characteristics of spatially targeted interventions of HIV, TB, Leprosy and Malaria

Reference, year	Setting	Study type	Intervention	Control
HIV (one study)				
Goswami2012(19)	Wake County, North Carolina USA	Implementation study	Community screening of HIV at sites in the hotspot areas by community nurses and disease intervention specialists from the HIV, syphilis and TB clinics at the county health department.	HIV screening of patients going to STD clinic outside of hotspot areas in the same county.
Tuberculosis (three studies)				
Moonan, 2006(20)	Tarrant County, north central Texas, USA	Implementation study	Community tuberculin skin testing and active TB screening by community organisations in hotspots (period 2002-2004)	No control or comparison group.
Goswami2012(19)	Wake County, North Carolina USA	Implementation study	Community screening of latent TB at sites in the hotspot areas by community nurses and disease intervention specialists from the HIV, Syphilis and TB clinics at the county health department.	Screening latent TB in patients going to a TB clinic outside of hotspot areas in the same county.
Cegielski2013 (21)	Smith County, Texas USA	Implementation study	Community screening of tuberculin skin testing and treatment of diagnosed in the hotspots (1996)	No control or comparison group
Leprosy (three studies)				

DE Souza Dias2007(22)	Municipality of Mossoro, Rio Grande do Norte, Brazil	Implementation study	Study team members moved door to door in the hotspots to identify people with symptoms of leprosy and referred them to the nearest primary health clinic for diagnosis (period 2005)	No control group
Jim2010(23)	state of Pohnpei in the Federated State of Micronesia.	Implementation study	Door to door community Screening of Leprosy cases in the hotspots (2007 to 2009)	No control group
Barreto2015(24)	Municipalities of Castanhal and Oriximinal in the state of Para, Brazil	Implementation study	Community leprosy case screening in two public schools located in the hotspot	Prevalence of leprosy in children from randomly selected schools not from the hotspot areas.
Malaria (three studies)				
Srivastava2009(25)	central Indian state of Madhya Predesh	Implementation study	Not specifically listed. (2007)	No control

Herdiana2013(26)	Island district of Sabang in Indonesia	Implementation study	Twice a month visit households. At each visit screen for malaria from household members with fever or recent history of fever. Treat all confirmed malaria cases.(May 2010 onwards)	Once a month to households in the hotspots. At each visit screen for malaria from household members with fever or recent history of fever. Treat all confirmed malaria cases. (May 2010 onwards)
Bousema2016(27)	Rachuonyo, a western district in Kenya	Cluster randomised trial	5 intervention cluster. Weekly larviciding of stagnant water bodies, long lasting treated nets, Indoor residual spraying and mass drug administration to households with confirmed malaria case.(2012)	5 cluster control areas. Annual residual spraying, distribution of long lasting treated nets at antenatal clinics. No drug mass administration

Abbreviations: HIV Human Immunodeficiency Virus, STD Sexually Transmitted Disease,TB Tuberculosis.

Table 3.4: Characteristics of spatially targeted interventions for HIV, TB, Leprosy and Malaria

Reference, year	Outcome	Effect of intervention[1]	Risk of bias assessment
HIV			
Goswami 2012 (29)	Comparison of case detection yield between hotspot intervention areas and a county STD clinic over the same period (2009-2011)	HIV prevalence was higher among community screened participants (8/240, 3%, 95% CI:1.4-6.5) compared to the Wake County STD clinic (64/15936, 0.4%, 95%CI:0.3-0.5) with a risk ratio of (8.3, 95% CI: 4.0-17.1), p<0.001.	Moderate
Tuberculosis			
Moonan2006 (24)	Yield of TB case detection in hotspots.	Targeted screening identified one person with TB for every 83 screened and one person with LTBI for every five screened. The yield of the targeted approach was considered to be more than what would be expected in a county with an active TB notification rate of 5.7 per 100,000 population year.	Critical
Goswami 2012 (29)	Comparison of case detection yield of LTBI between hotspot areas and county TB clinic over the same period (2009-2011)	LTBI prevalence was higher among community screened participants (36/234, 15%, 95% CI:11.0-21.7) versus (541/9024, 6%, 95%CI=5.6-6.6) at the TB clinic with a risk ratio of (2.5, 95% CI: 1.9-3.5), p<0.001.	Moderate
Cegielski 2013 (31)	A before and after intervention comparison of mapped TB notification rates between 1985-1995 and 1996-2006	TB notification rates in the targeted hotspots declined from 39.6 per 100,0000 people per year (95% CI:30.4-48.8) from 1985 to 1995 to zero from 1996 to 2006 (p<0.001)	Serious
Leprosy			
De Souza Dias 2007 (32)	Percent of notified cases attributable to the intervention, and before and after intervention case notification rate comparison.	Active case finding identified 50% of the total cases that were diagnosed in 2005. The case notification rate in 2005 was higher compared to pre	Serious

		intervention year 2004, 9.34 per 10000 versus 5.16 per 10000 respectively.	
Jim 2010 (33)	Yield of leprosy case detection during intervention period compared to the pre-intervention period and reduction in households that needed to be screened.	Eight-fold decrease in the number of households that needed to be screened from 2007 to 2009. While still identifying a similar number of new cases to pre-spatially targeted active case finding period 2002 to 2006.	Moderate
Barreto 2015 (25)	The yield of case detection of leprosy cases in school children in hotspot intervention schools versus in children from randomly selected schools.	In the hotspot school's (11/134, 8.2%, 95% CI:3.5-13.0) students with a mean age of 10 years were diagnosed with leprosy. While (63/1592, 3.9%, 95% CI:3.0-4.9) students from randomly selected schools with a mean age of 12 years were diagnosed with leprosy with a risk ratio of (2.1, 95% CI: 1.1-3.8), $p<0.05$.	Moderate
Malaria			
Srivastava 2009 (34)	Difference in absolute numbers of notified malaria cases between 2006 and 2007.	An absolute reduction in numbers of notified cases in 2007 (N = 90829) from the notified cases in 2006 (N = 96042), (5.7%, 95% CI:4.7-6.7), $p < 0.05$.	Critical
Herdiana2013 (35)	Change of malaria notification rates from pre-intervention to post-intervention period.	30-fold reduction in malaria notifications from 3.83 per 1000 in 2008 to 0.13 per 1000 in 2011.	Serious
Bousema2016 (36)	Change in parasite prevalence in the evaluation zones (1-500m from hotspots) of intervention clusters versus control clusters	The first evaluation zone 1-249m at 8^{th} week (3.6%, 95% CI: -2.6 to 9.7%), p=0.216 and the second evaluation zone 250-500m at 8^{th} week (3.8%, 95% CI: -2.4 to 10.0%), p=0.187. The first evaluation zone 1-249m at 16^{th}	Low

		week (1.0%, 95% CI: -7.0 to 9.1%), p=0.713 and the second evaluation zone 250-500m at 16th week (1.0%, 95% CI: -8.3 to 10.4%), p=0.809	

[1] Results calculated from the data published in the papers.

Abbreviations: TB Tuberculosis, LTBI Latent Tuberculosis Infection, m Metres, N Number, STD Sexually Transmitted Disease.

Table 3.5: Geolocation of cases and hotspot identification

Reference, year	Geolocation of cases	Hotspot identification
HIV (one study)		
Goswami2012 (29)	Notified cases of either Tuberculosis(TB) (N=150), HIV (N=665) or syphilis (155) between 1 Jan 2005 and 31 Dec 2007 were geolocated to households. The method for geolocation of the cases was not described in the paper.	A kernel density map of the cases was produced. Areas with the highest densities of three diseases of HIV, Syphilis and TB (greater than 10 cases per square mile) were classified as hotspots. Two hotspot neighbourhoods were identified in the county.
Tuberculosis (three studies)		
Moonan, 2006 (24)	Notified TB cases (N=991) from 1 Jan 1993 to 31 Dec 2000 in Tarrant County, north central Texas, USA were geolocated to zip codes using residential addresses and zip codes that patients gave at the time of diagnosis with the aid of a GIS software.	Areas with the highest TB notification rates and high percentage of genotypically clustered TB isolates were identified as hotspots. Three neighbourhood hotspots were identified.
Goswami2012 (29)	Notified cases of TB (N=150), HIV (N=665) and syphilis (N= 155) that were notified between 1 Jan 2005 and 31 Dec 2007 were geolocated to households. The method for geolocation of the households was not described in the paper	A kernel density map was developed. Areas with the highest densities of three diseases of HIV, Syphilis and TB (greater than 10 cases per square mile) were classified as hotspots. Two hotspot neighbourhoods were identified in the county.
Cegielski 2013 (31)	Notified TB cases between 1985 to 1995 (N=128) and all notified LTBI from 1993 to 1995 (N=311) were geocoded to their households using the addresses that patients gave at the time of diagnosis. In addition, field workers tracked addresses to households to get household coordinates of addresses that failed to geolocate.	The points of cases were plotted on a map and areas with the densest clusters of points of cases were identified as hotspots, 2 neighbourhoods were identified in the county.

Leprosy (three studies)		
DE Souza Dias2007 (32)	Notified leprosy cases that occurred between 1998 and 2002 (N=368) in the municipality of geocoded to households. The method for geolocation of cases was not described.	Density map with a radius of 100m of the notified leprosy cases was produced. Four hotspot areas were identified
Jim2010(33)	Notified leprosy cases from 2002 to 2006 (N=502) were geolocated to households. Field workers visited all notified cases to get household GPS coordinates using a GIS device.	A density map based on 1-mile radius of the notified leprosy cases. Areas with high concentration of cases classified as hotspots.
Barreto2015 (25)	Notified leprosy cases from January 2004 to February 2010 (n=633) were geocoded to households. Field workers visited households of registered cases to collect GPS coordinates.	Hotspots were identified using the Kulldorff's spatial scan statistic and by stratification of the leprosy notified rates. Two hotspots were identified.
Malaria (three studies)		
Srivastava2009 (34)	Notified malaria cases between 2000 to 2005 were obtained from the State Department of Health based on the cases notified in clinics in the blocks or districts.	Blocks or districts with a percentage of plasmodium falciparum malaria cases of all notified cases that was either 100% or consistently greater than 30% from 2000-2005 or greater than 70% in 2005
Herdiana2013 (35)	Notified malaria cases from 2007 to 2008 in addition to other self-reported malaria cases that were found during a survey (n=319) were geocoded to households. Field workers obtained the GPS coordinates of households using GIS devices.	Villages that had the majority of malaria cases were classified as hotspots. 14 out of 18 villages were identified
Bousema2016 (36)	June and July 2011, 17,503 individuals tested in a malaria prevalence survey for the prevalence of P. falciparum antibodies (AMA-1 or MSP-10). Field workers collected the GPS coordinates of the households using GPS devices.	Segments of the study area were scanned in the 2 x 4 km rolling windows and areas with higher ($p < 0.05$) prevalence of antibodies and age-adjusted antibody density than the local average values were identified as hotspots.

Abbreviations: GIS Geographical Information System, GPS Global Positioning System, HIV Human Immunodeficiency Virus, m Metres, N Number, TB Tuberculosis, STD Sexually Transmitted Disease.

Declarations

Ethics approval and consent to participate

This review used published data, and ethical review was not required.

Consent for publication

Not applicable

Availability of data and materials

The search strategy and summary data tables for this systematic review have all been included in the main text and in the appendix section.

Competing interests

There are no conflicts of interest to declare.

Funding

This work was supported by Overseas Collaborative Site Staff-Studentship included as part of Professor Elizabeth Corbett's (ELC) Senior Research Fellowship in Clinical Science from Wellcome Trust ("Sustainable Community Action for Lung hEalth (SCALE): a cluster-randomised trial in Blantyre, Malawi." WT200901/Z/16/Z.). The funders had no role in study design; in the collection, analysis, and interpretation of data; in the writing of the report; and in the decision to submit the paper for publication. Peter MacPherson (PM) is funded by Wellcome Trust (206575)

Authors' contributions

MK, PM, ELC, and James Carpenter (JC) designed the study. MK, PM, JC and MN contributed to the data collection and management. MK, PM, JC and MN contributed to the data analysis. MK, PM, ELC, JC, Ted Cohen (TC) and MN contributed to the data interpretation. MK wrote the first draft. MK, PM, ELC, JC, TC and MN reviewed, contributed to, and approved the final draft. All authors read and approved the final manuscript

Acknowledgements

We would like to thank the LSHTM library staff for assisting us with the formulation of the search strategy.

References

1. United Nations Joint Programme on HIV/AIDS (UNAIDS). UNAIDS data 2018 [Internet]. Geneva; 2018. Available from: http://www.unaids.org/sites/default/files/media_asset/unaids-data-2018_en.pdf
2. World Health Organisation. Global tuberculosis report 2018. Geneva; 2018.
3. World Health Oganisation. Weekly leprosy epidemiological report 31AUG2018 [Internet]. Geneva; 2018 [cited 2019 Feb 1]. Available from: http://apps.who.int/iris/bitstream/hand
4. World Health Organisation. World malaria report 2018 [Internet]. Geneva; 2018 [cited 2019 Feb 1]. Available from: www.who.int/malaria
5. World Health Organisation. Global leprosy strategy 2016-2020. Geneva; 2015.
6. World Health Organisation. Frequently asked questions on leprosy [Internet]. Geneva; 2013. Available from: http://apps.searo.who.int/PDS_DOCS/B5044.pdf
7. Almeida ACG, Kuehn A, Castro AJM, Vitor-Silva S, Figueiredo EFG, Brasil LW, et al. High proportions of asymptomatic and submicroscopic Plasmodium vivax infections in a peri-urban area of low transmission in the Brazilian Amazon. Parasit Vectors [Internet]. 2018 Dec 20 [cited 2019 Feb 1];11(1):194. Available from: http://www.ncbi.nlm.nih.gov/pubmed/29558985
8. Ngwira LG, Dowdy DW, Khundi M, Barnes GL, Nkhoma A, Choko AT, et al. Delay in seeking care for tuberculosis symptoms among adults newly diagnosed with HIV in rural Malawi. INT J TUBERC LUNG DIS [Internet]. 2018 Mar 1 [cited 2019 Feb 1];22(3):280–6. Available from: http://dx.doi.org/10.5588/ijtld.17.0539
9. Fomundam HN, Tesfay AR, Mushipe SA, Mosina MB, Boshielo CT, Nyambi HT, et al. Prevalence and predictors of late presentation for HIV care in South Africa. S Afr Med J [Internet]. 2017 Nov 27 [cited 2019 Feb 1];107(12):1058–64. Available from: http://www.ncbi.nlm.nih.gov/pubmed/29262956
10. Sulat JS, Prabandari YS, Sanusi R, Hapsari ED, Santoso B. The impacts of community-based HIV testing and counselling on testing uptake. J Heal Res [Internet]. 2018;JHR-01-2018-015. Available from: http://www.emeraldinsight.com/doi/10.1108/JHR-01-2018-

015

11. Ho J, Fox GJ, Marais BJ. Passive case finding for tuberculosis is not enough. Int J Mycobacteriology [Internet]. 2016 Dec 1 [cited 2019 Mar 7];5(4):374–8. Available from: https://www.sciencedirect.com/science/article/pii/S2212553116301595?via%3Dihub
12. World Health Organisation. A framework for malaria elimination global Malaria programme [Internet]. Geneva; 2017 [cited 2019 Mar 13]. Available from: https://apps.who.int/iris/bitstream/handle/10665/254761/9789241511988-eng.pdf;jsessionid=FEBA2DD0A83E29D8833B238DAD060E96?sequence=1
13. Brian J. Coburn, Justin T. Okano SB, Coburn BJ, Okano JT, Blower S. Using geospatial mapping to design HIV elimination strategies for sub-Saharan Africa. Sci Transl Med [Internet]. 2017;9(383 PG-). Available from: NS -
14. Cuadros DF, Li J, Branscum AJ, Akullian A, Jia P, Mziray EN, et al. Mapping the spatial variability of HIV infection in Sub-Saharan Africa: Effective information for localized HIV prevention and control. Sci Rep. 2017;7(1):1–11.
15. Shaweno D, Karmakar M, Alene KA, Ragonnet R, Clements AC, Trauer JM, et al. Methods used in the spatial analysis of tuberculosis epidemiology: a systematic review. BMC Med [Internet]. 2018 [cited 2019 Feb 20];16(1):193. Available from: https://bmcmedicine.biomedcentral.com/articles/10.1186/s12916-018-1178-4
16. Monteiro LD, Martins-Melo FR, Brito AL, Alencar CH, Heukelbach J. Spatial patterns of leprosy in a hyperendemic state in Northern Brazil, 2001-2012. Rev Saude Publica. 2015;49.
17. Alencar CH, Ramos AN, dos Santos ES, Richter J, Heukelbach J. Clusters of leprosy transmission and of late diagnosis in a highly endemic area in Brazil: Focus on different spatial analysis approaches. Trop Med Int Heal. 2012;17(4):518–25.
18. Sluydts V, Heng S, Coosemans M, Van Roey K, Gryseels C, Canier L, et al. Spatial clustering and risk factors of malaria infections in Ratanakiri Province, Cambodia. Malar J [Internet]. 2014 Dec 30 [cited 2019 Feb 1];13(101139802):387. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med10&NEWS=N&AN=25269827

19. Shaweno D, Trauer JM, Denholm JT, McBryde ES. The role of geospatial hotspots in the spatial spread of tuberculosis in rural Ethiopia: a mathematical model. R Soc open Sci [Internet]. 2018;5(9):180887. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=prem&NEWS=N&AN=30839742
20. Lessler J, Azman AS, McKay HS, Moore SM. What is a Hotspot Anyway? Am J Trop Med Hyg [Internet]. 2017 Jun [cited 2019 Feb 18];96(6):1270–3. Available from: http://www.ncbi.nlm.nih.gov/pubmed/28719289
21. Shrestha S, Chatterjee S, Rao KD, Dowdy DW. Potential impact of spatially targeted adult tuberculosis vaccine in Gujarat, India. J R Soc Interface [Internet]. 2016;13(116). Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med12&NEWS=N&AN=27009179
22. Lonnroth, Knut. Systematic screening for active tuberculosis Principles and recommendations [Internet]. [cited 2019 May 27]. Available from: https://www.who.int/tb/publications/Final_TB_Screening_guidelines.pdf
23. United Nations Human Settlements Programme. SLUM ALMANAC 2015 2016 Tracking Improvement in the Lives of Slum Dweller [Internet]. Nairobi; 2015. Available from: https://unhabitat.org/slum-almanac-2015-2016/
24. Moonan PK, Oppong J, Sahbazian B, Singh KP, Sandhu R, Drewyer G, et al. What is the outcome of targeted tuberculosis screening based on universal genotyping and location? Am J Respir Crit Care Med [Internet]. 2006;174(5):599–604. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med5&NEWS=N&AN=16728707
25. Barreto JG, Bisanzio D, Frade MAC, Moraes TMP, Gobbo AR, Guimarães L de S, et al. Spatial epidemiology and serologic cohorts increase the early detection of leprosy. BMC Infect Dis [Internet]. 2015 Nov 16 [cited 2019 Mar 12];15(1):1–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26573912
26. Wold Health Organisation. Systematic screening for active tuberculosis. Geneva; 2013.

27. Higgins JPT, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ. 2011 Oct 29;343(7829).
28. Sterne JA, Hernán MA, Reeves BC, Savović J, Berkman ND, Viswanathan M, et al. ROBINS-I: A tool for assessing risk of bias in non-randomised studies of interventions. BMJ. 2016 Oct 12;355.
29. Goswami ND, Hecker EJ, Vickery C, Ahearn MA, Cox GM, Holland DP, et al. Geographic information system-based screening for TB, HIV, and syphilis (GIS-THIS): a cross-sectional study. Pai M, editor. PLoS One [Internet]. 2012 Oct 2 [cited 2019 Mar 29];7(10):e46029. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med8&NEWS=N&AN=23056227
30. Moonan PK, Bayona M, Quitugua TN, Oppong J, Dunbar D, Jost KCJ, et al. Using GIS technology to identify areas of tuberculosis transmission and incidence. Int J Health Geogr [Internet]. 2004;3(1):23. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=medp&NEWS=N&AN=15479478
31. Cegielski JP, Griffith DE, McGaha PK, Wolfgang M, Robinson CB, Clark PA, et al. Eliminating tuberculosis one neighborhood at a time. Am J Public Health [Internet]. 2013 Jul [cited 2019 Feb 5];103(7):1292–300. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23078465
32. De Souza Dias MCF, Dias GH, Nobre ML, M.C.F. DSD, G.H. D, M.L. N. The use of Geographical Information System (GIS) to improve active leprosy case finding campaigns in the Municipality of Mossoro, Rio Grande do Norte State, Brazil. Lepr Rev [Internet]. 2007;78(3):261–9. Available from: http://www.lepra.org.uk/lr/Sept07/Lep261-269.pdf
33. Jim R, Johnson E, Pavlin BI. Role of GIS technology during leprosy elimination efforts in Pohnpei. Pac Health Dialog [Internet]. 2010;16(1):109–14. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med7&NEWS=N&AN=20968242

34. Srivastava A, Nagpal BN, Joshi PL, Paliwal JC, Dash AP, A. S, et al. Identification of malaria hot spots for focused intervention in tribal state of India: a GIS based approach. Int J Health Geogr [Internet]. 2009;8(101152198):30. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med6&NEWS=N&AN=19457227

35. Herdiana H, Fuad A, Asih PB, Zubaedah S, Arisanti RR, Syafruddin D, et al. Progress toward malaria elimination in sabang, aceh, indonesia. Am J Trop Med Hyg [Internet]. 2012;87(5 SUPPL. 1):47. Available from: http://www.ajtmh.org/content/87/5_Suppl_1/1.1.full.pdf+html

36. Bousema T, Stresman G, Baidjoe AY, Bradley J, Knight P, Stone W, et al. The Impact of Hotspot-Targeted Interventions on Malaria Transmission in Rachuonyo South District in the Western Kenyan Highlands: A Cluster-Randomized Controlled Trial. PLoS Med. 2016;13(4):e1001993–e1001993.

37. Tripepi G, Jager KJ, Dekker FW, Zoccali C. Selection bias and information bias in clinical research. Nephron - Clin Pract. 2010;115(2).

38. Morris U, Msellem MI, Mkali H, Islam A, Aydin-Schmidt B, Jovel I, et al. A cluster randomised controlled trial of two rounds of mass drug administration in Zanzibar, a malaria pre-elimination setting-high coverage and safety, but no significant impact on transmission. BMC Med [Internet]. 2018;16(1):215. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=medl&NEWS=N&AN=30526588

39. Mosha JF, Sturrock HJW, Greenwood B, Sutherland CJ, Gadalla NB, Atwal S, et al. Hot spot or not: a comparison of spatial statistical methods to predict prospective malaria infections. Malar J [Internet]. 2014;13(101139802):53. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med10&NEWS=N&AN=24517452

40. Hsiang MS, Ntshalintshali N, Kang Dufour M-S, Dlamini N, Nhlabathi N, Vilakati S, et al. Active case-finding for malaria: A three-year national evaluation of optimal approaches to detect infections and hotspots through reactive case detection in the low transmission

setting of Eswatini. Clin Infect Dis [Internet]. 2019; Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=medp&NEWS=N&AN=31095677

41. Fatima R, Qadeer E, Yaqoob A, Haq MU, Majumdar SS, Shewade HD, et al. Extending "Contact Tracing" into the Community within a 50-Metre Radius of an Index Tuberculosis Patient Using Xpert MTB/RIF in Urban, Pakistan: Did It Increase Case Detection?. PLoS One [Internet]. 2016;11(11):e0165813. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med12&NEWS=N&AN=27898665
42. Elliott P, Wartenberg D. Spatial epidemiology: Current approaches and future challenges. Vol. 112, Environmental Health Perspectives. Public Health Services, US Dept of Health and Human Services; 2004. p. 998–1006.
43. Dowdy DW, Dye C, Cohen T. Data Needs for Evidence-Based Decisions: A TB Modeler's "Wish List" [cited 2019 Feb 25]; Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4041555/pdf/nihms584157.pdf
44. Cudahy PGT, Andrews JR, Bilinski A, Dowdy DW, Mathema B, Menzies NA, et al. Spatially targeted screening to reduce tuberculosis transmission in high-incidence settings. Lancet Infect Dis [Internet]. 2018 Dec 13 [cited 2019 Feb 16];0(0). Available from: http://www.ncbi.nlm.nih.gov/pubmed/30554997
45. MacPherson P, Khundi M, Nliwasa M, Choko AT, Phiri VK, Webb EL, et al. Disparities in access to diagnosis and care in Blantyre, Malawi, identified through enhanced tuberculosis surveillance and spatial analysis. BMC Med [Internet]. 2019 Dec 29 [cited 2019 Feb 6];17(1):21. Available from: https://bmcmedicine.biomedcentral.com/articles/10.1186/s12916-019-1260-6
46. Borgdorff MW. New measurable indicator for tuberculosis case detection. Emerg Infect Dis [Internet]. 2004 Sep [cited 2019 Feb 6];10(9):1523–8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15498151
47. Department of HIV/AIDS, Stop TB Department, World Health Organization (WHO). Guidelines for intensified tuberculosis case-finding and isoniazid preventive therapy for

people living with HIV in resource-constrained settings [Internet]. Geneva; 20111 [cited 2019 May 24]. Available from: https://apps.who.int/iris/bitstream/handle/10665/44472/9789241500708_eng.pdf;jsessionid=21FF5A6C6174558D73F38D25777B56F0?sequence=1

48. Crampin AC, Glynn JR, Traore H, Yates MD, Mwaungulu L, Mwenebabu M, et al. Tuberculosis transmission attributable to close contacts and HIV status, Malawi. Emerg Infect Dis. 2006;12(5):729–35.

49. Chamie G, Kato-Maeda M, Emperador DM, Wandera B, Mugagga O, Crandall J, et al. Spatial overlap links seemingly unconnected genotype-matched TB cases in rural Uganda. 2018 [cited 2019 Jan 30]; Available from: https://doi.org/10.1371/journal.pone.0192666

50. Busgeeth K, Rivett U. The use of a spatial information system in the management of HIV/AIDS in South Africa. Int J Health Geogr [Internet]. 2004;3(13). Available from: http://www.ij-healthgeographics.com/content/3/1/13/abstract

51. Bousema T, Griffin JT, Sauerwein RW, Smith DL, Churcher TS, Takken W, et al. Hitting hotspots: spatial targeting of malaria for control and elimination. PLoS Med [Internet]. 2012 Jan [cited 2019 Mar 14];9(1):e1001165. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22303287

52. Stresman G, Bousema T, Cook J. Malaria Hotspots: Is There Epidemiological Evidence for Fine-Scale Spatial Targeting of Interventions? Vol. 35, Trends in Parasitology. Elsevier Ltd; 2019. p. 822–34.

53. WHO. Module 2: Screening WHO operational handbook on tuberculosis Systematic screening for tuberculosis disease [Internet]. 2021 [cited 2021 Apr 6]. Available from: https://reliefweb.int/report/world/who-consolidated-guidelines-tuberculosis-module-2-screening-systematic-screening

4 TB Case fatality analysis

4.1 Introduction

In this chapter, I investigated the clinical, health system, and neighbourhood determinants of TB case fatality in urban Blantyre, Malawi. Globally, TB disease was the second leading cause of death from a single infectious agent in 2021. TB treatment reduces the risk of death and helps individuals recover from the disease. Unfortunately, some individuals die while on TB treatment. In this analysis, we set out to investigate the risk factors for death on TB treatment. The identified risk factors could be intervened upon to reduce the risk of death in individuals that are categorised as having a higher risk of death.

In a multi-level spatial regression modelling analysis of TB case fatality rates in Blantyre identified through a citywide enhanced TB surveillance system, we found strong evidence that older age, HIV status, and distance to the TB treatment clinic were all associated with an increased risk of death while on TB treatment. Distance to a health facility is a proxy indicator of how easy it is to access health care. In our study population, distance increased the odds of death only for patients that were registered at the referral clinic but not at a primary health care clinic. This is consistent with the hypothesis that high quality facility-based TB screening and care is complementary to community-based interventions in reducing TB mortality.

This paper was submitted on 9 February 2021 to Epidemiology and Infection Journal and was published on 27 July 2021 and has been described verbatim below.

4.2 The TB case-fatality manuscripts

LONDON SCHOOL of HYGIENE & TROPICAL MEDICINE

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT

T: +44 (0)20 7299 4646
F: +44 (0)20 7299 4656
www.lshtm.ac.uk

RESEARCH PAPER COVER SHEET

Please note that a cover sheet must be completed for each research paper included within a thesis.

SECTION A – Student Details

Student ID Number	lsh1805281	**Title**	Mr
First Name(s)	McEwen Joseph		
Surname/Family Name	Khundi		
Thesis Title	Towards a framework for improved targeting of tuberculosis interventions: a spatial analysis of patient notification and prevalence survey data from Blantyre, Malawi		
Primary Supervisor	James Carpenter		

If the Research Paper has previously been published please complete Section B, if not please move to Section C.

SECTION B – Paper already published

Where was the work published?	Epidemiology and Infection		
When was the work published?	27 July 2021		
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion			
Have you retained the copyright for the work?*	**No**	Was the work subject to academic peer review?	**Yes**

*If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be published?	
Please list the paper's authors in the intended authorship order:	

Stage of publication	Choose an item

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	I came up with the research question, led the analysis, led the writing of the manuscript, and submitted it for publication. I am the first author.

SECTION E

Student Signature	[redacted]
Date	7 July 2022

Supervisor Signature	[redacted]
Date	21 · VII · 2022

Clinical, health systems and neighbourhood determinants of tuberculosis case fatality in urban Blantyre, Malawi: a multilevel epidemiological analysis of enhanced surveillance data

Authors:

*McEwen Khundi[1,2] (MK), Peter MacPherson[1,2,3] (PM), Helena R A Feasey[1,2] (HRAF), Rebeca Nzawa Soko[1,2] (RNS), Marriott Nliwasa (MN)[1,4], Elizabeth L Corbett[1,2] (ELC), James Carpenter[2] (JC)

Affiliations:

1. Malawi-Liverpool-Wellcome Trust Clinical Research Programme, Blantyre, Malawi
2. London School of Hygiene and Tropical Medicine, UK
3. Department of Clinical Sciences, Liverpool School of Tropical Medicine, UK
4. Helse Nord TB Initiative, College of Medicine, University of Malawi

Corresponding author:

McEwen Khundi

Malawi,

mcewenkhundi@gmail.com

Tel: +265 997 615 243

Abstract text word count: 199

Main text word count: 3604

Figures: 3

Tables:4

Abstract

We investigated whether household to clinic distance was a risk factor for death on tuberculosis (TB) treatment in Malawi. Using enhanced TB surveillance data, we recorded all TB treatment initiations and outcomes between 2015 and 2018. Household locations were geolocated, and distances measured by straight line or shortest road network. We constructed Bayesian multi-level logistic regression models to investigate associations between distance and case fatality. A total of 479/4397 (10.9%) TB patients died. Greater distance was associated with higher (odds ratio (OR): 1.07 per kilometre (km) increase, 95% credible interval (CI): 0.99, 1.16) odds of death in TB patients registered at the referral hospital, but not among TB patients registered at primary clinics (OR: 0.98 per km increase, 95%CI: 0.92, 1.03). Age (OR: 1.02 per year increase, 95%CI: 1.01, 1.02), and HIV-positive status (OR: 2.21, 95%CI: 1.73, 2.85), were also associated with higher odds of death. Model estimates were similar for both distance measures. Distance was a risk factor for death among patients at the main referral hospital, likely due to delayed diagnosis and suboptimal healthcare access. To reduce mortality, targeted community TB screening interventions for TB disease and HIV, and expansion of novel sensitive diagnostic tests are required.

Keywords: epidemiology, HIV, multilevel modelling, statistics, treatment outcomes, tuberculosis

Key message

- Tuberculosis (TB) causes more deaths than HIV/AIDS and malaria combined.
- There is limited evidence on whether household to clinic distance is a risk factor for death on TB treatment.
- We used two measures of household to clinic distance (Cartesian distance, and shortest road network distance).
- For both distance measures, household to clinic distance was a risk factor for death among patients treated at the main referral hospital only, likely due to delayed diagnosis and suboptimal healthcare access.

Introduction

Tuberculosis (TB) remains the leading cause of death from a single infectious disease in the world, causing more deaths than HIV/AIDS and malaria combined [1]. Slow progress has been made in reducing TB mortality, and the World Health Organization's (WHO) End TB strategy target of achieving a 90% reduction in TB deaths between 2015 and 2030 [1] is unlikely to be met.

In the WHO Africa (AFRO) region, TB treatment success rates (cured or completed treatment) in 2017 were 86% for HIV-negative people and 78% for HIV-positive people [1]. Despite these improvements from 2007 (when overall TB treatment success was estimated to be 79% in Africa), a considerable fraction of people with TB symptoms are delayed in accessing TB diagnosis and care at health facilities [1,2]. The reasons for delayed access to treatment are complex and multi-layered, but include: health care seeking behaviours, clinical and geographical factors, and suboptimal care quality in health facilities [3].

Distance to health facilities is a well-recognised access barrier to prompt diagnosis [4–7]. Individuals who live far from health facilities are at an increased risk of unfavourable health outcomes [4–6] and at an increased risk of death on TB treatment [8–10]. However, there is limited evidence about associations between distance and death on TB treatment in urban African settings.

Using prospectively collected data from people initiating TB treatment at health facilities in urban Blantyre, Malawi in the era of high antiretroviral therapy coverage for HIV, we hypothesised that people initiating TB care at facilities at greater distances from their homes might be at greater risk of death on TB treatment compared to people who lived nearer to clinics. Additionally, as

accurate measurement of clinic-distance may be challenging under routine programmatic conditions, we compared distance measurement using two approaches that could be used by health planners and epidemiologists: Cartesian distance and shortest road network distance.

Methods

Study setting and design

Blantyre District is located in the southern region of Malawi. Blantyre city, in the centre of the district, has large areas of densely-populated informal settlements. Blantyre's 2018 census population was 1,264,304 [11] and HIV prevalence was 18% [12].

Blantyre enhanced TB monitoring and evaluation

In Malawi, patients diagnosed with TB register to receive treatment at primary health care centres and hospitals. In Blantyre, TB registration clinics include one referral hospital (Queen Elizabeth Hospital Central Hospital [QECH]), one large private church supported clinic (Mlambe), three private clinics and seven government public primary health care clinics [13]. QECH is a tertiary referral hospital for the whole southern region of Malawi and offers inpatient care including TB diagnosis and treatment [13].

In a joint project between the Malawi-Liverpool-Wellcome Trust Clinical Research Programme, Blantyre District Health Office, and the Malawi National TB Control Programme (NTP), TB Officers (a cadre of health worker employed by the Ministry of Health of Malawi) received training to strengthen the TB surveillance system in Blantyre. Since 2015, TB Officers stationed at TB treatment clinics in Blantyre provided TB treatment, HIV testing and linkage to treatment in accordance with Malawi guidelines. They further recorded TB registration data into NTP TB registers and additionally recorded individual-level clinical, sociodemographic and household data using an electronic data collection application (ePaL), which we have previously developed and validated (18). Using ePaL, TB officers obtained global positioning satellite (GPS) coordinates for

the location of TB patients' households (18). A spot sputum sample was also collected from all patients starting TB treatment (where able to provide) that underwent smear microscopy and mycobacteria growth indicator tube culture. On a quarterly basis, study and TB programme registers were reconciled.

TB outcomes were defined according to mutually-exclusive WHO TB treatment outcome definition guidelines [14], with patients classified as either: cured; completed treatment; treatment failed; confirmed died; lost to follow up; and not evaluated (typically because still on treatment). The outcome of interest for this analysis was confirmed death on TB treatment; we did not follow-up patients to confirm death. The analysis was limited to study participants that were registered for treatment from 1st January 2015 to 30th December 2018.

Statistical analysis

Baseline characteristics

We compared the characteristics of participants, recorded at TB treatment registration, between those who died during TB treatment and those who were alive at the end of treatment. For categorical variables we calculated percentages and used the chi-square test for comparisons; for continuous variables, we calculated medians (and interquartile ranges [IQR]), mean (and standard deviation), and compared between groups using the Kruskal-Wallis test.

Distance estimation

We estimated the distance from study participants' households to their TB treatment initiation clinic using two approaches (Figure 4.1). The first approach was to estimate the distance based on as a "straight line" distance (Cartesian distance). In the second approach, we used Blantyre urban road network downloaded from OpenStreetMap (OpenStreetMap Foundation) to calculate the shortest road network distance using the stplanr R package [15].

Statistical models

To estimate the causal relationship between clinic distance and TB case fatality, we constructed a directed acyclic graph (Figure 4.2). The minimum adjustment set of confounders identified from the DAG were: sex, age, HIV, hospital admission (QECH) and household wealth score. We included a term for the interaction between household to clinic distance and the clinic at which TB registration occurred (QECH, the city's referral hospital, versus other primary health care centres). This was because QECH is a large referral hospital known to have a patient population with more advanced TB disease, more complicated TB disease, and patients that are likely to have travelled a longer distance compared to patients that attend other clinics.

Using asset ownership data, we created a household wealth score variable, calculated using a proxy means test developed for urban populations from the Malawi Integrated Household Survey [16]. For individuals for whom components needed to calculate their wealth score were missing, their wealth score was imputed using multiple interval imputation, with the observed wealth score providing the lower bound.

We constructed Bayesian multi-level logistic regression models (Equation 1) to investigate the association between participants' household to clinic distance and their risk of death, adjusting for confounders and an interaction between household to clinic distance and clinic of registration. Weakly regularising priors, refined by inspecting prior predictive plots, were assigned to model intercepts and slopes. Model convergence was assessed by visual inspection of trace plots, effective sample numbers, and Gelman-Rubin statistics.

Equation 1

$$D_{ij} \sim Binomial(1, p_{ij})$$

$$logit(p_{ij}) = \alpha_{clinic\ of\ registration\_j} + \beta_{sex\_ij} + +\beta_{age\_ij} + \beta_{wealthscore\_ij} + \beta_{distance\_ij} + \beta_{QECH\_ij} + \beta_{distance_{ij} * QECH\_ij}$$

PRIORS

$$\beta_{sex} \sim Normal(0, 0.5)$$

$$\beta_{age} \sim Normal(0, 0.5)$$

$$\beta_{hivstatus} \sim Normal(0, 1)$$

$$\beta_{wealthscore} \sim Normal(0, 0.5)$$

$$\beta_{distance} \sim Normal(0, 0.5)$$

$$\beta_{\text{QECH}} \sim Normal(0, 0.5)$$

$$\beta_{distance* \text{ QECH}} \sim Normal(0, 0.5)$$

$$\beta_{clinic\ of\ registration\_j} \sim \alpha + u_j$$

$$\alpha \sim Normal(0, 0.5)$$

$$u_j \sim Normal(0, \sigma_\alpha)$$

$$\sigma_\alpha \sim HalfNormal(0, 0.5)$$

Where D_{ij} indicates death on TB treatment for the i^{th} patient in clinic j, sex indicates sex (male or female), age indicates age in years, hivstatus indicates HIV status (positive or negative), wealthscore indicates household wealth score, distance indicates distance to clinic in kilometres, and QECH indicates whether participant registered for treatment at QECH or the other primary health centres, *distance** QECH is the interaction effect between distance and QECH and clinic of registration indicates registering for treatment at the j^{th} TB treatment registration heath centre.

Four thousand samples with a burn-in of 600 iterations were drawn from each model posterior distribution using Markov Chain Monte Carlo methods implemented within Stan via the brms R package [17]. Posterior means and 95% credible intervals (CI) were calculated on the log-odds scale and were exponentiated to give odds ratios. In sensitivity analysis, we refitted models recoding participants who were lost-to follow-up or transferred out during

TB treatment as having died since it was possible that some of these patients had actually died [18]. Additionally, we restricted analysis to those with microbiologically-confirmed TB. To investigate the predictive accuracy of the two distance estimates we used models that just had distance and the interaction effect of distance and registering for TB treatment at QECH. We plotted the difference in predicted probability of death for the 100 patients with the greatest difference in distance from the two measurement methods. Analysis was conducted using R version 3.5.2 (R Foundation for Statistical Computing, Vienna).

Ethical considerations

Ethical approval was granted by the London School of Hygiene and Tropical Medicine (16228) and the College of Medicine, University of Malawi Research Ethics Committee (P.12/18/2556). Participants provided oral consent to participate in TB surveillance with a waiver for written consent granted by both research ethics committees.

Results

Baseline characteristics

A total of 4461/5199 (85.8%) patients that initiated TB treatment at the 12 study clinics were included in this analysis. 64/4461 (1.4%) had a missing TB outcome and were excluded from the primary analysis (Table 4.s1). Among participants that had treatment outcomes, 479/4397 (10.9%) died while taking TB treatment, and 258/4397 (5.9%) were reported as lost to follow up or transferred out (Table 1)**.** The percentage of participants who died did not substantially differ across the study years (*P* = 0.313).

More deaths were reported among participants who initiated TB treatment at QECH (315/2170, 14.5%) than in the other 11 TB clinics (164/2227, 7.4%, *P* < 0.001) – Table 1. Similar percentages of men (303/2789, 10.9%) and women (176/1608, 10.9%, *P* = 0.934) died during TB treatment**.** The median age of participants who died was older (median: 37 years, IQR: 30.0-45.0) compared to those who were alive at the end of TB treatment (35 years, IQR: 28.0-41.0, *P* < 0.001)**.** A higher proportion of deaths occurred in patients who were HIV-positive (400/2981, 13.4%) compared to HIV-negative (79/1416, 5.6%, *P* < 0.001)**.**

Participants who did not have microbiologically-confirmed TB (318/2382, 13.4%) were considerably more likely to die compared to participants with either sputum smear, Xpert or culture positive disease (161/2015, 8.0%, *P* < 0.001).The median road network household to clinic distance (median: 5.2km, IQR: 3.2, 7.7) was consistently higher than the Cartesian household to clinic distance estimates (median: 3.8km, IQR: 2.1-5.9). Both household to clinic distance estimates showed that patients who died on TB treatment lived further from clinics where they initiated TB treatment than

those that were alive at the end of TB treatment. For the road network distance measure, those who died on TB treatment had greater clinic distances (median: 5.8km, IQR: 4.1, 9.0) compared to participants who were still alive at TB treatment completion (median: 5.1km, IQR: 3.1, 7.5, $P < 0.001$). Likewise, those who died on TB treatment had a longer Cartesian distance (median: 4.4km, IQR: 2.9, 6.8) versus participants who were still alive at the end of TB treatment (median: 3.8km, IQR: 2.0, 5.7, $P < 0.001$). Patients who did not live within geographically-mapped areas of the city were excluded 738/5199 (14.2%) because we did not have data to ascertain their household location, and so could not calculate their household to clinic distance (Table 4.s1).

Unadjusted analysis of the association of household to clinic distance and risk of death

TB patients who lived further from the clinic at which they registered for treatment had higher odds of death compared to those who lived nearer, whether measured by the road network distance method (odds ratio [OR]: 1.05 per km increase, 95%CI: 1.03, 1.08) or the Cartesian distance method (OR: 1.06 per km increase, 95%CI: 1.03, 1.09) – Table 4.2.

Unadjusted analysis for confounders

In unadjusted analysis, each one year increase in age was associated with a 2% increase (OR: 1.02, 95%CI: 1.01, 1.02) in the odds of death. Being HIV-positive was associated with a 2.5-times (OR vs. HIV-negative status: 2.53, 95%CI: 2.0, 3.25) increase in the odds of death on TB treatment. TB patients who registered at QECH had two-times higher (OR: 2.10, 95%CI: 1.74, 2.56) odds of death compared to people registering at the other primary clinics. Each one unit increase in wealth score was associated with a 40% (OR: 1.40, 95% CI: 0.98, 2.04) increase in the odds of death.

Adjusted analysis for household to clinic distance models

Greater household to clinic distance (road network) had a higher odds of death in TB patients registered at QECH (OR: 1.07 per km increase, 95%CI: 0.99, 1.16), but was not associated with odds of death in TB patients registered at other clinics (OR: 0.98 per km increase, 95%CI: 0.92, 1.03). The average risk of death varied between clinics by 6% (95%CI: 4%, 11%) holding categorical variables at their base level and continuous variables at their average value, with a standard deviation of 0.67 (0.32, 1.18) - Table 4.2.

For the model containing Cartesian distance to clinic, model coefficients and uncertainty bounds were very similar to those from the road distance model, with distance from household to clinic of treatment registration associated with higher odds of death in TB patients registered at QECH (OR: 1.09 per km increase, 95%CI: 1.0, 1.21), but was not associated with odds of death on TB treatment in patients registered at other clinics (OR: 0.97 per km increase, 95%CI: 0.89, 1.04).

Comparison of model predicted risk of death among 100 participants with biggest difference between their road network distance and Cartesian distance estimates

The majority of the 100 TB patients with the greatest difference in distance between the two household to clinic distance measurements registered for treatment at QECH or Mlambe private clinic (Figure 4.3). For these participants who had the greatest measurement difference, there was, as expected, a trend of increased difference in probability of death with greater difference between the two distance measures. The greatest differences were seen at Mlambe private clinic, reflecting greater distances travelled by road. However, the maximum difference in predicted probability of death over these 100 patients was only 0.79%.

Sensitivity analysis

When we reclassified TB patients who were lost to follow-up or had transferred out during TB treatment as having died, model estimates were very similar to our primary analysis (Table 4.3). Analysis restricted to participants with microbiologically-confirmed disease did not alter our model estimates (Table 4.4).

Discussion

Our primary hypothesis was that greater household to clinic distance would be associated with increased risk of death for people starting treatment for TB in urban Blantyre, Malawi. The results of the pre-specified unadjusted analysis showed clinic distance was a risk factor for death after starting TB treatment. However, in adjusted analysis based on our causal graph, greater clinic distance was a risk factor for death only for participants initiating care at the tertiary referral hospital (QECH). Treatment initiation at the city's tertiary hospital, older age and HIV-positive status were important predictors of death whilst taking TB treatment. We additionally found moderate variation in the risk of death between treatment clinics that was not explained by TB patient clinical and sociodemographic characteristics, indicating that clinic quality of care might be an important residual determinant of treatment outcomes.

We used two different measures (Cartesian distance, and shortest road network distance) of household to clinic distance that could be applied by public health planners and epidemiologists, recognising that road network distance may be considerably more challenging to estimate under routine programmatic conditions. Our analysis found that models using either distance measurement reached very similar conclusions, and therefore we recommend that in low-resource settings, the simpler measurement (Cartesian distance) should likely be more appropriate for routine programme use.

For TB patients that were treated at the city's tertiary hospital (QECH), greater household to clinic distance might be a proxy for lack of access to quality health care service [8,19]. Individuals who live in the peri-urban areas far from the centre of Blantyre have limited resources for transport,

and may struggle to access timely tertiary-level health care compared to those living nearer to the centre of the city [20]. Delayed presentation for care will then result in admission to hospital with more advanced disease and at a higher risk of death [5,21].

To reduce the high risk of death on TB treatment, prompt diagnosis and treatment for people presenting to health facilities is required [21,22]. At health facilities in high HIV-prevalence settings, TB symptoms are common (up to 60% reporting at least one of the WHO four cardinal TB symptoms) [23]. Therefore, novel screening approaches—including triage testing using high sensitivity initial tests such as chest X-ray or C-reactive protein, followed by highly specific tests such as Xpert for those who triage test positive, require evaluation [5,24].

QECH is a tertiary referral hospital, and most patients treated for TB at QECH will have previously sought diagnosis—likely on multiple occasions before referral—at a primary health centre [25]. Consequently, earlier diagnosis in primary health care could reduce hospital admission and potentially reduce mortality [1]. We need to improve the TB diagnostic capacity of primary health facilities in the outskirts of the city and ensure that primary health care facilities can make prompt referrals for hospital care where required [1,5,8].

In addition there is a need to evaluate community interventions of TB screening in areas further away from the centre of the city (QECH) since prevalent cases identified in community interventions are usually diagnosed at earlier stage of disease [1,5]. Our previous work has shown that notifications decline with greater distance from the centre of Blantyre city; this might be because of lack of access to quality health care, including TB diagnosis, at areas in the periphery of the city [20]. Health promotion activities to promote early treatment seeking, and quality

improvement activities with health workers to support early diagnosis and reduce variation in practice between clinics could contribute to reduced TB case fatality, whilst strengthening universal healthcare provision within health facilities [1,5].

Malawi has made tremendous progress towards achieving the WHO 90-90-90 HIV targets [26]: 90% of individuals with HIV are aware of their HIV status, 79% of individuals with a diagnosis are on antiretroviral therapy (ART) and 72% of individuals on ART treatment have viral suppression [27]. Nevertheless, in this study, HIV-positive status remained strongly associated with increased risk of death during TB treatment. Nearly all TB patients in this study were aware of their HIV-positive status, and 89% were taking ART at TB treatment initiation. Persistently high case fatality for HIV-positive people despite high ART coverage suggests that HIV positive people have severe acute illness [28], have profound delays in TB diagnosis and treatment initiation [29,30], or have virological failure to HIV treatment, which itself confers a high risk of death. In a study among HIV-positive people taking ART who were admitted to hospital in Zomba District, Malawi 32% had virological treatment failure, and resistance to first-line ART drugs was near universal [31]. Improvements in viral load monitoring and rapid treatment changes linked to supportive adherence interventions could reduce the number of patients that present with severe disease at the start of TB treatment and save more lives [31,32]. Additionally, implementation of point-of-care HIV viral load monitoring for people living with HIV who are admitted to hospital could identify treatment failure earlier. Following WHO recommendations, the Malawi National HIV Programme has recommended that all people living with HIV take a Dolutegravir containing regimen, including by switching treatment regimens for those already taking [31,33]. It will be

important to investigate the effect of this change on TB case fatality among patients treated at primary clinics and in hospital (QECH) [33].

A key strength of this study was that we used prospectively collected data at TB registration through our citywide enhanced TB surveillance system. Our routine monitoring and evaluation records have consistently high agreement with national TB treatment registers. We additionally captured TB patients' household GPS coordinates at the point of registration for TB treatment; most previous studies have attempted to retrospectively geolocate patients' households using physical addresses, a method that is prone to error [34,35]. Our analysis of household to clinic distance was analysed at a continuous scale rather than on a transformed categorical scale with arbitrary value cut off points [36,37]. To investigate the causal relationship between household to clinic distance and risk of death, we selected confounding variables for adjustment by constructing a directed acyclic graph [38]. Sensitivity analysis restricted to TB patients with microbiologically-confirmed TB and where patients who were lost to follow-up or transferred out to another clinic did not alter our findings. However, timing of death was not collected meaning survival analysis could not be done. A small number of TB patients may not have been able to accurately geolocate their household, or may have deliberately misidentified their household; monthly quality assurance checking of a 5% random sample of households attempted to mitigate this. According to Table 4.s1, TB patients who registered for treatment who were residents in the areas of the city not mapped by our GPS system had some different characteristics compared to those living in the mapped areas. We excluded TB patients from the unmapped areas from the analysis. Nevertheless, the main objective of the analysis was to predict TB case fatality using data obtained during treatment registration and there is no reason to believe that the underlying

relationships should differ between the TB cases that were included in the analysis versus those that were excluded.

In conclusion, in prespecified multilevel modelling using a citywide enhanced surveillance data linked to our satellite GPS location system, we found that household to clinic distance was associated with increased risk of death for patient starting TB treatment at the city's central hospital. This could be explained by barriers in accessing prompt diagnosis and treatment, and by variable quality of care at primary health facilities. HIV-positive status and older age remain important risk factors for death. Therefore, interventions that improve access to TB diagnosis through community based active case finding, and improved quality of health facility TB screening, and prioritise HIV-positive people with potentially high-levels of viral failure for TB screening and ART optimisation are required to reduce the unacceptably high case fatality.

Funding

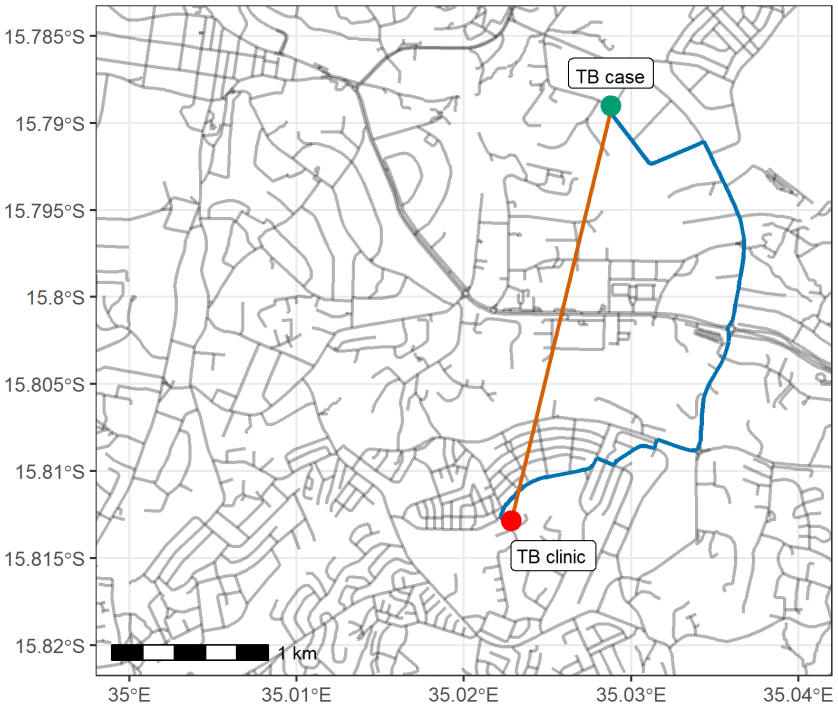
This work was supported by two grants from Wellcome Trust (E.L.C. grant number WT200901/Z/16/Z) and (P.M. grant number 206575).

Conflict of Interests: None

Availability of data

Data will be made available upon request.

Figure 4.1. Illustration of the two methods for measurement of household to clinic distance: Cartesian distance (2.7km), shortest road network distance (4.5km)



Note: The patient place of residence is a randomly generated point for illustration and does not correspond to any patient in the dataset.

Figure 4.2. Directed Acyclic Graph (Causal Diagram)

Illustrating the relationship between distance to TB clinic and risk of death on TB treatment and other covariates. The variables sex, age, HIV, Queens Elizabeth hospital registration (QECH) versus registration at other clinics and poverty were selected as the minimum adjustment set.

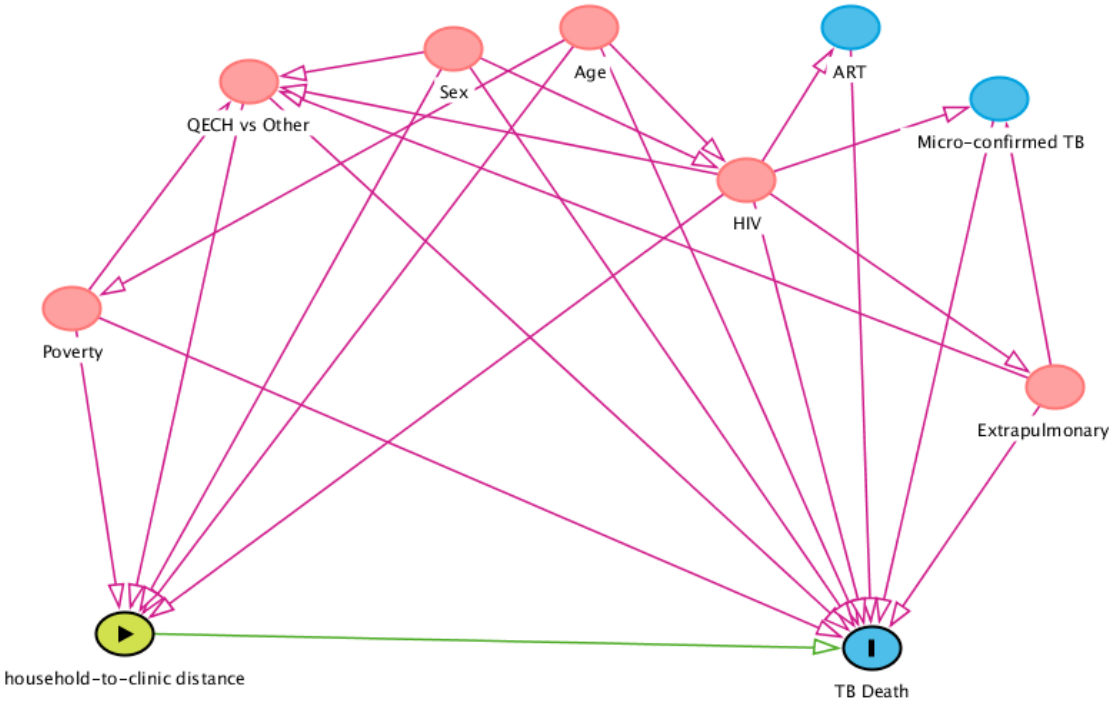
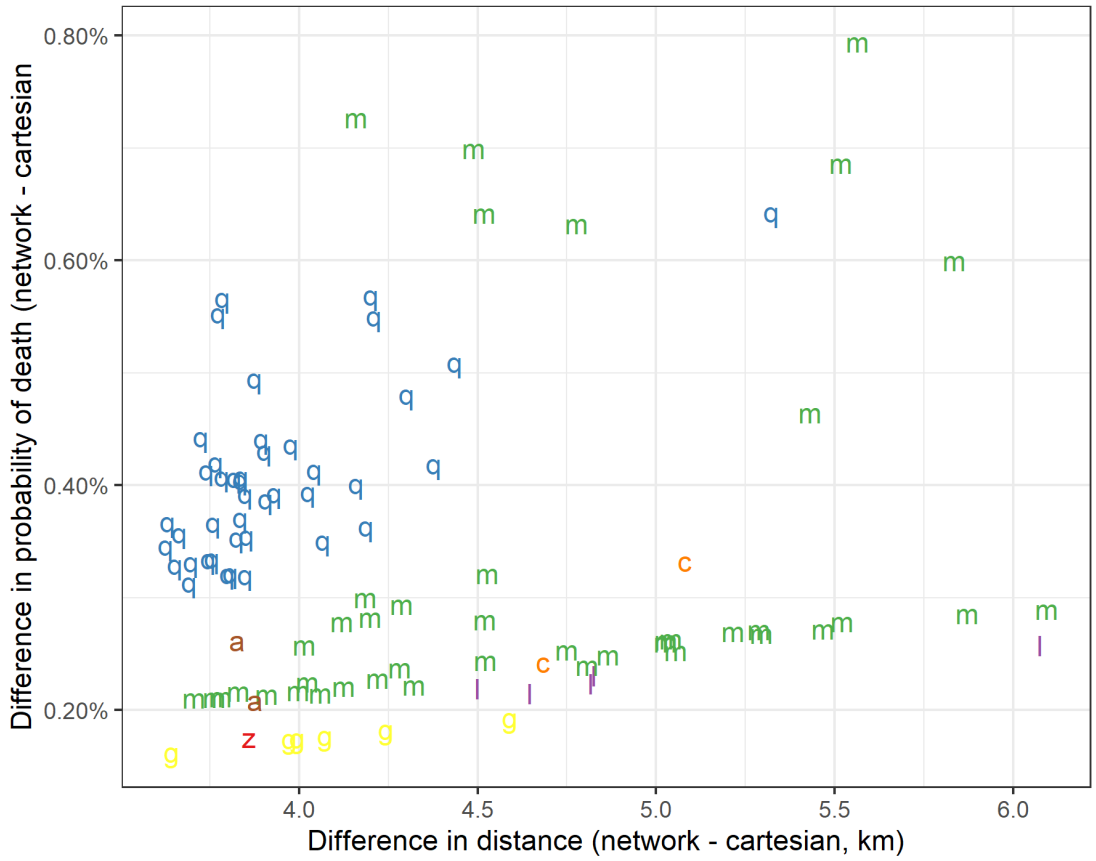


Figure 4.3. Plot of difference between network and Cartesian distance, versus fitted probability of death, for the 100 notified TB cases with the largest distance differences



Abbreviations: q, Queen Elizabeth Central Hospital; m, Mlambe Private Hospital; l, Limbe Health Centre; g, Bangwe Health Centre; z, Zingwangwa Health Centre; a, Blantyre Adventist Hospital; c, Chilomoni Health Centre

Table 4.1: Characteristics for notified Tuberculosis cases in Urban Blantyre Malawi, 2015-2018

	Alive (N=3918)	Died (N=479)	Total (N=4397)	P-value
Year of registration				0.313
2015	860 (89.1%)	105 (10.9%)	965 (100.0%)	
2016	1187 (89.5%)	139 (10.5%)	1326 (100.0%)	
2017	1238 (88.0%)	169 (12.0%)	1407 (100.0%)	
2018	633 (90.6%)	66 (9.4%)	699 (100.0%)	
Gender				0.934
Female	1432 (89.1%)	176 (10.9%)	1608 (100.0%)	
Male	2486 (89.1%)	303 (10.9%)	2789 (100.0%)	
Age, years				< 0.001
Median (IQR)	35.0 (28.0,	37.0 (30.0,	35.0 (28.0,	
Mean (SD)	35.1 (13.4)	38.3 (13.7)	35.4 (13.4)	
HIV status				< 0.001
Negative	1337 (94.4%)	79 (5.6%)	1416 (100.0%)	
Positive	2581 (86.6%)	400 (13.4%)	2981 (100.0%)	
TB classification				< 0.001
Extrapulmonary TB	1375 (85.5%)	233 (14.5%)	1608 (100.0%)	
Pulmonary TB	2543 (91.2%)	246 (8.8%)	2789 (100.0%)	
Micro confirmed TB				< 0.001
Not microbiologically - confirmed TB	2064 (86.6%)	318 (13.4%)	2382 (100.0%)	
Microbiologically-confirmed TB	1854 (92.0%)	161 (8.0%)	2015 (100.0%)	
Distance Km (Cartesian)				< 0.001
Median (IQR)	3.8 (2.0, 5.7)	4.4 (2.9,6.8)	3.8 (2.1, 5.9)	
Mean (SD)	4.3 (3.1)	4.9 (3.1)	4.3 (3.1)	
Distance Km (Road network)				< 0.001
Median (IQR)	5.1 (3.1, 7.5)	5.8 (4.1, 9.0)	5.2 (3.2, 7.7)	
Mean (SD)	5.7 (3.7)	6.5 (3.7)	5.8 (3.7)	
Wealth score[a]				0.461
Missing	3446	417	3863	
Median (IQR)	2.5 (2.2, 3.0)	2.6 (2.2, 3.1)	2.5 (2.2, 3.0)	
Mean (SD)	2.6 (0.5)	2.6 (0.6)	2.6 (0.5)	
Clinic of registration				< 0.001
Other clinics	2063 (92.6%)	164 (7.4%)	2227 (100.0%)	

Queen Elizabeth Central Hospital	1855 (85.5%)	315 (14.5%)	2170 (100.0%)

Abbreviations: HIV, Human immunodeficiency virus; km, Kilometres; (Q1,Q3) interquartile range; SD standard deviation; TB tuberculosis.

[a]Wealth score, household wealth developed using household asset ownership.

Table 4.2: Statistical model results for the main analysis, notified TB cases in Urban Blantyre, Malawi from 2015 - 2018

Variable	Unadjusted OR (95% CI)	Adjusted OR (95% CI)	Adjusted OR Network distance[e] (95% CI)	Adjusted OR Cartesian distance[f] (95% CI)
Sex (Male)	0.99 (0.81, 1.21)		{1.06 (0.87, 1.29)	{1.06 (0.87, 1.30)
Age (Years)[a]	1.02 (1.01, 1.02)		1.02 (1.01, 1.02)	1.02 (1.01, 1.02)
HIV Positive	2.53 (2.00, 3.25)		2.21 (1.73, 2.85)	2.21 (1.73, 2.86)
Wealth score[a,b]	1.40 (0.98, 2.04)		1.15 (0.80, 1.73)	1.14 (0.80, 1.72)
Queen Elizabeth Central Hospital	2.10 (1.74, 2.56)			
Network distance (km)[a]	1.05 (1.03, 1.08)			
Cartesian distance (km)[a]	1.06 (1.03, 1.09)			
Road network distance		{1.02 (0.99, 1.06)	0.98 (0.92, 1.03)	
Queen Elizabeth Central Hospital		1.93 (1.57, 2.38)	2.00 (0.93, 4.25)	
Network distance (km)* Queen Elizabeth Central Hospital		1.02 (0.96, 1.08)}[c]	1.07 (0.99, 1.16)	
Cartesian distance		{1.03 (0.99, 1.07)		0.97 (0.89, 1.04)
Queen Elizabeth Central Hospital		1.91 (1.55, 2.35)		2.00 (0.92, 4.27)
Cartesian distance (km)* Queen Elizabeth Central Hospital		1.03 (0.96, 1.11)}[d]		1.09 (1.00, 1.21)

Clinic of registration intercept	0.06 (0.04, 0.11)	0.06 (0.04, 0.11)
Clinic of registration intercept standard deviation	0.67 (0.32, 1.18)}	0.69 (0.33, 1.19)}

Abbreviations: km, Kilometres; OR, Odds ratio.

[a]Age in years, household to clinic distance and household wealth score were cantered by subtracting by the mean.

[b]Wealth score of household wealth developed using household asset ownership.

[c,d]Adjusted for TB treatment registration at Queens Elizabeth Central Hospital (QECH) vs. the other health care facilities, household to clinic distance, and the interaction effect of QECH and household to clinic distance

[e,f]Adjusted for Sex (Male vs. Female), Age in years, HIV status (HIV positive vs HIV negative), Wealth score the household wealth score, TB treatment registration at QECH vs. the other health care facilities, distance household to clinic distance in kilometres, distance (km)* QECH interaction effect of QECH and household to clinic distance and a term for the random intercept of clinic of registration.

The analysis was done for all the data at once (N = 4397) using Bayesian multi-level logistic regression models.

Table 4.3: Statistical model results for the sensitivity analysis: patients with loss to follow-up or transfer out treatment status recoded as having died, in Urban Blantyre, Malawi, 2015-2016

Variable	Unadjusted OR (95% CI)	Adjusted OR (95% CI)	Adjusted OR Network distance[e] (95% CI)	Adjusted OR Cartesian distance[f] (95% CI)
Sex (Male)	1.10 (0.94, 1.28)		{1.06 (0.87, 1.29)	{1.06 (0.87, 1.29)
Age (Years)[a]	1.01 (1.01, 1.02)		1.02 (1.01, 1.03)	1.02 (1.01, 1.03)
HIV Positive	1.76 (1.47, 2.11)		2.21 (1.73, 2.85)	2.21 (1.73, 1.84)
Wealth score[a,b]	1.39 (0.97, 2.04)		1.15 (0.80, 1.72)	1.15 (0.80, 1.72)
Queen Elizabeth Central Hospital	1.44 (1.23, 1.68)			
Network distance (km)[a]	1.03 (1.01, 1.05)			
Cartesian distance (km)[a]	1.01 (0.98, 1.05)			
Road network distance		{1.02 (0.99, 1.04)	0.98 (0.92, 1.03)	
Queen Elizabeth Central Hospital		1.37 (1.16, 1.64)	2.02 (0.94, 4.31)	
Network distance (km)* Queen Elizabeth Central Hospital		1.01 (0.96, 1.07)}[c]	1.07 (0.99, 1.16)	
Cartesian distance		{1.02 (0.99, 1.05)		0.97 (0.89, 1.04)
Queen Elizabeth Central Hospital		1.37 (1.17, 1.61)		2.00 (0.92, 4.26)
Cartesian distance (km)* Queen Elizabeth Central Hospital		1.02 (0.96, 1.09)}[d]		1.09 (1.00, 1.21)

Clinic of registration intercept	0.06 (0.04, 0.11)	0.06 (0.04, 0.11)
Clinic of registration intercept standard deviation	0.68 (0.33, 1.17)}	0.69 (0.33, 1.19)}

Abbreviations: km, Kilometres; OR, Odds ratio.

[a]Age in years, household to clinic distance and household wealth score were centered by subtracting by the mean.

[b]Wealth score of household wealth developed using household asset ownership.

[c,d]Adjusted for TB treatment registration at Queens Elizabeth Central Hospital (QECH) vs. the other health care facilities, household to clinic distance, and the interaction effect of QECH and household to clinic distance.

[e,f]Adjusted for Sex (Male vs. Female), Age in years, HIV status (HIV positive vs HIV negative), Wealth score the household wealth score, TB treatment registration at QECH vs. the other health care facilities, distance household to clinic distance in kilometres, distance (km QECH interaction effect of QECH and household to clinic distance and a term for the random intercept of clinic of registration.

The analysis was done for all the data at once (N = 4397) using Bayesian multi-level logistic regression models with participants that defaulted and those that transferred out (N = 258) recoded as died.

Table 4.4: Statistical model results for the sensitivity analysis: Restricted to patients with microbiologically-confirmed Tuberculosis disease only in Urban Blantyre Malawi, 2015-2018

Variable	Unadjusted OR (95% CI)	Adjusted OR (95% CI)	Adjusted OR Network distance[e] (95% CI)	Adjusted OR Cartesian distance[f] (95% CI)
Sex (Male)	0.93 (0.69-1.28)		{0.99 (0.72-1.38)	{0.99 (0.72-1.38)
Age (Years)[a]	1.02 (1.00-1.03)		1.01 (1.00-1.03)	1.01 (1.00-1.03)
HIV Positive	3.06 (2.12-4.59)		2.65 (1.79-4.06)	2.66 (1.78-4.03)
Wealth score[a,b]	1.85 (0.95-3.49)		1.63 (0.84-3.04)	1.62 (0.84-3.37)
Queen Elizabeth Central Hospital	2.22 (1.62-3.04)			
Network distance (km)[a]	1.07 (1.02-1.11)			
Cartesian distance (km)[a]	1.06 (0.98-1.14)			
Road network distance		{0.97 (0.88-1.04)	0.97 (0.88-1.04)	
Queen Elizabeth Central Hospital		1.67 (0.86-3.42)	1.67 (0.86-3.42)	
Network distance (km)* Queen Elizabeth Central Hospital		1.14 (1.01-1.29)}[c]	1.14 (1.01-1.29)	
Cartesian distance		{1.01 (0.94-1.08)		0.98 (0.87-1.06)
Queen Elizabeth Central Hospital		1.73 (1.18-2.43)		1.66 (0.86-3.37)
Cartesian distance (km)* Queen Elizabeth Central Hospital		1.14 (1.00-1.29)}[d]		1.16 (1.01-1.35)

Clinic of registration intercept	0.04 (0.02-0.09)	0.04 (0.02-0.0.09)
Clinic of registration intercept standard deviation	0.45 (0.02-1.22)}	0.43 (0.02-1.20)}

Abbreviations: km, Kilometres; OR, Odds ratio.

[a]Age in years, household to clinic distance and household wealth score were centered by subtracting by the mean.

[b]Wealth score of household wealth developed using household asset ownership.

[c,d]Adjusted for TB treatment registration at Queens Elizabeth Central Hospital (QECH) vs. the other health care facilities, household to clinic distance, and the interaction effect of QECH and household to clinic distance

[e,f]Adjusted for Sex (Male vs. Female), Age in years, HIV status (HIV positive vs HIV negative), Wealth score the household wealth score, TB treatment registration at QECH vs. the other health care facilities, household to clinic distance in kilometres, distance (km)* QECH interaction effect of QECH and household to clinic distance and a term for the random intercept of clinic of registration.

The analysis was done using N= 2015 participants that had a microbiologically confirmed TB diagnosis using Bayesian multi-level logistic regression models.

Table 4.s1: **Characteristics for notified Tuberculosis cases in Urban Blantyre Malawi, 2015-2018. Comparing participants who were included in the study versus those who were not part of the study but were from Blantyre**

	In study (N=4461)	Not in study (N=738)	Total (N=5199)
Year of registration			
2015	966 (21.7%)	149 (20.2%)	1115 (21.4%)
2016	1330 (29.8%)	191 (25.9%)	1521 (29.3%)
2017	1437 (32.2%)	222 (30.1%)	1659 (31.9%)
2018	728 (16.3%)	176 (23.8%)	904 (17.4%)
Gender			
Male	2827 (63.4%)	405 (54.9%)	3232 (62.2%)
Female	1634 (36.6%)	333 (45.1%)	1967 (37.8%)
Age, years			
Median (Q1, Q3)	35.000 (28.0, 42.0)	35.0 (21.3, 45.0)	35.0 (27.0, 42.0)
Mean (SD)	35.5 (13.4)	34.206 (19.0)	35.3 (14.4)
HIV status			
HIV negative	1428 (32.0%)	333 (45.1%)	1761 (33.9%)
HIV positive	3033 (68.0%)	405 (54.9%)	3438 (66.1%)
TB classification			
Extrapulmonary TB	1637 (36.7%)	466 (63.1%)	2103 (40.5%)
Pulmonary TB	2824 (63.3%)	272 (36.9%)	3096 (59.5%)
Micro confirmed TB			
Not microb-confirmed TB	2420 (54.2%)	621 (84.1%)	3041 (58.5%)
Microb-confirmed TB	2041 (45.8%)	117 (15.9%)	2158 (41.5%)
TB treatment outcome			
N-Miss	64	15	79
Cured	916 (20.8%)	31 (4.3%)	947 (18.5%)
Treatment completed	2715 (61.7%)	419 (58.0%)	3134 (61.2%)
Died	479 (10.9%)	94 (13.0%)	573 (11.2%)
Treatment failure	29 (0.7%)	2 (0.3%)	31 (0.6%)
Lost to follow up	258 (5.9%)	177 (24.5%)	435 (8.5%)
Clinic of registration			
Other clinics	2257 (50.6%)	113 (15.3%)	2370 (45.6%)
Queen Elizabeth Central Hospital	2204 (49.4%)	625 (84.7%)	2829 (54.4%)

Abbreviations: HIV, Human immunodeficiency virus; (Q1, Q3) interquartile range; SD standard deviation; TB tuberculosis

References

1. **World Health Organization**. *Global tuberculosis report 2018*. *WHO*. Geneva, 2019.
2. **World Health Oganisation**. *Global Tuberculosis Control Report 2009*. Geneva, 2010.
3. **Heunis JC, *et al.*** *Risk factors for mortality in TB patients: A 10-year electronic record review in a South African province*. *BMC Public Health*. BioMed Central Ltd., 2017, p. 38.
4. **Mayer CM, *et al.*** Distance to clinic is a barrier to PrEP uptake and visit attendance in a community in rural Uganda. *Journal of the International AIDS Society* Switzerland: John Wiley and Sons Inc., 2019; **22**: e25276.
5. **Wold Health Organisation**. *Systematic screening for active tuberculosis*. Geneva, 2013.
6. **Momenyan S, *et al.*** Spatial inequalities and predictors of HIV/AIDS mortality risk in Hamadan, Iran: a retrospective cohort study. *Epidemiology and health* 2018; **40**: e2018038.
7. **O'Donnell O**. *Access to health care in developing countries: Breaking down demand side barriers*. *Cadernos de Saude Publica*. Fundacao Oswaldo Cruz, 2007, p. 2820–2834.
8. **Tripathy JP, *et al.*** Is physical access an impediment to tuberculosis diagnosis and treatment? A study from a rural district in North India. *Public Health Action* International Union Against Tuberculosis and Lung Disease, 2013; **3**: 235–239.
9. **Fluegge K, *et al.*** Impact of geographic distance on appraisal delay for active TB treatment seeking in Uganda: A network analysis of the Kawempe Community Health Cohort Study. *BMC Public Health* BioMed Central Ltd., 2018; **18**: 798.
10. **Cai J, *et al.*** Factors associated with patient and provider delays for tuberculosis diagnosis and treatment in Asia: A systematic review and meta-analysis. *PLoS ONE* Public Library of Science, 2015; **10**Published online: 25 March 2015.doi:10.1371/journal.pone.0120088.
11. **National Statistical Office of Malawi (NSO)**. *Malawi population census preliminary report*. Zomba, 2018.
12. **Malawi Ministry of Health**. Malawi Population-Based HIV Impact Assessment (MPHIA) 2015-2016: Final Report. Lilongwe, Ministry of Health. October 2018.
13. **Makwero MT**. Delivery of primary health care in Malawi. *African Journal of Primary Health Care and Family Medicine* AOSIS (pty) Ltd, 2018; **10**: 1799.
14. **World Health Organization (WHO)**. TB case and treatment outcome definitions. World Health Organization, 2014; Published online: 2014.
15. **Lovelace R, Ellison R**. *stplanr: A Package for Transport Planning*.

16. **Payongayong E, *et al.*** *Simple household poverty assessment models for Malawi: proxy means test from the 1997–98 Malawi Integrated Household Survey. Lilongwe, Malawi: National Statistical Office, Government of Malawi, 2006*.

17. **Bürkner P-C**. brms: An R Package for Bayesian Multilevel Models Using Stan. *Journal of Statistical Software* 2017; **80**: 1–28.

18. **Squire SB, *et al.*** 'Lost' smear-positive pulmonary tuberculosis cases: Where are they and why did we lose them? *International Journal of Tuberculosis and Lung Disease* 2005; **9**: 25–31.

19. **Secretary of Jan Swasthya Sahyog, Laux TS, Patil S**. Predictors of tuberculosis treatment outcomes among a retrospective cohort in rural, Central India. *Journal of Clinical Tuberculosis and Other Mycobacterial Diseases* Elsevier Ltd, 2018; **12**: 41–47.

20. **MacPherson P, *et al.*** Disparities in access to diagnosis and care in Blantyre, Malawi, identified through enhanced tuberculosis surveillance and spatial analysis. *BMC Medicine* BioMed Central, 2019; **17**: 21.

21. **Virenfeldt J, *et al.*** Treatment delay affects clinical severity of tuberculosis: A longitudinal cohort study. *BMJ Open* BMJ Publishing Group, 2014; **4**: 4818.

22. **Lonnroth, Knut**. *Systematic screening for active tuberculosis Principles and recommendations*.

23. **Chihota VN, *et al.*** Missed Opportunities for TB Investigation in Primary Care Clinics in South Africa: Experience from the XTEND Trial. *PloS one* Public Library of Science, 2015; **10**: e0138149.

24. **van't Hoog AH, *et al.*** Screening Strategies for Tuberculosis Prevalence Surveys: The Value of Chest Radiography and Symptoms. Herrmann JL, ed. *PLoS ONE* Public Library of Science, 2012; **7**: e38691.

25. **Chikovore J, *et al.*** 'For a mere cough, men must just chew Conjex, gain strength, and continue working': The provider construction and tuberculosis care-seeking implications in Blantyre, Malawi. *Global Health Action* Co-Action Publishing, 2015; **8**Published online: 2015.doi:10.3402/gha.v8.26292.

26. **United Nations Joint Programme on HIV/AIDS (UNAIDS)**. Joint United Nations Programme on HIV/AIDS. UNAIDS data 2020. Geneva, Switzerland. 2020; : 436.

27. **United Nations Joint Programme on HIV/AIDS (UNAIDS)**. *Malawi country HIV/AIDS factsheet 2019*. (https://www.unaids.org/en/regionscountries/countries/malawi). Accessed 29 October 2020.

28. **Kwan C, Ernst JD**. *HIV and tuberculosis: A deadly human syndemic*. *Clinical Microbiology*

Reviews. American Society for Microbiology (ASM), 2011, p. 351–376.

29. **Nogueira BMF, *et al.*** Factors associated with tuberculosis treatment delay in patients co-infected with HIV in a high prevalence area in Brazil. *PLoS ONE* Public Library of Science, 2018; **13**Published online: 1 April 2018.doi:10.1371/journal.pone.0195409.

30. **Gesesew H, *et al.*** The prevalence and associated factors for delayed presentation for HIV care among tuberculosis/HIV co-infected patients in Southwest Ethiopia: A retrospective observational cohort. *Infectious Diseases of Poverty* BioMed Central Ltd., 2016; **5**: 96.

31. **Gupta-Wright A, *et al.*** *Virological failure, HIV-1 drug resistance, and early mortality in adults admitted to hospital in Malawi: an observational cohort study*. *The Lancet HIV*. 2020.

32. **Craik A, *et al.*** Challenges with targeted viral load testing for medical inpatients at Queen Elizabeth Central Hospital in Blantyre, Malawi. *Malawi Medical Journal* Malawi Medical Journal, 2016; **28**: 179–181.

33. **World Health Oganisation**. *UPDATE OF RECOMMENDATIONS ON FIRST-AND SECOND-LINE ANTIRETROVIRAL REGIMENS*. 2019.

34. **Cegielski JP, *et al.*** Eliminating tuberculosis one neighborhood at a time. *American journal of public health* American Public Health Association, 2013; **103**: 1292–1300.

35. **Robsky KO, *et al.*** Is distance associated with tuberculosis treatment outcomes? A retrospective cohort study in Kampala, Uganda. *BMC Infectious Diseases* BioMed Central Ltd., 2020; **20**: 406.

36. **Royston P, Altman DG, Sauerbrei W**. Dichotomizing continuous predictors in multiple regression: A bad idea. *Statistics in Medicine* Stat Med, 2006; **25**: 127–141.

37. **Altman DG, Royston P**. *The cost of dichotomising continuous variables*. *British Medical Journal*. BMJ Publishing Group, 2006, p. 1080.

38. **Textor J, *et al.*** Robust causal inference using directed acyclic graphs: The R package 'dagitty'. *International Journal of Epidemiology* Oxford University Press, 2016; **45**: 1887–1894.

5 TB hotspot identification

5.1 Introduction

This chapter was aimed at developing a modelling strategy for identifying hotspot neighbourhoods of burden of undiagnosed TB cases. This was achieved by analysing of neighbourhood prevalence-to-notification ratios for adult bacteriologically-confirmed tuberculosis using data from systematic surveillance of notified TB cases and a citywide TB prevalence survey form Urban Blantyre, Malawi. I led the analysis of this manuscript and also led the writing of the paper. The paper was submitted on 4 November 2021 to Plos One and was published on 6 May 2022. The paper has been described verbatim below.

In summary, I found that, overall, there remains a substantial burden of undiagnosed TB in urban Africa. However, there is significant variation by neighbourhood, implying that ACF interventions should be prioritised to reach communities with higher barriers to accessing TB care at health facilities. Since these communities are likely to have fewer resources to afford the diagnosis offered through routine care, community screening using spatially targeted interventions would help achieve equitable access to TB diagnosis. This model should enable other places with similar characteristics to urban Blantyre to highlight potential hotspot neighbourhoods without the need to carry out a full prevalence survey.

5.2 TB hotspots manuscript

LONDON SCHOOL of HYGIENE & TROPICAL MEDICINE

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT

T: +44 (0)20 7299 4646
F: +44 (0)20 7299 4656
www.lshtm.ac.uk

RESEARCH PAPER COVER SHEET

Please note that a cover sheet must be completed for each research paper included within a thesis.

SECTION A – Student Details

Student ID Number	lsh1805281	**Title**	Mr
First Name(s)	McEwen Joseph		
Surname/Family Name	Khundi		
Thesis Title	Towards a framework for improved targeting of tuberculosis interventions: a spatial analysis of patient notification and prevalence survey data from Blantyre, Malawi		
Primary Supervisor	James Carpenter		

If the Research Paper has previously been published please complete Section B, if not please move to Section C.

SECTION B – Paper already published

Where was the work published?	PLOS ONE		
When was the work published?	6 May 2022		
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion			
Have you retained the copyright for the work?*	**No**	Was the work subject to academic peer review?	**Yes**

*If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be published?	
Please list the paper's authors in the intended authorship order:	

Stage of publication	Choose an item

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	I developed the research question, led the analysis, led the writing of the manuscript, and submitted it for publication. I am the first author.

SECTION E

Student Signature	[redacted]
Date	7 July 2022

Supervisor Signature	[redacted]
Date	[illegible]

Neighbourhood prevalence-to-notification ratios for adult bacteriologically-confirmed tuberculosis reveals hotspots of underdiagnosis in Blantyre, Malawi

McEwen Khundi[1,2]*, James R Carpenter[1], Elizabeth L Corbett[1,2], Helena R A Feasey[1,2], Rebecca Nzawa Soko[1,2], Marriott Nliwasa[1,3], Hussein Twabi[3], Lingstone Chiume[1], Rachael M Burke[1,2], Katherine C Horton[2], Peter J Dodd[4], Ted Cohen[5], Peter MacPherson[1,2,6]

[1]Malawi-Liverpool-Wellcome Trust Clinical Research Programme, Blantyre, Malawi,

[2]London School of Hygiene and Tropical Medicine, UK,

[3]Helse Nord TB Initiative, College of Medicine, University of Malawi,

[4]School of Health and Related Research, University of Sheffield, Sheffield, UK,

[5]Yale School of Public Health, New Haven, USA,

[6]Department of Clinical Sciences, Liverpool School of Tropical Medicine, UK

*Corresponding author

E-mail: mcewenkhundi@gmail.com

Abstract (279)

Local information is needed to guide targeted interventions for respiratory infections such as tuberculosis (TB). Case notification rates (CNRs) are readily available, but systematically underestimate true disease burden in neighbourhoods with high diagnostic access barriers. We explored a novel approach, adjusting CNRs for under-notification (P:N ratio) using neighbourhood-level predictors of TB prevalence-to-notification ratios. We analysed data from 1) a citywide routine TB surveillance system including geolocation, confirmatory mycobacteriology, and clinical and demographic characteristics of all registering TB patients in Blantyre, Malawi during 2015-19, and 2) an adult TB prevalence survey done in 2019. In the prevalence survey, consenting adults from randomly selected households in 72 neighbourhoods had symptom-plus-chest X-ray screening, confirmed with sputum smear microscopy, Xpert MTB/Rif and culture. Bayesian multilevel models were used to estimate adjusted neighbourhood prevalence-to-notification ratios, based on summarised posterior draws from fitted adult bacteriologically-confirmed TB CNRs and prevalence. From 2015-19, adult bacteriologically-confirmed CNRs were 131 (479/371,834), 134 (539/415,226), 114 (519/463,707), 56 (283/517,860) and 46 (258/578,377) per 100,000 adults per annum, and 2019 bacteriologically-confirmed prevalence was 215 (29/13,490) per 100,000 adults. Lower educational achievement by household head and neighbourhood distance to TB clinic was negatively associated with CNRs. The mean neighbourhood P:N ratio was 4.49 (95% credible interval [CrI]: 0.98- 11.91), consistent with slow/underdiagnosis of TB, and was most pronounced in informal peri-urban neighbourhoods. Here we have demonstrated a method for the identification of neighbourhoods with high levels of under-diagnosis of TB without the

requirement for a prevalence survey; this is important since prevalence surveys are expensive and logistically challenging. If confirmed, this approach may support more efficient and effective targeting of intensified TB and HIV case-finding interventions aiming to accelerate elimination of urban TB.

Introduction

Despite substantial investment under global health initiatives, progress towards ending tuberculosis (TB) epidemics has been disappointingly slow. TB remained the leading infectious cause of adult death in 2019, with an estimated 1.4 million deaths, and was second only to COVID-19 in 2020 [1,2]. Estimated incidence has been falling, but not rapidly enough to meet the EndTB Strategy goals [1,2]. The World Health Organization (WHO) African Region achieved a 19% reduction in TB incidence between 2015 and 2020, mainly attributable to improving HIV and TB prevention and treatment services [1,2]. Early diagnosis of TB is essential for prevention of TB deaths and new infections, as undiagnosed TB patients can remain infectious for many years if not effectively treated [3]. In 2020, however, 4.1 million (or 41% of all incident TB patients), globally, were estimated to remain undiagnosed or unnotified – with substantial increases in these "missing millions" partly because of COVID-19 disruptions in TB diagnostic services [1,2,4].

Reaching the ambitious WHO EndTB Strategy targets for incidence (90% reduction from 2015) and death (95% reduction from 2015) from TB by 2035 will require innovative strategies [2,5]. Efficient diagnosis of self-presenting patients reporting TB symptoms at health facilities, although critical to patient management, is unlikely to be sufficient unless accompanied by community-based interventions [3,6]. Active case-finding for undiagnosed TB disease (ACF), using approaches such as door-to-door enquiry for chronic cough, can rapidly reduce undiagnosed TB prevalence, but is limited to very high prevalence populations by the cost and performance of currently available TB diagnostics [3,6–8].

Previous attempts to define “hotspots” of TB disease in urban and rural Africa and Asia have used locally-resolved case notification rates [9] that cannot distinguish poorer access to TB diagnostic services from true low disease burden [10,11]. True TB burden is highly heterogeneous but, even in the same District or City, substantial heterogeneity in routine TB service access [3,12] tends to obscure hotspots, which typically represent the combined effects of adverse social and environmental determinants with high barriers to accessing health services [3,13,14].

Data from TB prevalence surveys can provide detailed neighbourhood-level data on TB determinants and undiagnosed TB burden to guide National TB Programmes and District Health Officers [9,12,15,16], allowing neighbourhoods with delayed detection and incomplete detection [16,17] of TB to be identified through high prevalence-to-notification (P:N) ratios [17,18]. TB prevalence surveys are, however, costly and logistically demanding. Identification of neighbourhoods with high levels of under-diagnosis without requirement for prevalence surveys could be of major benefit to National TB Programmes. Here, we aimed to develop simple, accurate models that could be used by researchers and TB Programme Managers using high-quality spatially-resolved TB neighbourhood level data from urban Blantyre, Malawi. We used multilevel Bayesian modelling to generate neighbourhood level P:N ratios [9,17], aiming to smooth over sparse prevalence data and borrow strength across neighbourhoods and make predictions beyond available prevalence data to support better prioritisation of community-based ACF.

Methods

Study setting

Blantyre District is in the Southern Region of Malawi, Central Africa. Blantyre City, in the centre of the District, is a major commercial centre and has several densely-populated informal settlements, as well as more-established urban and peri-urban neighbourhoods where rates of poverty are high and access to municipal and health services are limited [19]. In the 2018 Malawi National census the Blantyre City population was 800,264 (502,018 adults ≥ 15y) [20] and adult HIV prevalence was 18% in a recent population-based survey [21].

Blantyre enhanced TB monitoring and evaluation of TB notifications

Patients diagnosed with TB in Malawi are registered by the National TB Programme. TB registration clinics in Blantyre include a government referral hospital, free public clinics and a small number of private health facilities [22]. TB Officers (a cadre of health workers employed by Malawi's Ministry of Health with responsibility for delivering TB services) [22] were supported to strengthen the TB notification surveillance system in Blantyre as part of a joint project between the Malawi-Liverpool-Wellcome Trust Clinical Research Programme (MLW), the Blantyre District Health Office, and the Malawi National TB Control Programme (NTP). All TB patients are offered provider-initiated HIV testing and antiretroviral therapy if newly diagnosed with HIV [22].

From 2015, TB Officers have been supported to use an electronic data capture application (ePaL) to collect additional clinical, sociodemographic, and household level data and a confirmatory sputum for microscopy and culture as part of the citywide enhanced surveillance

system [4,11]. ePAL supports capture of global positioning satellites (GPS) coordinates identifying the place of residence for each TB patient, using high resolution satellite maps with locally-captured reference locations within each neighbourhood of the city. The ePAL application has previously been validated and described elsewhere [11,23]. All patients starting TB treatment were asked to provide an additional single spot sputum sample for smear microscopy and mycobacteria growth indicator tube (MGIT) culture, performed at the MLW/University of Malawi College of Medicine TB Research Laboratory. The enhanced TB surveillance data and NTP registers were reconciled on a quarterly basis, and monthly 5% of patients were traced to home for data validation purposes.

TB prevalence survey

In 2019, a TB prevalence survey was carried out in Blantyre City by the MLW study team at the start of a planned cluster-randomised trial of community-based TB screening interventions, subsequently interrupted by COVID-19 (ISRCTN11400592). 72 neighbourhood clusters were defined, each comprised of several community health workers (CHW) areas, with the goal of having approximately 4000 adults in each neighbourhood. CHW areas are the smallest health administrative unit in the city, and each is affiliated to a primary health clinic [11,24]. Using Google Earth, a geographical information specialist captured the GPS coordinates of all the houses in the 72 neighbourhood clusters to be used as the prevalence survey sampling frame. In each neighbourhood, 115 households were selected at random for participation into the prevalence survey with the aim of recruiting 215 adults (≥18 years old) per neighbourhood.

Adults from the randomly selected households were visited and invited to attend a study tent located at a central point within the neighbourhood for TB and HIV investigations. TB screening was provided at the tent using a digital chest radiograph that was immediately read by an experienced radiographer trained in TB prevalence surveys, and with interpretation supported by computer assisted diagnostic software (Qure.ai version 2.0). Participants who had an abnormal chest X-ray or reported cough of any duration were asked to provide two spot sputum samples for Xpert MTB/Rif, smear microscopy, and MGIT culture. Positive Xpert MTB/Rif and smear microscopy results were provided within two days, and culture results within approximately six weeks; a prevalent TB case was defined as a positive result for Xpert MTB/Rif or smear microscopy, with a positive MGIT culture result that was speciated as *Mycobacterium tuberculosis*. HIV testing was offered using both OraQuick (OraSure Technologies, manufactured in Thailand) oral HIV test kits and a rapid fingerprick kits (Determine 1/2, Alere, USA) in parallel. Positive HIV results were confirmed using Uni-Gold (Trinity Biotech, Ireland). If participants verbally reported being HIV positive, only a Uni-Gold confirmation test was done. All participants with newly diagnosed HIV were provided with post-test counselling and assisted to register for HIV treatment at their nearest primary care clinic.

Neighbourhood populations

In 2015, the MLW study team conducted a population census in all of the city's CHW areas [11]; in an independent exercise the Malawi National Statistical Office (NSO) conducted the Malawi Population and Household National Census in 2018 that included Blantyre City [20]. Population denominators were derived from the MLW study team's 2015 census and the NSO's 2018 census. The 2018 NSO and 2015 MLW study census data were used to calculate annual

population growth rates, which were then used to estimate annual population denominators for the 72 study neighbourhoods from 2015 to 2019. The growth rates were calculated separately for ages 0 to 4yrs, 5 to 14yrs, 15yrs or older adult males and 15yrs and older adult females. Growth rates were assumed to be the same for all the neighbourhoods.

Calculation of empirical TB notification and prevalence rates and predictors

Empirical neighbourhood-level TB case notification rates (CNR) from 2015 to 2019 and 2019 TB prevalence rates were calculated by summing the adult (≥18y) TB notifications or prevalent cases at neighbourhood level and dividing by respective adult population denominators and multiplying by 100,000 to scale the rate to per 100,000 population. The percentage of adults in each neighbourhood and the percentage of male adults aged 15 years or older were calculated using data from the 2015 study team Blantyre City census and were assumed to be consistent from 2016 to 2019. The distance to the nearest TB clinic was estimated by calculating the cartesian (straight line) distance between the centroid of each neighbourhood and the nearest TB clinic; this served as a proxy indicator for access to TB diagnosis and treatment [25]. Neighbourhood HIV prevalence was calculated using prevalence survey data, and the percentage of households whose head never finished primary school — a proxy variable for poverty — was estimated using data from the TB prevalence study. According to the Malawi Integrated Household Survey, household head education level was closely associated with household poverty [26,27].

Neighbourhood baseline characteristics

We report the neighbourhood-level percentage of adults aged 15 years or older, percentage of male adults (15 years or older), distance to nearest TB clinic, percentage of households with head of house who never finished primary school and HIV prevalence, summarised by their mean, range, and standard deviation (sd). We plotted the spatial distribution of the covariates across the 72 neighbourhood clusters on choropleth maps.

Statistical modelling

We fitted Bayesian multilevel models to estimate neighbourhood-level adult (≥ 18y) annual bacteriologically-confirmed TB case notification rates for each year between 2015 and 2019, and separately for prevalence rates for 2019. Models were fitted using Markov chain Monte Carlo (MCMC) sampling using the brms package as an interface to Stan in R, with inference based on three chains of 14,000 posterior samples after discarding 1000 burn-in samples [28]. Since the notification and prevalence data were from two different and independent datasets they were modelled separately to allow greater control in exploring covariates. The bacteriologically-confirmed case notification data were modelled using a Poisson response distribution (Equation 5.1), we included a dummy variable for year with the reference level of 2019.The prevalence data were modelled using a zero-inflated-Poisson distribution (Equation 5.2) to account for overdispersion and excess neighbourhoods with zero prevalent cases. Neighbourhood-level random effects were modelled with a spatial intrinsic conditional autoregressive (ICAR) term (5.S1 and 5.S2 Equations) or a random intercept term (Equation 5.1 and 5.2), but not with both random terms in the same model. Possible combination of

neighbourhood-level variables were derived for the TB case notification and prevalence data. The set of variables considered for inclusion in the model were selected based on previous research [11], ease of measurement for TB programmes, and availability within the datasets. The models' predictive performance were evaluated using leave-one-out (LOO) cross-validation [29]. The best fitting models were selected based on their expected log pointwise predictive density (ELPD) LOO statistic [29] (5.S5 and 5.S6 Tables). Weakly regularising priors were assigned to model intercepts and slopes. Model convergence was assessed by visual inspection of trace plots, effective sample sizes and Gelman-Rubin statistics [28].

Equation 5.1

Let $Y \sim Pois(\mu)$

$$\Pr(Y_{ij} = y_{ij}) = \left(\frac{\mu_{ij}^{y_{ij}}\, exp(-\mu_{ij})}{y_{ij}!}\right)$$

$$\log(\mu_{ij}) = \alpha + \alpha_i + \beta_1 x_{1i} + \beta_2 x_{2i} \ldots + \beta_k x_{ki} + \beta_{year2015} year2015_j + \beta_{year2016} year2016_j + \beta_{year2017} year2017_j + \beta_{year2018} year2018_j + \log(Pop_{ij})$$

$$\alpha \sim Normal(\mu_\alpha = 0, \sigma_\alpha^2 = 10)$$

$$\beta_k \sim Normal(\mu_\beta = 0, \sigma_\beta^2 = 10)$$

$$\alpha_i \sim Normal(0, \sigma_{\alpha_i}^2)$$

$$\sigma_{\alpha_i} \sim HalfCauchy(0, 1)$$

The expectation of $Y_{ij} = y_{ij}$ given by:

$$E(Y_{ij} = y_{ij}) = \mu_{ij}$$

Where i refers to neighbourhood i for i =1,2,3..72, and *j* indexes year (2015,…2019), $(\alpha + \alpha_i)$ intercept parameters and $\beta_1, \beta_2 \ldots \beta_k$ are unknown regression coefficients that are estimated from the data for the cluster level covariates x_{1i}, x_{2i} … x_{ki} , and sigma (σ_{α_i}) is the standard deviation for the random intercept for neighbourhood i. Pop_{ij} is the total population of neighbourhood in year *j* that is used as the offset. Here $year2015_j$, $year2016_j$, $year2017_j$ and $year2018_j$ are dummy variables which take the value of one for that year and take the value of zero for the other years (the baseline year was 2019).

Equation 5.2

To define the zero-inflated Poisson model, let $Z \sim Bern(1-p)$ (so $\Pr\{Z = 0\} = p$) and independently $W \sim Pois(\mu)$. Then the data are modelled by the zero-inflated variable Y which is defined as

$$Y = ZW$$

$$\Pr(Y_i = y_i) = \begin{cases} \left(p + (1-p)exp(-\mu_i)\right) if\ y_i = 0 \\ \left((1-p)\frac{\mu_i^{y_i}\, exp(-\mu_i)}{y_i!}\right) if\ y_i \geq 1 \end{cases}$$

$$logit(p_i) = \vartheta$$

$$\vartheta \sim Normal\left(\mu = 0, \sigma_{\vartheta}^2 = 10\right)$$

$$\log(\mu_i) = \alpha + \alpha_i + \beta_1 x_{1i} + \beta_2 x_{2i} \ldots + \beta_k x_{ki} + \log(Pop_i)$$

$$\alpha \sim Normal(\mu_\alpha = 0, \sigma_\alpha^2 = 10)$$

$$\beta_k \sim Normal\left(\mu_\beta = 0, \sigma_\beta^2 = 10\right)$$

$$\alpha_i \sim Normal\left(0, \sigma_{\alpha_i}^2\right)$$

$$\sigma_{\alpha_i} \sim HalfCauchy(0, 1)$$

The expectation of $\mathrm{Y}_i = y_i$ is given by:

$$E(Y_i = y_i) = (1 - p)\,\mu_i$$

Where i refers to neighbourhood i for i =1,2,3…72 (note we only have prevalence data for one year, 2019). ϑ, $(\alpha + \alpha_i)$ intercept parameters and $\beta_1, \beta_2 \ldots \beta_k$ are unknown regression coefficients that are estimated from the data for the covariates $x_{1i}, x_{2i} \ldots x_{ki}$, and sigma (σ_{α_i}) is the standard deviation for the random intercept for neighbourhood i. Pop_i is the total population of neighbourhood that is used as the offset.

The intercept, mean rate ratios and 95% credible intervals (CrI) of the selected models are presented in a model summary table. We drew 42,000 posterior samples of the prevalence and notification rates from the selected models. The posterior bacteriologically-confirmed P:N ratios were calculated by dividing the posterior prevalence rates by the posterior notification rates (Equation 5.3). We obtained the posterior mean of the P:N ratios and their 95% posterior Crl, and summarised their distribution in a choropleth map and a caterpillar plot [30]. Analysis was conducted using R version 4.0.3 (R Foundation for Statistical Computing, Vienna).

Equation 5.3

$$\text{Posterior notification rate}_{ij} = \exp(\alpha_j + \alpha_{ij}) * 100{,}000 \text{ (From equation 5.1)}$$

$$\text{Posterior prevalance rate}_{ij} = (1 - p_j) * \exp(\alpha_j + \alpha_{ij}) * 100{,}000 \text{ (From equation 5.2)}$$

Where j indexes the posterior sample $j = 1, 2, 3, \ldots 42{,}000$ and i is for neighbourhood i = 1,2,3 … 72.

The P:N ratio posterior was calculated by first drawing a sample of 42,000 posterior samples of the posterior notification and prevalence rates per 100,000. The j^{th} posterior P:N ratio was calculated by dividing the j^{th} posterior prevalence rate by the j^{th} posterior notification rate.

Sensitivity analysis

Neighbourhood-level TB prevalence was post-stratified according to age-sex groups in neighbourhood populations from WorldPop in order to correct for under-participation of some age-sex groups in the prevalence survey [31]. WorldPop population estimates were used because they had more granular age-sex groups than our census-based denominators. The post stratified TB prevalence was used to reproduce P:N ratios. We additionally undertook sensitivity analysis to estimate neighbourhood P:N ratio for all adult forms of TB, including both bacteriologically-confirmed and clinically-diagnosed cases.

Ethical considerations

Ethical approval was granted by the London School of Hygiene and Tropical Medicine (16228) and the College of Medicine, University of Malawi Research Ethics Committee (P.12/18/2556). Participants in both the prevalence survey and MLW study census provided written informed consent.

Results

Neighbourhood characteristics

From 2015 to 2019, the estimated total population of the 72 study neighbourhoods increased from 612,792 to 905,419, with 60.90 % (range: 54.80-70.60, sd: 3.00) adults (≥15 years). There was substantial neighbourhood-level variability in the mean percentage of adults who were men (51.54%, range: 46.89-55.23, sd: 1.55), living in households headed by someone who had not completed primary education (low education: 16.90%, range 4.30-32.40%, sd: 6.10) and HIV prevalence 13.80% (range: 4.21-27.44%, sd: 4.32). The mean distance from the centroid of the neighbourhoods to the nearest TB clinic was (1.74 km, range: 0.36-3.68 km, sd: 0.89) Table 5.1. Neighbourhoods located near the centre of the city had higher population density, and higher proportions of adults and male residents (Figure 5.1). The peri-urban neighbourhoods tended to have higher distance to the nearest TB clinic, higher HIV prevalence, and a greater percentage of household heads without primary education.

Figure 5.1. Choropleth maps all covariates considered in predictive models of TB case prevalence and notification rates

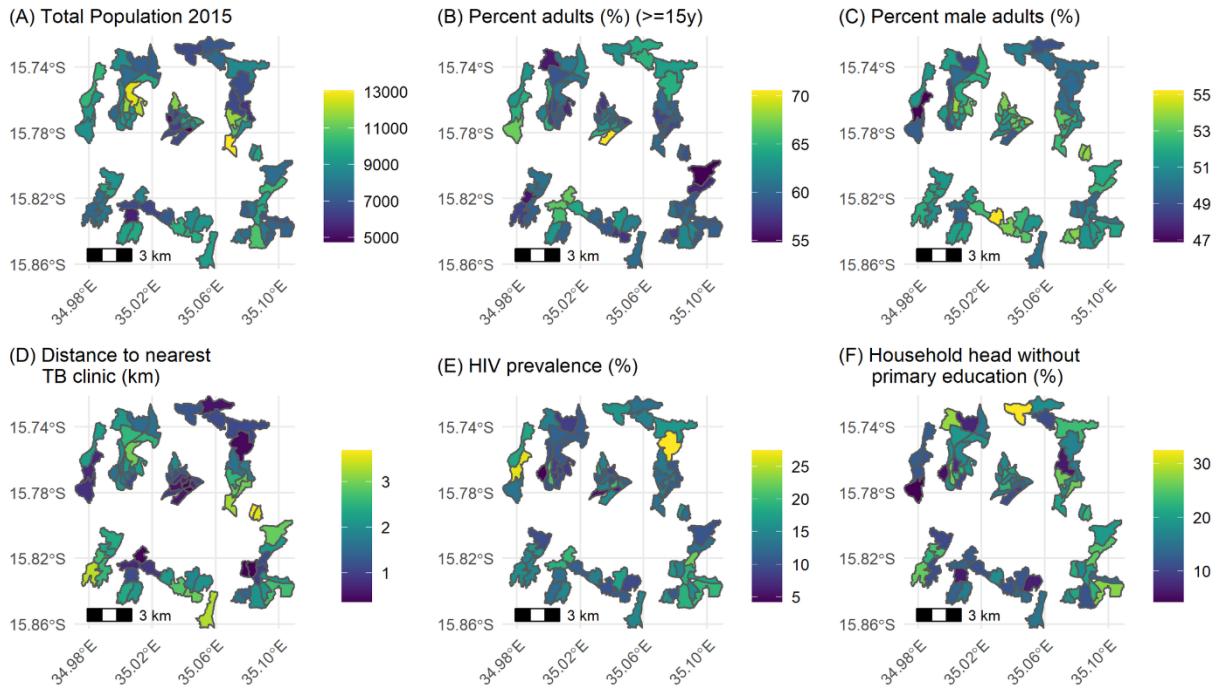


Table 5.1: Neighbourhood-level summary data for the 72 neighbourhoods

Characteristic	Mean (sd)	Range	n	N
2015 bacteriologically-confirmed adult TB notification rate (per 100,000)	131 (77)	0—383	479	371,834
2016 bacteriologically-confirmed adult TB notification rate (per 100,000)	134 (70)	32—328	539	415,226
2017 bacteriologically-confirmed adult TB notification rate (per 100,000)	114 (57)	28—291	519	463,707
2018 bacteriologically-confirmed adult TB notification rate (per 100,000)	56 (38)	0—167	283	517,860
2019 bacteriologically-confirmed adult TB notification rate (per 100,000)	46 (29)	0—144	258	578,377
2019 adult bacteriologically-confirmed TB prevalence rate (per 100,000)	215 (335)	0—1,415	29	13,490
2015 adult TB notification rate† (per 100,000)	242 (119)	23—639	884	371,834
2016 adult TB notification rate† (per 100,000)	273 (128)	47—629	1,106	415,226
2017 adult TB notification rate† (per 100,000)	251 (101)	42—451	1,157	463,707
2018 adult TB notification rate† (per 100,000)	127 (66)	13—287	642	517,860
2019 adult TB notification rate† (per 100,000)	150 (65)	28—349	849	578,377
Percentage of adults (≥15y) (%)	60.90 (3.00)	54.80—70.60	371,834	612,792
Percentage of male adults (%)	51.54 (1.55)	46.89—55.23	191,855	371,834
Household head without primary education (%)	16.90 (6.10)	4.30—32.40	2,700	15,897
Distance to TB clinic (km)	1.74 (0.89)	0.36—3.68	NA	NA
HIV prevalence (%)	13.80 (4.32)	4.21—27.44	1631	11705

n (numerator), N (denominator), range (minimum—maximum), sd (standard deviation, sum (total), NA (not applicable), km (kilometre). Numerator and denominators for TB notifications and TB prevalence limited to adults.

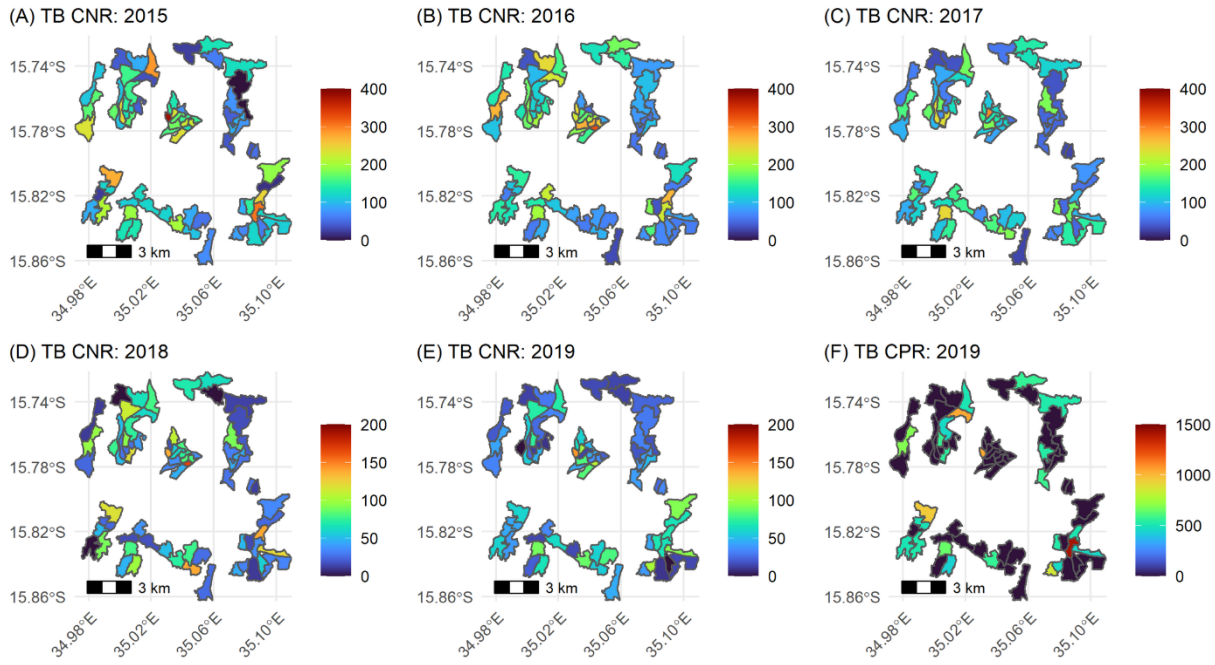
†All forms of TB

Empirical neighbourhood adult bacteriologically-confirmed TB cases notification and prevalence rates

During 2015-19 there were a total of 2,078 adults aged 18 years or older with bacteriologically - confirmed TB registered for treatment from the study neighbourhoods. CNR's declined during this period; neighbourhood mean adult bacteriologically-confirmed TB CNR's from 2015 to 2019 were 131 (range: 0-383, sd: 77), 134 (range: 32-328, sd:70), 114 (range: 28-291, sd:57), 56 (range: 0-167, sd:38) and 46 (0-144, sd:29) per 1000,000 adults Table 5.1. HIV prevalence among TB cases was 65.45 % (3394/5260) and 64.56 % (1864/5260) of registered patients were men.

A total of 29 (range: 0-3, sd: 0.6) bacteriologically-confirmed previously undiagnosed adult TB patients were identified during the prevalence survey, with an empirical neighbourhood mean of bacteriologically-confirmed TB prevalence rate of 215 (29/13,490) per 100,000 (range: 0-1,415, sd: 335). The rates of empirical TB notifications were higher in the city centre and were lower in the city's outskirts, whereas empirical TB prevalence rates were higher on the outskirts of the city Fig 5.2.

Figure 5.2: Empirical bacteriologically-confirmed adult TB case notification rates (CNR) 2015-2019 (A-E), and TB case prevalence rates (CPR) 2019 (F); both per 100,000



Neighbourhood-level predicted TB prevalence and notifications rates

For analysis of microbiologically-confirmed adult TB CNRs, the model that was selected included a neighbourhood-level random intercept, and neighbourhood-level covariates including the percentage of adults (≥15y), distance to the nearest TB clinic, and percentage of household heads who had not completed primary school education (5.S2 and 5.S6 Tables).

There was an overall trend of a reduction in rates of annual adult bacteriologically-confirmed TB notification rates. In comparison to 2019, the years 2015 (rate ratio [RR]: 2.89, 95% CrI: 2.48–3.37), 2016 (RR: 2.91, 95% CrI: 2.51–3.38), 2017 (RR: 2.51, 95% CrI: 2.16–2.92) and 2018 (RR: 1.23, 95% CrI: 1.03–1.45) had substantially higher CNRs (Table 5.2).

Table 5.2: Parameter estimates for selected regression models for predicting neighbourhood level TB prevalence and notifications.

	Adult bacteriologically-confirmed TB notification model		**Adult bacteriologically - confirmed TB prevalence model**	
Fixed effects **Parameters**	**Mean rate ratio**	**95% CrI**	**Mean rate ratio**	**95% CrI**
Percentage of adult residents (≥15y)[a]	0.96	(0.93, 1.00)	0.94	(0.80, 1.10)
Distance to nearest TB clinic (km)[a]	0.78	(0.69, 0.88)		
Percentage of household heads that did not complete primary school[a]	0.98	(0.96, 0.99)		
Year: 2019	Reference			
Year: 2015	2.89	(2.48, 3.37)		
Year: 2016	2.91	(2.51, 3.38)		
Year: 2017	2.51	(2.16, 2.92)		
Year: 2018	1.23	(1.03, 1.45)		
Intercept	$50.88*10^{-5}$	($42.99*10^{-5}$, $60.00*10^{-5}$)	$232.08*10^{-5}$	($132.05*10^{-5}$, $404.96*10^{-5}$)
Zero inflation intercept			0.18	(0.01, 0.46)
Random effects SD: cluster	0.31	(0.24, 0.39)	0.33	(0.01, 0.90)

km kilometre; sd standard deviation; CrI Credible interval.

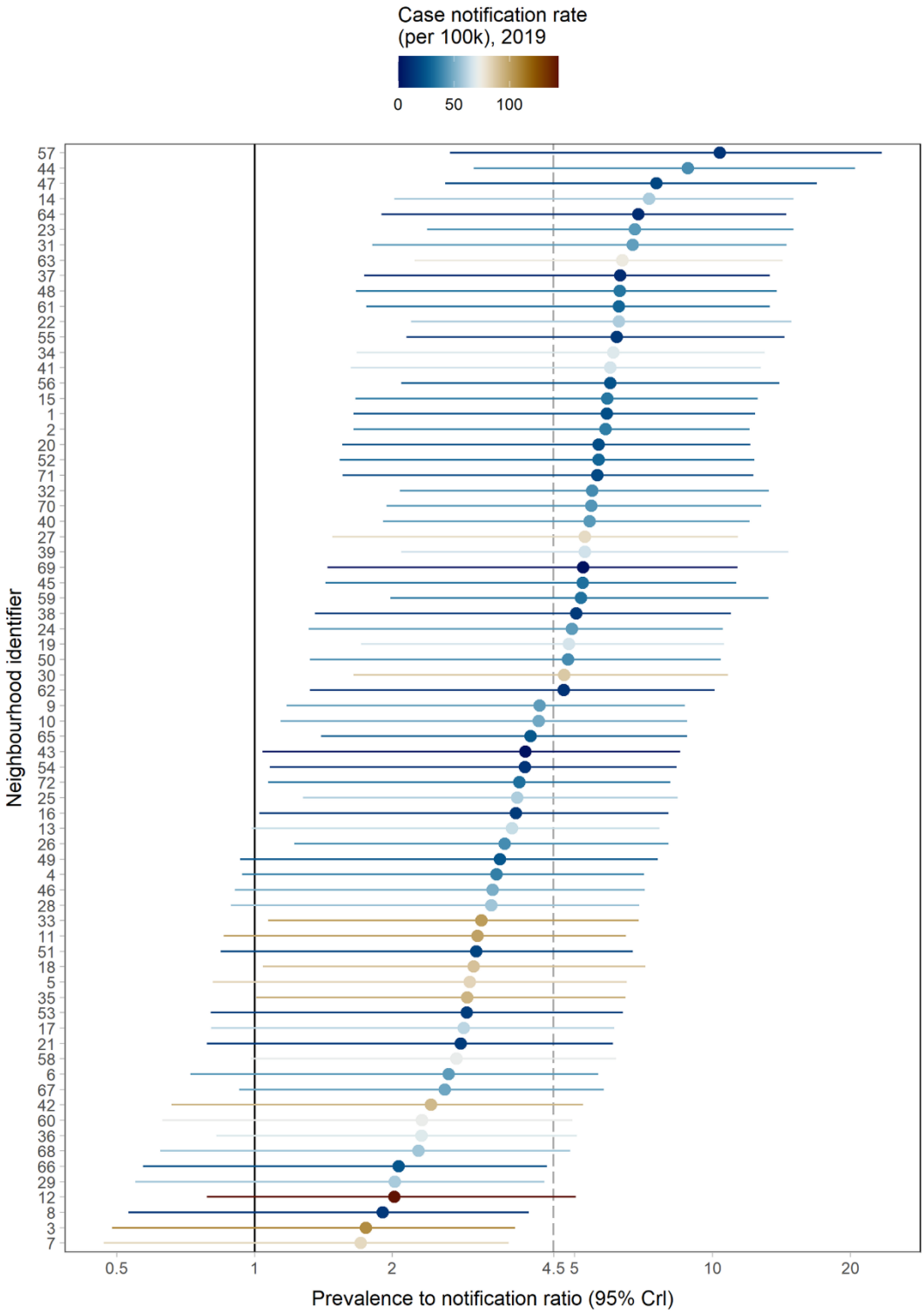
[a]Percentage of adults was centred by subtracting by its mean (60.90 %), Distance to nearest TB clinic (km) was centred by subtracting by 1km, Percentage of household head that did not complete primary school was centred by subtracting by its mean (16.90 %).

For the prevalence rates, the model with a random intercept for neighbourhood and neighbourhood percentage of adults (≥15y), was selected (5.S1 and 5.S5 Tables). In this model, there was no association between the neighbourhood TB prevalence rate and the percentage of neighbourhood adults (≥15y) (Table 5.2).

Neighbourhood level hotspots of TB underdiagnosis, ratio of prevalence to notifications

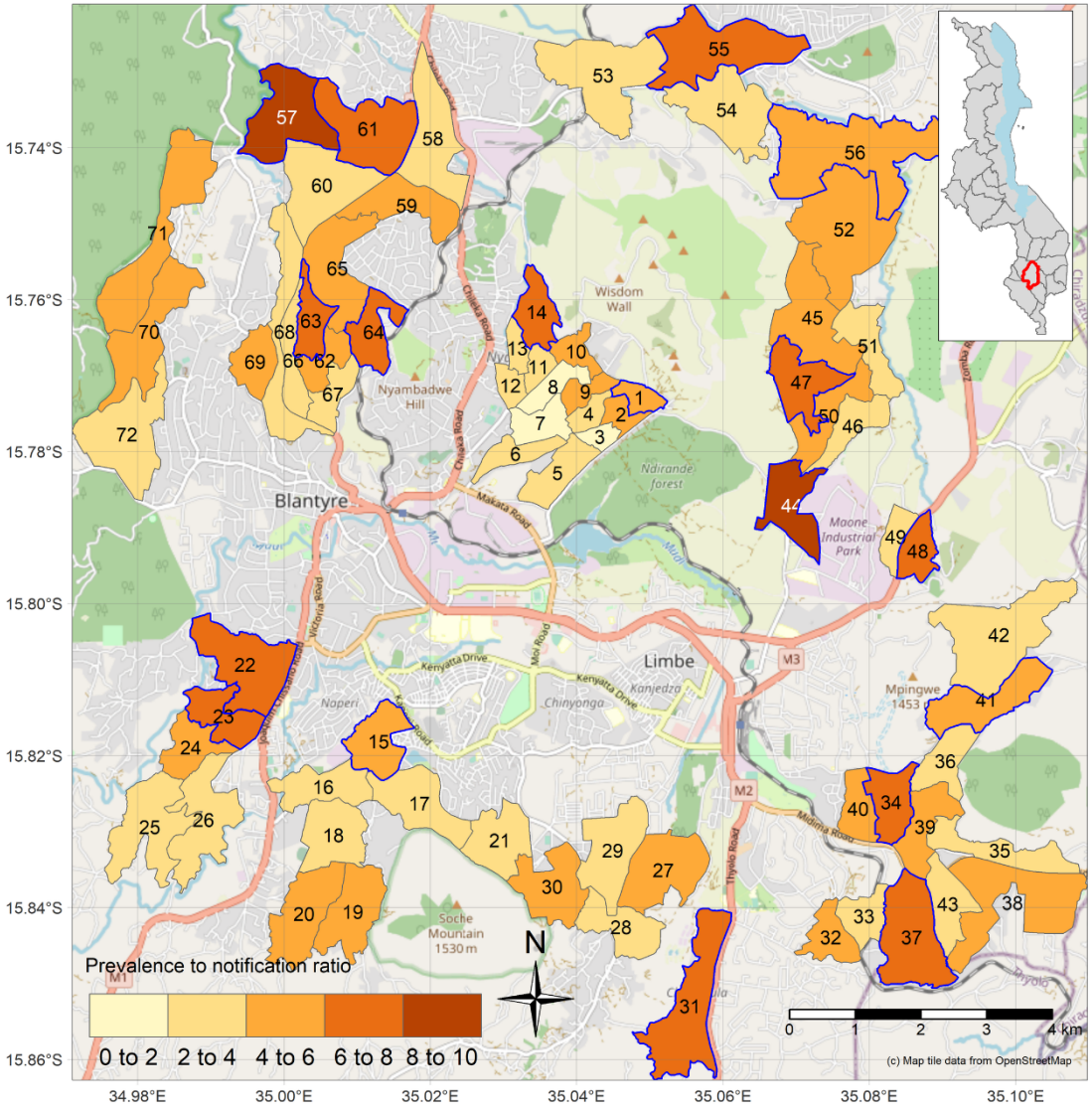
The mean neighbourhood posterior P:N ratios for adult bacteriologically-confirmed TB varied considerably between neighbourhoods (range: 1.70-10.40, sd: 1.79), with the mean posterior P:N ratio of the 72 neighbourhoods being 4.49 (95% CrI: 0.98-11.91) Figure 5.3.

Figure 5.3: Neighbourhood level TB prevalence to notification ratios (with 95% CrIs) using final models. The neighbourhoods were ordered according to prevalence to notification ratio size. The dashed line is the mean prevalence to notification ratio. CrI Credible interval.



Neighbourhoods with high-case notification rates and low P:N ratio were likely to have good access to TB diagnosis, and as a result, they were less likely to be hotspots of undiagnosed active TB. (Figure 5.3). Overall, P:N ratios were higher in neighbourhoods in the outskirts of the city, characterised by rapidly-growing informal settlements (Figure 5.4).

Figure 5.4: Map of TB prevalence to notification ratios predicted from final models including estimated neighbourhood random effects (Inset map of Malawi with Blantyre in red). Models include neighbourhood random effects. Neighbourhoods outlined in blue are in the highest quartile for P:N ratios. Inset map of Malawi with Blantyre District in red. Map tile data from OpenStreetMap.



Sensitivity analysis

The model coefficients were mostly similar to that in the primary analysis (5.S7 and 5.S8 Tables). The mean posterior P:N ratio of the 72 neighbourhoods was 5.04 (95% CrI: 1.86-10.26) and was 1.39 (95% CrI: 0.30-3.58) for the post stratified TB prevalence base analysis and the analysis based on all notified TB cases respectively. Overall, the distribution of the P:N ratios was similar to the primary analysis (5.S1-5.S5 Figures). The sensitivity analysis of all TB case notifications classified 14 neighbourhoods into the 4th quartile of the 18 neighbourhoods that were classified as having 4th quartile P:N ratios by the primary analysis, while the analysis based on post-stratified TB prevalence classified 16 neighbourhoods into 4th quartile of the 18 neighbourhoods that were classified as having 4th quartile P:N ratio by the primary analysis (5.S9 Table). Similarly to the primary analysis, the P:N ratios were higher in the outskirts of the city and in the informal urban settlement areas (5.S3 and 5.S5 Figures).

Discussion

Our main finding was that urban Blantyre in Malawi, a city with free health services and high coverage of treatment for HIV but high rates of poverty, has a significant burden of delayed and undiagnosed TB, with an overall estimated mean neighbourhood TB P:N ratio of 4.49:1. The 18/72 neighbourhoods with P:N ratios in the highest quartile were likely to be "hotspots" of delayed diagnosis or missed diagnosis of TB cases. Most neighbourhoods with high P:N ratios were located in the rapidly-growing informal settlements on the periphery of the city and were adjacent to forest reserves or mountainous terrain where the city is expanding. However, we did not directly capture informality, and further research is required to quantitatively define "the degree of formality of settlement"; these are residential areas that are not registered by the authorities or have makeshift housing structures and are likely to be important areas of focus for TB case finding activities. Our approach, which uses data that can be collected by District Health Officers in urban African cities (neighbourhood distance to clinic, percentage of male residents, and percentage of households where the head has not completed primary education), could be used to prioritise neighbourhoods for community-based TB ACF and prevention interventions, potentially a more efficient and effective way of delivering TB screening and prevention interventions [32].

A major strength of this analysis was that we used data from an enhanced TB surveillance system and a well-conducted subdistrict area prevalence survey. We used a systematic modelling strategy to identify a parsimonious model with variables that are predictive of neighbourhood TB notification and prevalence rates. Our strategy of combining notification and prevalence data offers important improvements in identifying high-burden areas compared to

approaches based solely on notification rates [11,12]. When only notifications are used, we might erroneously identify areas with easier access to TB diagnosis as hotspots and miss areas with a high burden of undiagnosed TB that have more limited access to TB diagnosis [17]. By using both notification and prevalence we can identify areas where the underlying notification of TB cases misses a higher proportion of undiagnosed TB cases [17]. But, by using the covariates identified in our models, our hope is that a new setting with similar characteristics to Blantyre might successfully rank its neighbourhoods to identify areas likely to have underdiagnosis. Hence our approach may reduce the need to carry out a relatively expensive across-the-board prevalence survey [33].

The End TB goal of 90 percent reduction in TB incidence between 2015 and 2035 is difficult to track in the majority of high-TB settings where TB incidence cannot be estimated directly from notifications but is estimated through inference methods that can produce imprecise estimates [1,2,34]. Developing methods for identifying areas of TB underdiagnosis is therefore critical to guide targeted intervention [12,35,36], which will be more important as epidemics become more concentrated. CNRs in Blantyre Malawi declined between 2015 to 2019, reflecting the general trend in other African countries in this period [1,2]. Blantyre has also achieved high coverage of HIV treatment and isoniazid preventive therapy [21,37].

The P:N ratio, which is used to assess TB burden in this study, is a proxy for the time between onset of TB infectiousness and diagnosis, and it is typically given the unit of years in mathematical modelling studies; its inverse is known as the patient diagnostic rate [17,38]. While the P:N ratio does not directly measure incidence, it can be used as an indicator of neighbourhood delayed TB case diagnoses or missed diagnosis [17]. We found overall across

Blantyre that the P:N ratio was high (4.49:1), indicating substantial underdiagnosis and delayed diagnosis of TB.

Our data suggests that the efficiency of selecting communities for community-based interventions, such as those based on poverty and population density, could be improved by incorporating neighbourhood-level prevalence survey and notification data which allows a more granular understanding of TB epidemiology [3,9,32,39]. District TB programmes benefit the most when they have information that enable them to prioritise their efforts, because they usually operate under resource constraints [3,39,40]. TB programmes must collect additional data to what is routinely collected through national TB case notification systems at the time of registering patients to gain a better understanding of the local TB epidemiology and guide public health interventions [15,32,36]. The majority of people who are disproportionately affected by TB, live in informal urban settlements that lack postal or zip codes [1,11,19]. Collection of neighbourhood location of TB patients' households may provide additional epidemiological insights for planning spatially-targeted interventions of TB hotspots [12,13,36,40,41].

Community-based TB ACF interventions have been shown to be most effective when conducted intensively, through repeated screening of communities over a short period of time [7,8]. For example, the ACT3 trial in Vietnam, conducted between 2014 and 2016, demonstrated reductions in TB prevalence through community ACF by offering annual Xpert MTB/Rif screening in 60 intervention neighbourhoods with approximately 54,000 adults over three years [8]. In a post intervention prevalence survey the intervention clusters showed a 44% reduction in TB prevalence [8]. TB screening interventions such as those tested in the ACT3 study would

be logistically and financially challenging in a setting like urban Blantyre, where health systems budgets are severely constrained [20,22]. Mathematical modelling work in Rio de Janeiro demonstrated that reducing TB transmission within TB hotspot neighbourhoods could reduce the city-wide TB incidence [13].

Both the distance to the nearest TB clinic and the percentage of household heads who did not finish primary school were positively associated with lower case notification rates. Both the notification and prevalence rate models had random intercept standard deviations with lower bounds of their 95% credible intervals that were well away from zero, indicating that TB epidemiology varied by city neighbourhood even after adjusting for covariates. We considered modelling neighbourhood random effects with the spatial ICAR term (5.S1 and 5.S2 Equations and 5.S3 and 5.S4 Tables), but models with random intercept terms gave a better description of the data because the spatial proximity of neighbourhoods with very different historical levels of interventions caused the ICAR model to over-smooth the data (5.S5 and 5.S6) Tables. In addition, we did not have enough data to support both random effects in the same model.

Our analysis had some limitations. The prevalence survey detected a low number of prevalent cases in the city. This meant that the models for prevalence had less power for identifying predictive covariates. Other work modelling regional TB prevalence has also found most potential covariates were unable to improve predictions [33]. The prevalence survey was also just done once; it is possible that if we had a repeat prevalence survey, we could have found a different distribution of TB prevalence than what was captured in this study [42]. Some groups were also under sampled by the prevalence trial, particularly men; we accounted for this in our post-stratified sensitivity analysis, showing similar results. In addition, P:N ratios based on

microbiologically-confirmed notified TB had higher mean rate ratios than in the sensitivity analysis that included all forms of TB, although the neighbourhoods identified with high P:N ratios were similar. The lower bound of the 95% CrI of some of the neighbourhood P:N ratios were less than one (Figure 3), although this represents low statistical power rather than a possibility of "overdiagnosis" of TB. More details on the goodness of fit of the models have been provided in the supplementary section (5.S3 Equation), which suggests that our models sufficiently describe the data well. Antiretroviral therapy (ART) coverage data was also not included in the model but there was a high percentage coverage of ART for people living with HIV across all neighbourhoods (mean: 95.41 %, range: 84.01- 95.41, sd: 0.03); we instead included HIV prevalence as people living with HIV are still at an increased risk of TB compared to HIV negative individuals even when they are on ART [1].

Countries in WHO Africa region need to accelerate the rate of TB incidence reduction from the current rate of about 4% per year to at least 10% by 2025 in order to meet the End TB goals [1,2,43]. For this to be achievable it is important that we have effective methods for prioritising communities for TB interventions to efficiently use the available resources [3]. National TB programmes that need to prioritise neighbourhood areas for TB interventions such as ACF, can collect the variables identified by our method which will be used by the model to predict the P:N ratios. By focusing on underserved communities, this will ensure universal health coverage for communities that are underserved by facility-based health care. The P:N ratios should be interpreted alongside the CNRs to obtain the full nature of the epidemic i.e., Figure 3. There is also a need to externally validate the model, as well as to investigate the effectiveness of spatially targeted interventions in randomised controlled trials [36].

Conclusion

Using a citywide enhanced surveillance data and prevalence survey data, we developed a predictive model to prioritise neighbourhoods for TB case detection and prevention activities based on readily available local data. In most low-resource settings, current active case-finding strategies are inefficient and resource-intensive. We have demonstrated a method for identifying neighbourhoods with high rates of underdiagnosis. Researchers and programme managers could prioritise identified TB hotspots for TB control and prevention interventions to focus efforts on urban TB elimination.

Acknowledgments

McEwen Khundi (MK), Peter MacPherson (PM), Elizabeth Corbett (ELC) and James Carpenter (JC) designed the study. Lingstone Chiume (LC), Rebecca Nzawa Soko (RNS), Hussein Twabi (HT), MK, Helena R A Feasey (HRAF) and Rachael M Burke (RMB) contributed to the data collection and management. MK, PM, JC, RBM and ELC contributed to the data analysis. MK, PM, JC, Peter J Dodd (PJD), Katherine C Horton (KCH), Theodore H Cohen (THC), RBM and ELC contributed to the data interpretation. MK wrote the first draft. MK, PM, ELC, JC, THC, LC, RNS, HT, HRAF, RMB, PJD, KCH, THC and Marriott Nliwasa reviewed and approved the final draft. All authors read and approved the final manuscript. We thank the National TB Programme and Blantyre District Health Office for their support and encouragement.

Conflict of Interests: None

Availability of data and code

Data and analysis code is available at GitHub repository

https://github.com/mcewenkhundi/TBhotspotclustersplos

Supporting information

5.S1 Equation.

Let $Y \sim Pois(\mu)$

$$\Pr(Y_{ij} = y_{ij}) = \left(\frac{\mu_{ij}^{y_{ij}}\, exp(-\mu_{ij})}{y_{ij}!}\right)$$

$$\log(\mu_{ij}) = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} \ldots + \beta_k x_{ki} + \beta_{year2015} year2015_j + \beta_{year2016} year2016_j + \beta_{year2017} year2017_j + \beta_{year2018} year2018_j + \log(Pop_{ij}) + \phi_i$$

$$\alpha \sim Normal(\mu_\alpha = 0, \sigma_\alpha^2 = 10)$$

$$\beta_k \sim Normal(\mu_\beta = 0, \sigma_\beta^2 = 10)$$

$$\phi_i \mid \phi_k, i \neq k, \sim Normal\left(\frac{\sum_{i\sim k} w_{ik}\phi_i}{d_i}, \frac{\sigma_i^2}{d_i}\right)$$

$$\sigma_i \sim HalfCauchy(0, 1)$$

The expectation of $Y_{ij} = y_{ij}$ given by:

$$E(Y_{ij} = y_{ij}) = \mu_{ij}$$

Where i refers to neighbourhood i for i =1,2,3..72, and *j* indexes year (2015,…2019), W is a 72 by 72 adjacency matrix where entries {i,i} are 0 and the off-diagonal elements are 1 if regions i and k are neighbours and 0 otherwise. d_i is the number of neighbours for neighbourhood$_i$, d_i was fixed to be 4. ϕ_i , α and $\beta_1, \beta_2 \ldots \beta_k$ are unknown regression coefficients that are estimated from the data for the covariates $x_{1i}, x_{2i} \ldots x_{ki}$, and sigma (σ_i) is the standard deviation for the spatial random term ϕ_i. Pop_{ij} is the total population of neighbourhood that is used as the offset. Here $year2015_j$, $year2016_j$, $year2017_j$ and $year2018_j$ are dummy variables which

take the value of one for that year and take the value of zero for the other years (the baseline year was 2019).

S2 Equation.

To define the zero-inflated Poisson model, let $Z \sim Bern(1-p)$ (so $\Pr\{Z=0\}=p$) and independently $W \sim Pois(\mu)$. Then the data are modelled by the zero-inflated variable Y which is defined as

$$Y = ZW$$

$$\Pr(Y_i = y_i) = \begin{cases} \left(p + (1-p)exp(-\mu_i)\right) if\ y_i = 0 \\ \left((1-p)\frac{\mu_i^{y_i}\, exp(-\mu_i)}{y_i!}\right) if\ y_i \geq 1 \end{cases}$$

$$logit(p_i) = \vartheta$$

$$\vartheta \sim Normal\left(\mu = 0, \sigma_\vartheta^2 = 10\right)$$

$$\log(\mu_i) = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} \ldots + \beta_k x_{ki} + \log(Pop_i) + \phi_i$$

$$\alpha \sim Normal(\mu_\alpha = 0, \sigma_\alpha^2 = 10)$$

$$\beta_k \sim Normal\left(\mu_\beta = 0, \sigma_\beta^2 = 10\right)$$

$$\phi_i \mid \phi_k, i \neq k, \sim Normal\left(\frac{\sum_{i\sim k} w_{ik}\phi_i}{d_i}, \frac{\sigma_i^2}{d_i}\right)$$

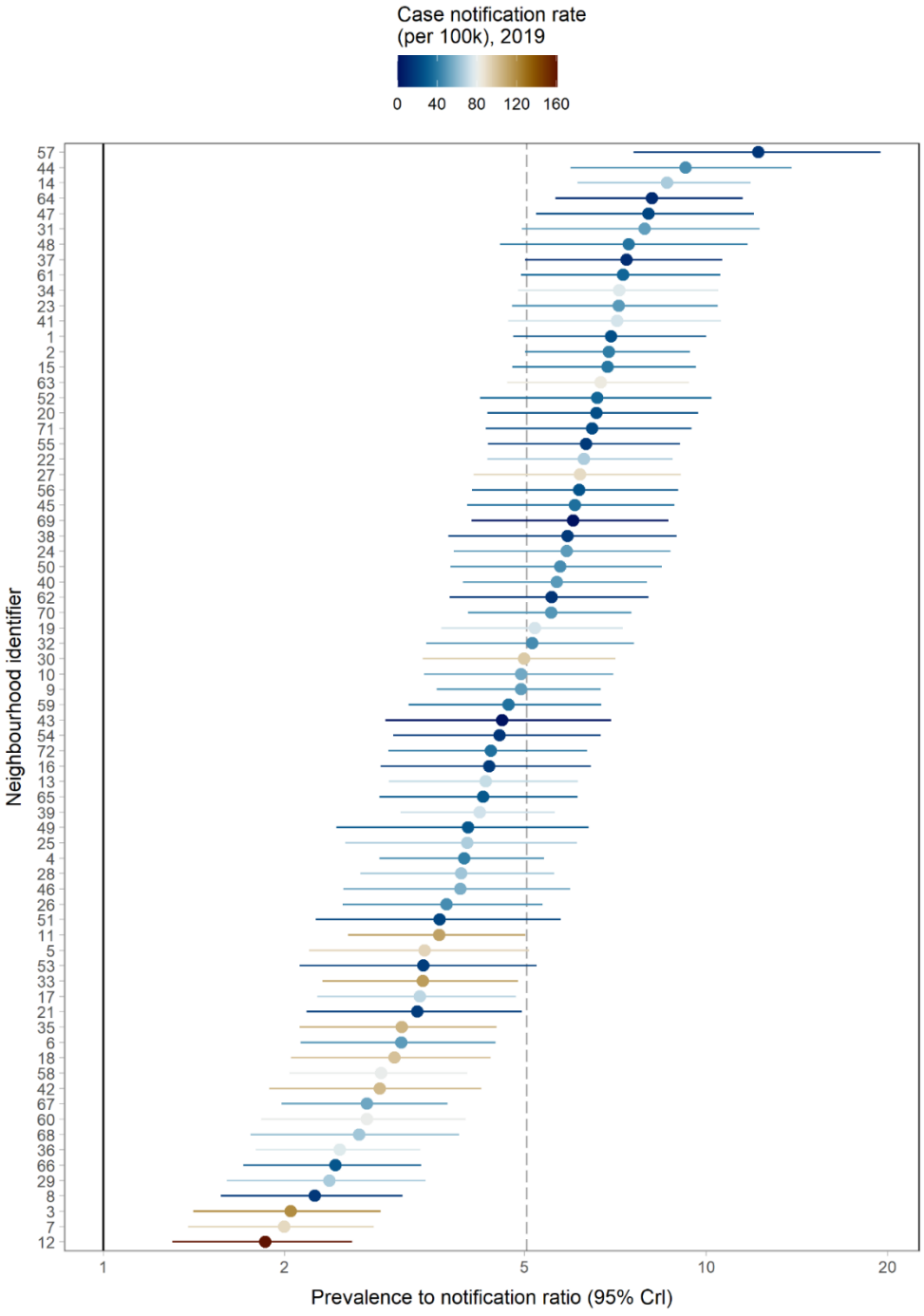
$$\sigma_i \sim HalfCauchy(0,1)$$

The expectation of $Y_i = y_i$ given by:

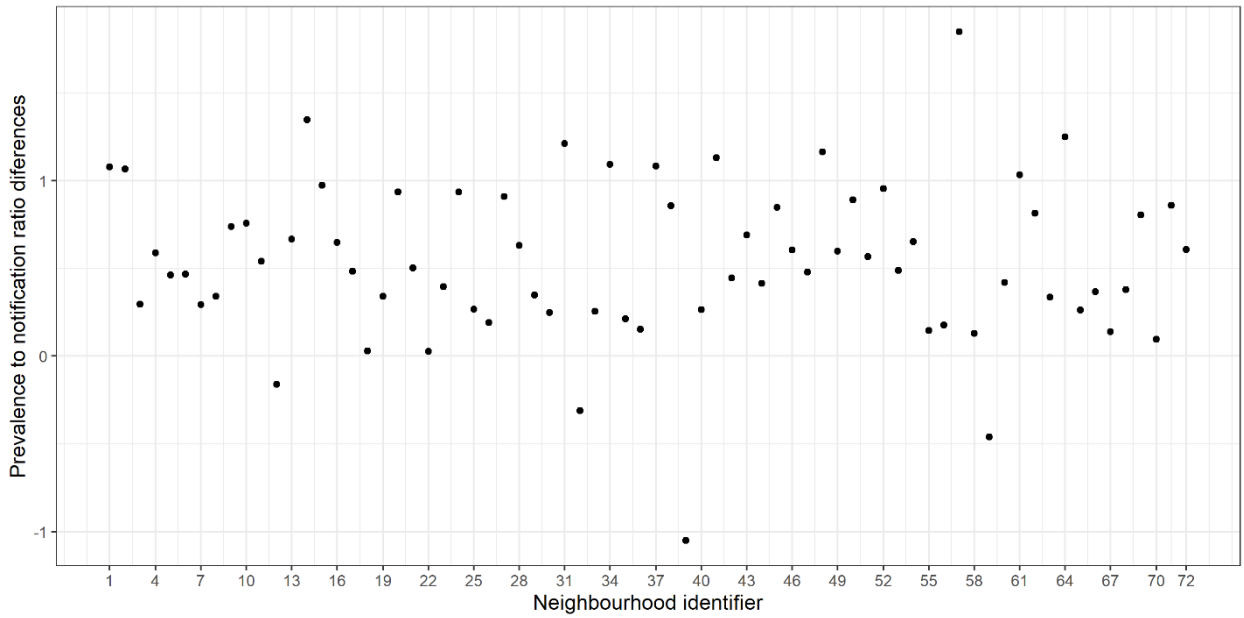
$$E(Y_i = y_i) = (1-p_i)\,\mu_i$$

Where i refers to neighbourhood i for i =1,2,3..72 (note we only have prevalence data for one year, 2019). W is a 72 by 72 adjacency matrix where entries {i,i} are 0 and the off-diagonal elements are 1 if regions i and k are neighbours and 0 otherwise. d_i is the number of neighbours for neighbourhood$_i$, d_i was fixed to be 4. ϕ_i, θ, α and $\beta_1, \beta_2 \ldots \beta_k$ are unknown regression coefficients that are estimated from the data for the covariates $x_{1i}, x_{2i} \ldots x_{ki}$, and sigma ($\sigma_i$) is the standard deviation for the spatial random term ϕ_i. Pop_i is the total population of neighbourhood that is used as the offset.

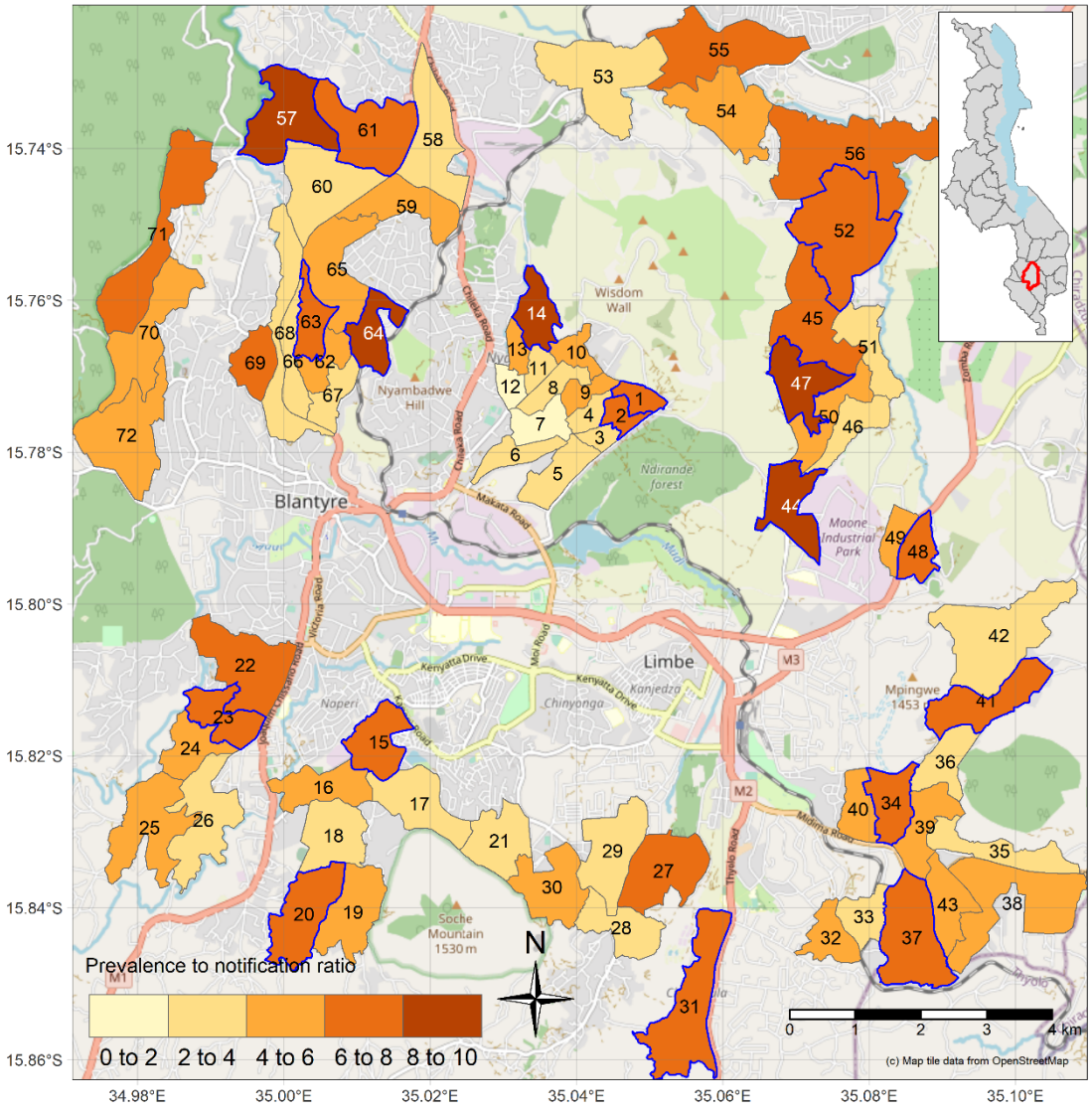
5.S1 Figure: Neighbourhood level TB prevalence to notification rate ratios (with 95% CIs) using final models. The neighbourhoods were ordered according to prevalence to notification ratio size. Analysis based on post stratified TB prevalence with microbiologically-confirmed TB notifications kept the same as in the primary analysis. The dashed line is the mean prevalence to notification ratio. Crl Credible interval.



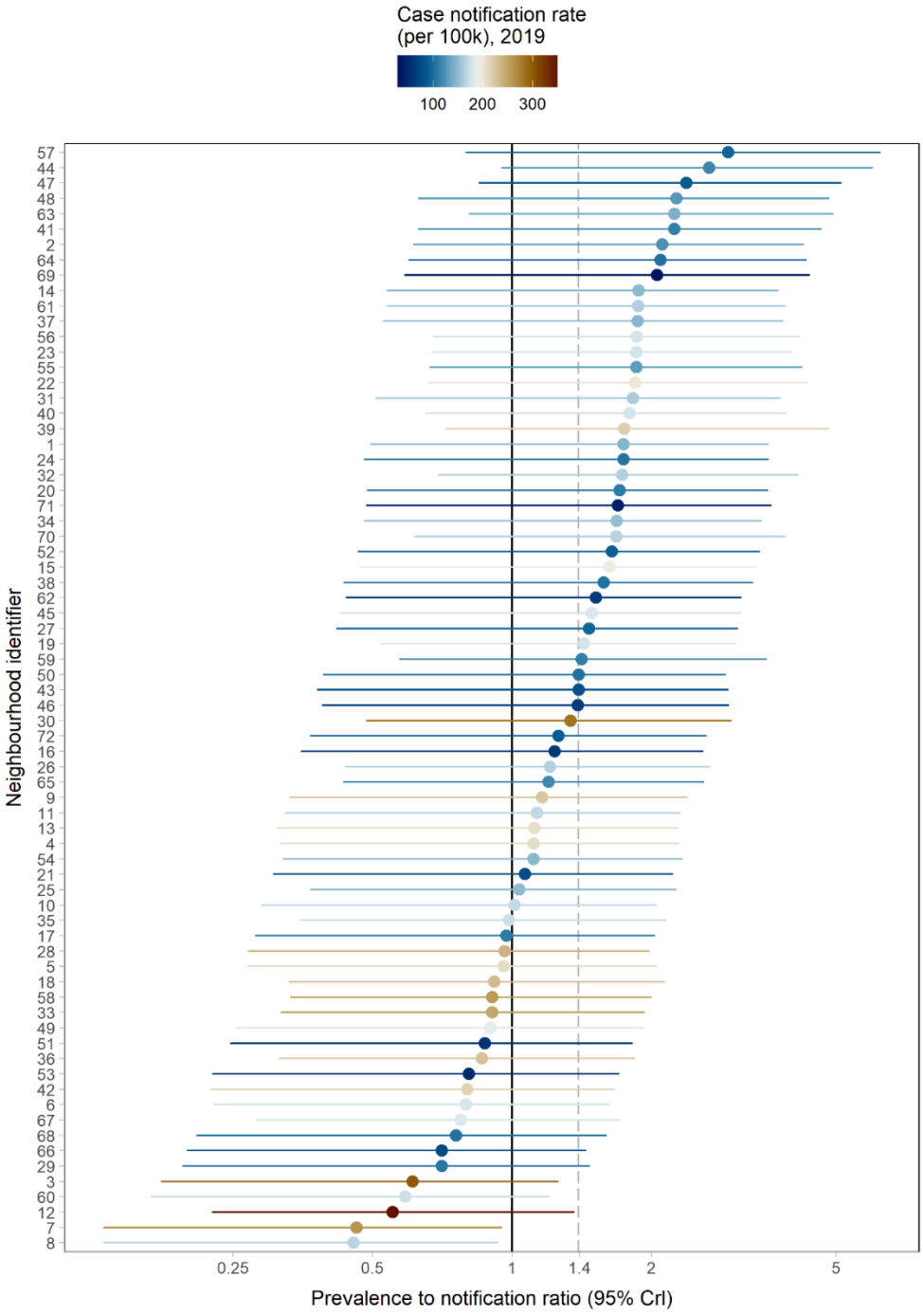
5.S2 Figure: Differences of neighbourhood level TB prevalence to notification rate ratios (P:N rate ratios), the P:N rate ratio based on post stratified prevalence TB subtracted by the P:N ratio based on the primary analysis.



5.S3 Figure: Map of TB prevalence to notification ratios predicted from final models (Inset map of Malawi with Blantyre in red). Analysis based on post stratified TB prevalence and with microbiologically-confirmed TB notifications kept the same as in the primary analysis. Models include neighbourhood random effects. Neighbourhoods outlined in blue are in the highest quartile for P:N ratios. Map tile data from OpenStreetMap.

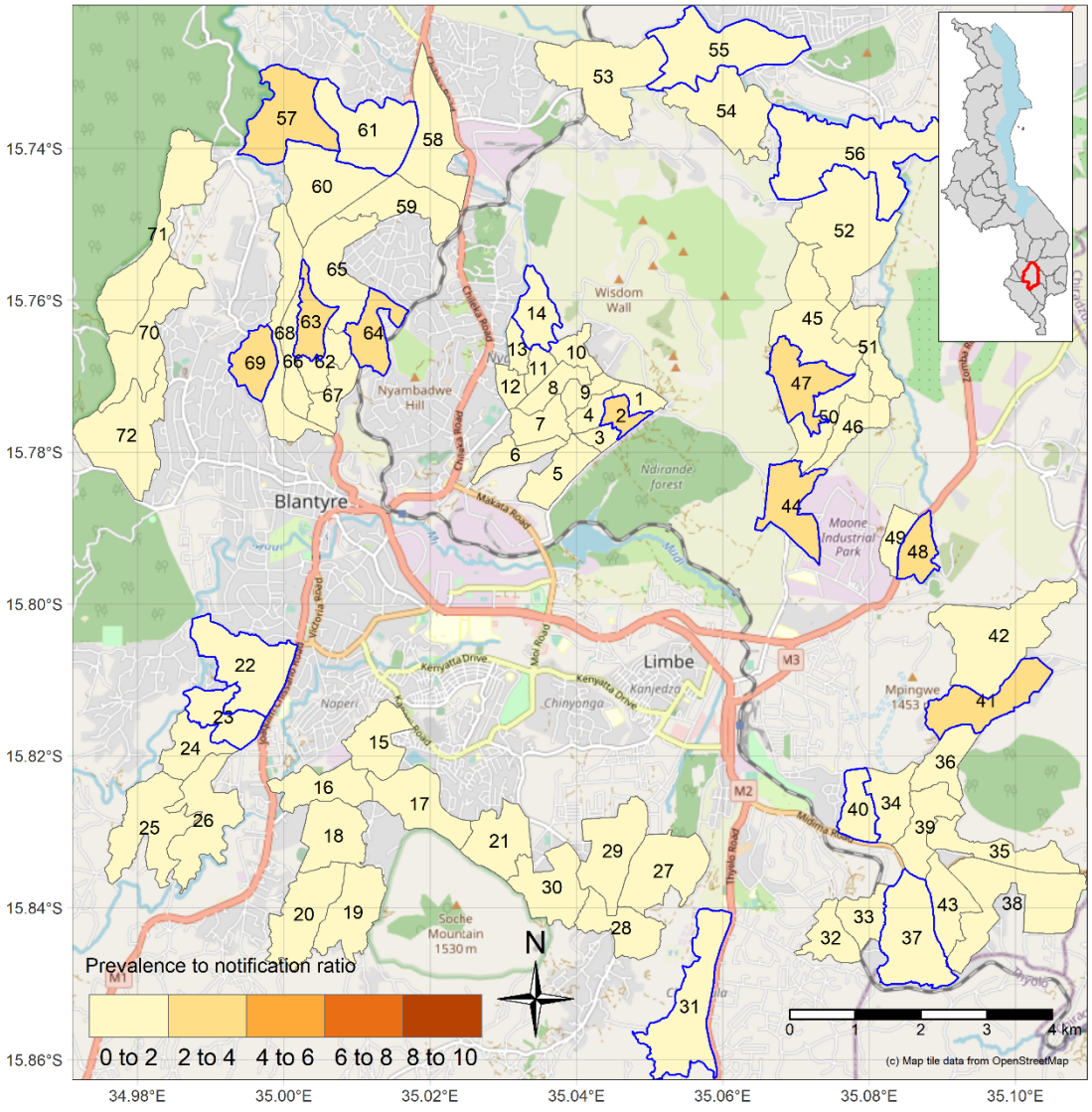


5.S4 Figure: Neighbourhood level TB prevalence to notification rate ratios (with 95% CIs) using final models. The neighbourhoods were ordered according to prevalence to notification ratio size. Analysis based on microbiologically-confirmed TB and clinically-diagnosed cases and with TB prevalence kept the same as in the primary analysis. The dashed line is the mean prevalence to notification ratio. CrI Credible interval.

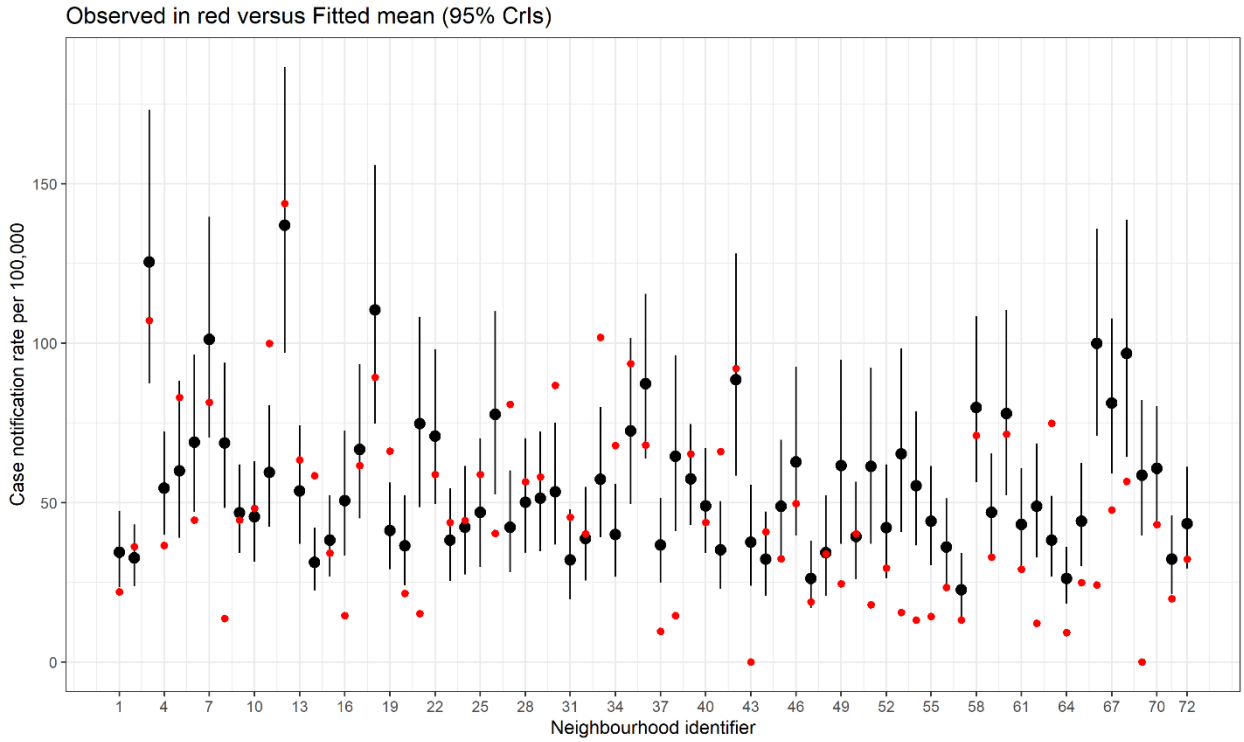


5.S5 Figure: Map of TB prevalence to notification ratios predicted from final models (Inset map of Malawi with Blantyre in red). Analysis based on microbiologically-confirmed TB and clinically-diagnosed cases and with TB prevalence kept the same as in the primary analysis.

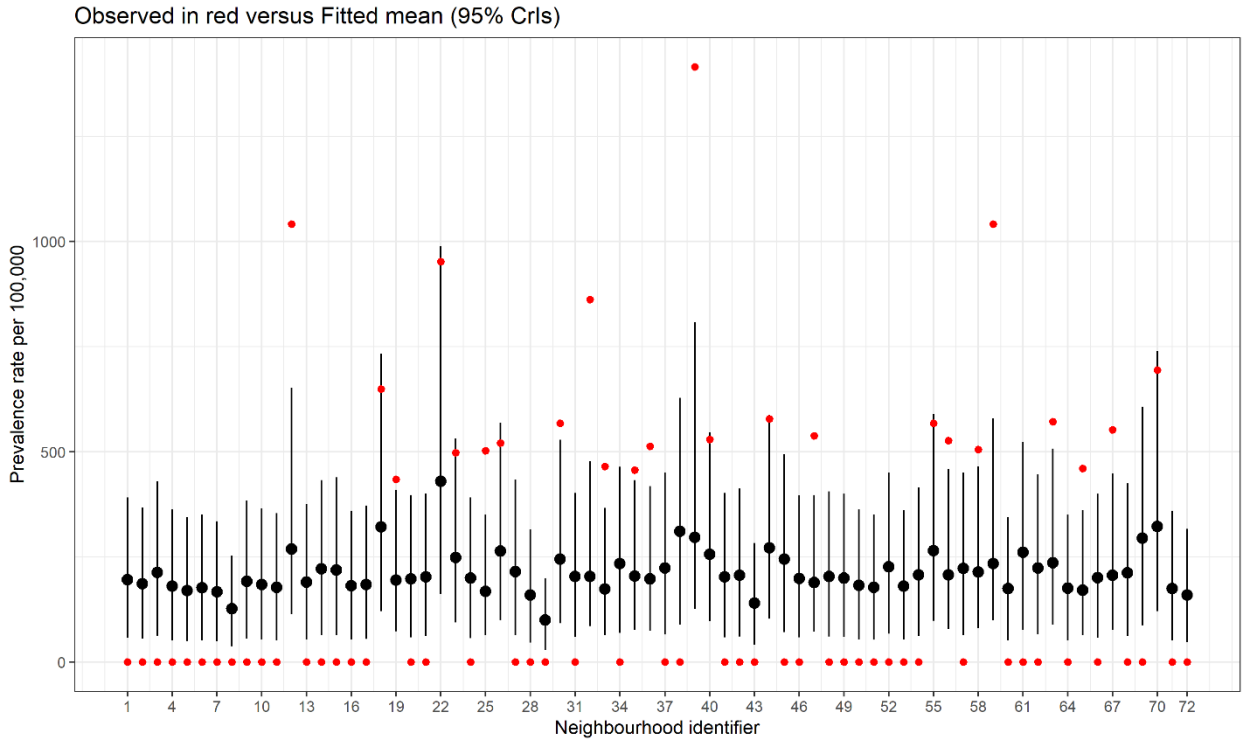
Models include neighbourhood random effects. Neighbourhoods outlined in blue are in the highest quartile for P:N ratios. Map tile data from OpenStreetMap.



5.S6 Figure: Observed versus predicted mean CNRs (95% CrIs). Analysis based on microbiologically-confirmed TB as in the primary analysis. CrI Credible interval.



5.S7 Figure: Observed versus predicted mean prevalence rates (95% CrIs). Analysis based on microbiologically-confirmed TB as in the primary analysis. CrI Credible interval.



5.S3 Equation. Model goodness of fit assessment.

We now present a chi-square-statistic goodness of fit statistic for the models as an approximate guidance to the goodness of fit of the models to the data.

Let Y_i be the observed number of cases and $\hat{\mu}_i$ the model predicated cases in neighbourhood *i* = 1, 2, 3, …72

$$\chi^2 = \sum_i \frac{(Y_i - \hat{\mu}_i)^2}{\hat{\mu}_i}$$

The value of chi-square was calculated separately for TB case notifications and also for prevalent TB cases. We also calculated the approximate degrees of freedom of each as

72 – number of parameters in each model.

The calculated chi-squared statistic of the case notifications was 73.06, while the degree of freedom was 64, giving an approximate p-value of 0.21

Similarly, the calculated chi-squared statistic of the prevalent cases was 56.81, while the degree of freedom was 70, giving an approximate p-value of 0.87.

While, because expectations in the 72 locations are often small, this only provides an approximate measure of goodness of fit, nevertheless taken together with our careful choice of models (e.g. zero-inflated Poisson for the prevalence) and systematic variable selection approach means are confident our models describe the data sufficiently well to justify their application.

5.S1 Table. Table of all the TB prevalence neighbourhood level models with a random intercept of clinic of treatment registration. Coefficients (mean rate ratio) were exponentiated and intercepts were multiplied by 100,000 (Equation 2).

Model names	Intercept	Percentage of adults (≥15y)	Percentage of household heads that did not complete primary school	Distance to nearest TB clinic (km)	Percentage of HIV prevalence	Percentage of male adults	Random effects SD: cluster	Probability Zi=0 (Refer to Equation 2)
prevalence model 1	218.46 (98.53-463.00)	0.90 (0.72-1.13)	0.99 (0.91-1.08)	0.98 (0.55-1.74)	1.03 (0.91-1.16)	0.92 (0.64-1.33)	0.39 (0.02-1.05)	0.19 (0.01-0.48)
prevalence model 2	210.26 (96.11-449.69)		1.01 (0.93-1.09)	1.08 (0.63-1.86)	1.02 (0.91-1.15)	0.98 (0.70-1.38)	0.38 (0.02-1.04)	0.19 (0.01-0.48)
prevalence model 3	220.73 (101.37-462.95)	0.91 (0.73-1.12)		0.99 (0.55-1.75)	1.02 (0.91-1.14)	0.92 (0.63-1.32)	0.37 (0.01-1.00)	0.19 (0.01-0.48)
prevalence model 4	212.01 (97.42-448.19)			1.09 (0.64-1.85)	1.03 (0.92-1.15)	0.98 (0.71-1.36)	0.36 (0.02-0.99)	0.19 (0.01-0.47)
prevalence model 5	217.90 (116.83-396.19)	0.90 (0.73-1.11)	0.99 (0.91-1.08)		1.03 (0.91-1.15)	0.92 (0.64-1.32)	0.37 (0.02-1.01)	0.19 (0.01-0.47)
prevalence model 6	226.71 (125.32-406.29)		1.01 (0.94-1.09)		1.02 (0.91-1.15)	0.99 (0.73-1.36)	0.36 (0.01-0.97)	0.19 (0.01-0.47)
prevalence model 7	220.83 (119.92-398.84)	0.91 (0.75-1.10)			1.02 (0.91-1.14)	0.92 (0.64-1.31)	0.36 (0.01-0.98)	0.18 (0.01-0.47)
prevalence model 8	230.62 (129.55-405.77)				1.02 (0.92-1.14)	1.00 (0.75-1.34)	0.34 (0.02-0.94)	0.18 (0.01-0.47)
prevalence model 9	223.22 (104.13-461.02)	0.91 (0.72-1.12)	1.00 (0.92-1.08)	0.97 (0.54-1.73)		0.89 (0.64-1.24)	0.37 (0.01-1.02)	0.19 (0.01-0.47)
prevalence model 10	214.86 (102.66-445.87)		1.01 (0.94-1.09)	1.07 (0.62-1.80)		0.95 (0.71-1.28)	0.36 (0.01-0.99)	0.18 (0.01-0.47)
prevalence model 11	226.52 (105.67-462.82)	0.91 (0.74-1.10)		0.97 (0.55-1.71)		0.89 (0.65-1.23)	0.36 (0.02-0.98)	0.18 (0.01-0.47)

Model names	Intercept	Percentage of adults (≥15y)	Percentage of household heads that did not complete primary school	Distance to nearest TB clinic (km)	Percentage of HIV prevalence	Percentage of male adults	Random effects SD: cluster	Probability Zi=0 (Refer to Equation 2)
prevalence model 12	216.23 (103.84-443.79)			1.08 (0.65-1.80)		0.95 (0.71-1.27)	0.35 (0.01-0.97)	0.18 (0.01-0.46)
prevalence model 13	221.57 (120.89-394.66)	0.91 (0.74-1.11)	1.00 (0.92-1.08)			0.89 (0.65-1.23)	0.36 (0.01-0.98)	0.18 (0.01-0.46)
prevalence model 14	229.64 (128.60-402.07)		1.01 (0.95-1.09)			0.96 (0.74-1.26)	0.35 (0.01-0.93)	0.18 (0.01-0.46)
prevalence model 15	225.02 (125.63-394.43)	0.91 (0.75-1.09)				0.89 (0.65-1.22)	0.35 (0.01-0.94)	0.18 (0.01-0.46)
prevalence model 16	233.95 (132.60-407.26)					0.97 (0.75-1.26)	0.33 (0.01-0.91)	0.18 (0.01-0.46)
prevalence model 17	223.37 (103.70-459.25)	0.92 (0.74-1.12)	0.99 (0.91-1.08)	0.97 (0.55-1.71)	1.04 (0.93-1.16)		0.36 (0.01-1.00)	0.19 (0.01-0.47)
prevalence model 18	214.82 (102.48-438.95)		1.01 (0.93-1.08)	1.07 (0.64-1.76)	1.03 (0.92-1.14)		0.35 (0.01-0.98)	0.18 (0.01-0.47)
prevalence model 19	227.46 (108.74-461.39)	0.93 (0.77-1.11)		0.97 (0.55-1.68)	1.03 (0.93-1.14)		0.35 (0.01-0.95)	0.18 (0.01-0.46)
prevalence model 20	216.39 (105.62-431.03)			1.08 (0.66-1.74)	1.03 (0.93-1.13)		0.35 (0.01-0.95)	0.18 (0.01-0.47)
prevalence model 21	220.72 (122.28-392.12)	0.92 (0.76-1.10)	0.99 (0.91-1.07)		1.04 (0.93-1.15)		0.35 (0.01-0.97)	0.18 (0.01-0.46)
prevalence model 22	228.49 (128.32-398.52)		1.01 (0.94-1.08)		1.02 (0.92-1.12)		0.35 (0.01-0.95)	0.18 (0.01-0.46)
prevalence model 23	224.93 (125.81-395.68)	0.93 (0.79-1.09)			1.03 (0.93-1.14)		0.34 (0.01-0.92)	0.18 (0.01-0.46)
prevalence model 24	233.05 (133.04-405.34)				1.02 (0.93-1.12)		0.33 (0.01-0.91)	0.18 (0.01-0.46)

Model names	Intercept	Percentage of adults (≥15y)	Percentage of household heads that did not complete primary school	Distance to nearest TB clinic (km)	Percentage of HIV prevalence	Percentage of male adults	Random effects SD: cluster	Probability Zi=0 (Refer to Equation 2)
prevalence model 25	236.34 (112.67-479.93)	0.93 (0.77-1.13)	1.00 (0.93-1.08)	0.94 (0.54-1.61)			0.36 (0.01-0.97)	0.19 (0.01-0.47)
prevalence model 26	225.87 (111.64-445.27)		1.01 (0.94-1.08)	1.02 (0.63-1.64)			0.35 (0.01-0.95)	0.18 (0.01-0.46)
prevalence model 27	240.36 (117.23-479.19)	0.93 (0.78-1.12)		0.94 (0.54-1.61)			0.34 (0.01-0.94)	0.18 (0.01-0.47)
prevalence model 28	226.74 (112.73-448.23)			1.04 (0.65-1.65)			0.34 (0.01-0.92)	0.18 (0.01-0.46)
prevalence model 29	227.68 (125.87-401.40)	0.94 (0.79-1.12)	1.00 (0.93-1.08)				0.35 (0.01-0.95)	0.18 (0.01-0.46)
prevalence model 30	233.71 (133.43-407.32)		1.01 (0.94-1.08)				0.33 (0.01-0.90)	0.18 (0.01-0.46)
prevalence model 31	232.08 (132.05-404.96)	0.94 (0.80-1.10)					0.33 (0.01-0.90)	0.18 (0.01-0.46)
prevalence model 32	237.63 (136.98-410.81)						0.32 (0.01-0.89)	0.17 (0.01-0.46)

5.S2 Table. Table of all the TB notified neighbourhood level models with a random intercept of clinic of treatment registration. Coefficients(mean rate ratio) were exponentiated and intercepts were multiplied by 100,000 (Equation 1).

Model names	Intercept	Percentage of adults (≥15y)	Percentage of household heads that did not complete primary school	Distance to nearest TB clinic (km)	Percentage of HIV prevalence	Percentage of male adults	Year: 2015	Year: 2016	Year: 2017	Year: 2018	Random effects SD: cluster
notification model 1	50.85 (42.97-59.94)	0.96 (0.92-1.00)	0.98 (0.96-0.99)	0.78 (0.69-0.88)	0.99 (0.97-1.02)	0.98 (0.91-1.05)	2.89 (2.48-3.36)	2.91 (2.51-3.38)	2.51 (2.16-2.92)	1.22 (1.03-1.45)	0.32 (0.24-0.40)
notification model 2	49.12 (41.57-57.84)		0.98 (0.97-1.00)	0.82 (0.73-0.91)	0.99 (0.97-1.01)	1.00 (0.94-1.07)	2.89 (2.49-3.37)	2.91 (2.51-3.38)	2.51 (2.16-2.92)	1.22 (1.03-1.45)	0.33 (0.25-0.41)
notification model 3	50.98 (42.89-60.45)	0.98 (0.95-1.02)		0.78 (0.69-0.88)	0.98 (0.96-1.01)	0.98 (0.91-1.05)	2.89 (2.48-3.37)	2.91 (2.51-3.38)	2.51 (2.16-2.92)	1.22 (1.03-1.45)	0.33 (0.26-0.42)
notification model 4	50.11 (42.33-59.15)			0.80 (0.71-0.89)	0.98 (0.96-1.01)	0.99 (0.93-1.06)	2.89 (2.49-3.37)	2.92 (2.52-3.39)	2.51 (2.16-2.93)	1.23 (1.04-1.45)	0.33 (0.26-0.42)
notification model 5	42.17 (36.26-48.91)	1.00 (0.95-1.04)	0.98 (0.96-1.00)		1.00 (0.97-1.02)	0.97 (0.90-1.05)	2.89 (2.48-3.36)	2.91 (2.51-3.38)	2.51 (2.16-2.92)	1.23 (1.03-1.45)	0.37 (0.29-0.46)
notification model 6	42.23 (36.36-49.07)		0.98 (0.96-0.99)		1.00 (0.97-1.02)	0.97 (0.91-1.04)	2.89 (2.49-3.36)	2.91 (2.51-3.38)	2.51 (2.17-2.92)	1.23 (1.04-1.45)	0.37 (0.29-0.46)
notification model 7	42.21 (36.14-49.03)	1.02 (0.98-1.06)			0.99 (0.96-1.01)	0.97 (0.90-1.05)	2.89 (2.49-3.37)	2.91 (2.51-3.38)	2.51 (2.17-2.92)	1.23 (1.04-1.45)	0.39 (0.31-0.48)

Model names	Intercept	Percentage of adults (≥15y)	Percentage of household heads that did not complete primary school	Distance to nearest TB clinic (km)	Percentage of HIV prevalence	Percentage of male adults	Year: 2015	Year: 2016	Year: 2017	Year: 2018	Random effects SD: cluster
notification model 8	42.32 (36.38-49.15)				0.99 (0.96-1.01)	0.95 (0.89-1.02)	2.89 (2.49-3.36)	2.91 (2.51-3.38)	2.51 (2.17-2.92)	1.22 (1.04-1.45)	0.39 (0.31-0.48)
notification model 9	50.76 (42.87-59.71)	0.96 (0.92-1.00)	0.98 (0.96-0.99)	0.78 (0.70-0.88)		0.98 (0.92-1.05)	2.89 (2.49-3.37)	2.91 (2.51-3.38)	2.51 (2.16-2.92)	1.23 (1.04-1.45)	0.31 (0.24-0.40)
notification model 10	48.99 (41.45-57.62)		0.98 (0.97-1.00)	0.82 (0.73-0.92)		1.01 (0.95-1.07)	2.89 (2.49-3.37)	2.91 (2.51-3.39)	2.51 (2.16-2.92)	1.22 (1.03-1.45)	0.32 (0.25-0.41)
notification model 11	50.77 (42.73-60.12)	0.98 (0.95-1.02)		0.78 (0.69-0.89)		0.99 (0.93-1.06)	2.89 (2.49-3.37)	2.91 (2.51-3.38)	2.51 (2.16-2.92)	1.23 (1.03-1.45)	0.34 (0.26-0.42)
notification model 12	49.89 (42.22-58.85)			0.80 (0.72-0.89)		1.00 (0.94-1.07)	2.89 (2.49-3.36)	2.91 (2.50-3.38)	2.51 (2.17-2.92)	1.22 (1.04-1.45)	0.34 (0.26-0.42)
notification model 13	42.19 (36.24-48.95)	1.00 (0.96-1.04)	0.98 (0.96-0.99)			0.97 (0.90-1.04)	2.89 (2.49-3.37)	2.91 (2.51-3.38)	2.51 (2.16-2.92)	1.23 (1.04-1.45)	0.37 (0.29-0.46)
notification model 14	42.21 (36.26-48.83)		0.98 (0.96-0.99)			0.97 (0.92-1.04)	2.89 (2.49-3.37)	2.91 (2.52-3.38)	2.51 (2.17-2.92)	1.23 (1.04-1.45)	0.36 (0.29-0.45)
notification model 15	42.24 (36.28-	1.02 (0.98-1.06)				0.98 (0.91-1.06)	2.89 (2.49-3.37)	2.92 (2.52-3.39)	2.51 (2.16-2.91)	1.23 (1.04-1.45)	0.38 (0.31-0.48)

Model names	Intercept	Percentage of adults (≥15y)	Percentage of household heads that did not complete primary school	Distance to nearest TB clinic (km)	Percentage of HIV prevalence	Percentage of male adults	Year: 2015	Year: 2016	Year: 2017	Year: 2018	Random effects SD: cluster
	49.13)										
notification model 16	42.33 (36.34-49.19)					0.96 (0.90-1.03)	2.89 (2.49-3.36)	2.91 (2.51-3.38)	2.51 (2.17-2.91)	1.22 (1.04-1.45)	0.38 (0.31-0.48)
notification model 17	50.89 (43.09-59.87)	0.97 (0.93-1.00)	0.98 (0.96-0.99)	0.78 (0.69-0.87)	1.00 (0.98-1.02)		2.89 (2.49-3.36)	2.91 (2.51-3.38)	2.51 (2.17-2.91)	1.23 (1.03-1.45)	0.31 (0.24-0.40)
notification model 18	49.14 (41.71-57.90)		0.98 (0.97-1.00)	0.82 (0.73-0.91)	0.99 (0.97-1.01)		2.89 (2.48-3.37)	2.91 (2.51-3.38)	2.51 (2.16-2.92)	1.23 (1.03-1.45)	0.32 (0.25-0.41)
notification model 19	51.11 (42.94-60.52)	0.99 (0.95-1.02)		0.78 (0.68-0.88)	0.99 (0.97-1.01)		2.89 (2.48-3.36)	2.91 (2.51-3.38)	2.51 (2.17-2.92)	1.23 (1.03-1.45)	0.33 (0.26-0.42)
notification model 20	50.35 (42.72-59.25)			0.79 (0.71-0.88)	0.99 (0.97-1.01)		2.89 (2.49-3.37)	2.91 (2.51-3.38)	2.51 (2.17-2.92)	1.22 (1.03-1.45)	0.33 (0.26-0.42)
notification model 21	42.22 (36.22-48.96)	1.00 (0.97-1.04)	0.98 (0.96-1.00)		1.00 (0.98-1.03)		2.89 (2.49-3.37)	2.91 (2.51-3.38)	2.51 (2.17-2.92)	1.23 (1.03-1.45)	0.37 (0.29-0.46)
notification model 22	42.24 (36.26-48.91)		0.98 (0.96-0.99)		1.00 (0.98-1.02)		2.89 (2.49-3.37)	2.91 (2.51-3.38)	2.51 (2.17-2.92)	1.23 (1.04-1.45)	0.37 (0.29-0.46)
notification	42.25	1.03 (0.99-			0.99 (0.97-		2.89 (2.48-	2.91 (2.51-	2.51 (2.16-	1.23 (1.03-	0.39 (0.31-

Model names	Intercept	Percentage of adults (≥15y)	Percentage of household heads that did not complete primary school	Distance to nearest TB clinic (km)	Percentage of HIV prevalence	Percentage of male adults	Year: 2015	Year: 2016	Year: 2017	Year: 2018	Random effects SD: cluster
model 23	(36.19-49.10)	1.06)			1.01)		3.36)	3.39)	2.92)	1.45)	0.48)
notification model 24	42.29 (36.29-49.16)				0.99 (0.97-1.02)		2.89 (2.48-3.36)	2.91 (2.51-3.38)	2.51 (2.16-2.92)	1.22 (1.04-1.45)	0.39 (0.31-0.48)
notification model 25	50.88 (42.99-60.00)	0.96 (0.93-1.00)	0.98 (0.96-0.99)	0.78 (0.69-0.88)			2.89 (2.48-3.37)	2.91 (2.51-3.38)	2.51 (2.16-2.92)	1.23 (1.03-1.45)	0.31 (0.24-0.39)
notification model 26	48.75 (41.40-57.19)		0.98 (0.97-1.00)	0.83 (0.74-0.91)			2.89 (2.49-3.37)	2.91 (2.51-3.39)	2.51 (2.17-2.92)	1.23 (1.03-1.45)	0.32 (0.25-0.40)
notification model 27	50.82 (42.83-60.12)	0.99 (0.95-1.02)		0.78 (0.69-0.88)			2.89 (2.49-3.36)	2.91 (2.52-3.38)	2.51 (2.17-2.92)	1.23 (1.03-1.45)	0.33 (0.26-0.42)
notification model 28	49.89 (42.27-58.58)			0.80 (0.72-0.89)			2.89 (2.48-3.37)	2.91 (2.51-3.38)	2.51 (2.16-2.92)	1.22 (1.03-1.45)	0.33 (0.26-0.42)
notification model 29	42.22 (36.35-48.93)	1.00 (0.97-1.04)	0.98 (0.96-0.99)				2.89 (2.49-3.36)	2.91 (2.51-3.38)	2.51 (2.17-2.92)	1.23 (1.04-1.45)	0.37 (0.29-0.45)
notification model 30	42.22 (36.24-48.94)		0.98 (0.96-0.99)				2.89 (2.49-3.37)	2.91 (2.51-3.38)	2.51 (2.16-2.92)	1.23 (1.03-1.45)	0.36 (0.29-0.45)

Model names	Intercept	Percentage of adults (≥15y)	Percentage of household heads that did not complete primary school	Distance to nearest TB clinic (km)	Percentage of HIV prevalence	Percentage of male adults	Year: 2015	Year: 2016	Year: 2017	Year: 2018	Random effects SD: cluster
notification model 31	42.27 (36.22-49.20)	1.03 (0.99-1.06)					2.89 (2.49-3.37)	2.91 (2.51-3.38)	2.51 (2.16-2.92)	1.23 (1.03-1.45)	0.38 (0.31-0.47)
notification model 32	42.33 (36.38-49.28)						2.89 (2.48-3.36)	2.91 (2.51-3.39)	2.51 (2.16-2.92)	1.22 (1.03-1.45)	0.39 (0.31-0.48)

5.S3 Table. Table of all the TB prevalence neighbourhood level models with spatial random effect. Coefficients (mean rate ratio) were exponentiated and intercepts were multiplied by 100,000 (S2 Equation).

Model names	Intercept	Percentage of adults (≥15y)	Percentage of household heads that did not complete primary school	Distance to nearest TB clinic (km)	Percentage of HIV prevalence	Percentage of male adults	Random effects SD: CAR	Probability Zi=0 (Refer to S2 Equation)
prevalence model 33	250.55 (128.71-481.99)	0.91 (0.73-1.13)	0.99 (0.92-1.08)	0.76 (0.42-1.39)	1.02 (0.90-1.14)	1.03 (0.72-1.47)	0.11 (0.01-0.50)	0.18 (0.01-0.45)
prevalence model 34	238.84 (121.98-469.25)		1.01 (0.94-1.08)	0.84 (0.48-1.48)	1.01 (0.90-1.13)	1.07 (0.78-1.49)	0.12 (0.01-0.53)	0.18 (0.01-0.45)
prevalence model 35	252.89 (130.36-480.74)	0.91 (0.75-1.11)		0.78 (0.42-1.42)	1.01 (0.91-1.13)	1.01 (0.71-1.45)	0.13 (0.01-0.53)	0.18 (0.01-0.45)
prevalence model 36	243.80 (126.45-467.87)			0.85 (0.48-1.50)	1.02 (0.91-1.13)	1.08 (0.78-1.49)	0.12 (0.01-0.50)	0.18 (0.01-0.45)
prevalence model 37	214.62 (124.68-375.95)	0.93 (0.76-1.14)	0.99 (0.92-1.07)		1.02 (0.91-1.15)	0.99 (0.70-1.39)	0.11 (0.01-0.47)	0.19 (0.01-0.47)
prevalence model 38	217.74 (128.17-380.36)		1.01 (0.94-1.08)		1.01 (0.90-1.14)	1.03 (0.76-1.39)	0.14 (0.01-0.55)	0.18 (0.01-0.45)
prevalence model 39	218.09 (125.32-385.46)	0.94 (0.78-1.12)			1.02 (0.91-1.14)	0.98 (0.70-1.37)	0.10 (0.00-0.45)	0.18 (0.01-0.46)
prevalence model 40	218.17 (128.15-				1.02 (0.92-1.13)	1.04 (0.79-1.38)	0.11 (0.01-0.45)	0.17 (0.01-0.44)

Model names	Intercept	Percentage of adults (≥15y)	Percentage of household heads that did not complete primary school	Distance to nearest TB clinic (km)	Percentage of HIV prevalence	Percentage of male adults	Random effects SD: CAR	Probability Zi=0 (Refer to S2 Equation)
	370.56)							
prevalence model 41	255.54 (132.62-486.46)	0.91 (0.74-1.12)	1.00 (0.93-1.07)	0.76 (0.41-1.37)		0.99 (0.72-1.37)	0.16 (0.01-0.60)	0.17 (0.01-0.45)
prevalence model 42	244.08 (128.82-462.85)		1.01 (0.95-1.08)	0.85 (0.48-1.48)		1.05 (0.78-1.42)	0.11 (0.00-0.47)	0.17 (0.01-0.44)
prevalence model 43	258.89 (136.73-481.87)	0.91 (0.75-1.10)		0.76 (0.42-1.36)		1.00 (0.73-1.38)	0.11 (0.01-0.45)	0.17 (0.01-0.44)
prevalence model 44	244.95 (128.15-469.47)			0.88 (0.49-1.60)		1.04 (0.76-1.41)	0.11 (0.00-0.47)	0.18 (0.01-0.46)
prevalence model 45	216.07 (126.61-368.62)	0.94 (0.77-1.14)	1.00 (0.93-1.07)			0.96 (0.70-1.30)	0.12 (0.00-0.52)	0.18 (0.01-0.45)
prevalence model 46	218.86 (129.77-367.12)		1.01 (0.94-1.08)			1.00 (0.78-1.30)	0.13 (0.01-0.53)	0.17 (0.01-0.44)
prevalence model 47	221.65 (130.22-380.42)	0.94 (0.78-1.11)				0.95 (0.70-1.28)	0.13 (0.00-0.54)	0.18 (0.01-0.45)
prevalence model 48	221.61 (131.66-378.89)					1.01 (0.78-1.31)	0.11 (0.00-0.47)	0.17 (0.01-0.44)

Model names	Intercept	Percentage of adults (≥15y)	Percentage of household heads that did not complete primary school	Distance to nearest TB clinic (km)	Percentage of HIV prevalence	Percentage of male adults	Random effects SD: CAR	Probability Zi=0 (Refer to S2 Equation)
prevalence model 49	245.92 (127.78-462.08)	0.91 (0.74-1.10)	0.99 (0.92-1.07)	0.78 (0.44-1.39)	1.01 (0.91-1.13)		0.12 (0.01-0.50)	0.17 (0.01-0.44)
prevalence model 50	236.68 (124.73-448.18)		1.01 (0.94-1.09)	0.89 (0.53-1.51)	1.00 (0.90-1.11)		0.11 (0.00-0.47)	0.18 (0.01-0.45)
prevalence model 51	252.77 (131.70-476.30)	0.91 (0.76-1.09)		0.79 (0.43-1.43)	1.01 (0.91-1.12)		0.14 (0.01-0.52)	0.17 (0.01-0.44)
prevalence model 52	234.16 (125.30-435.37)			0.91 (0.55-1.49)	1.01 (0.91-1.11)		0.11 (0.01-0.47)	0.17 (0.01-0.44)
prevalence model 53	215.44 (126.71-366.04)	0.93 (0.78-1.11)	0.99 (0.92-1.07)		1.03 (0.92-1.14)		0.15 (0.01-0.55)	0.17 (0.01-0.45)
prevalence model 54	221.54 (130.86-378.80)		1.01 (0.94-1.08)		1.01 (0.92-1.11)		0.11 (0.00-0.51)	0.17 (0.01-0.44)
prevalence model 55	215.88 (127.63-366.28)	0.94 (0.80-1.10)			1.02 (0.93-1.12)		0.10 (0.00-0.45)	0.17 (0.01-0.44)
prevalence model 56	221.22 (131.40-372.99)				1.01 (0.92-1.11)		0.12 (0.01-0.50)	0.16 (0.01-0.44)
prevalence model 57	257.33 (136.61-	0.91 (0.75-1.09)	1.00 (0.93-1.07)	0.75 (0.42-1.32)			0.10 (0.01-0.43)	0.17 (0.01-0.43)

Model names	Intercept	Percentage of adults (≥15y)	Percentage of household heads that did not complete primary school	Distance to nearest TB clinic (km)	Percentage of HIV prevalence	Percentage of male adults	Random effects SD: CAR	Probability Zi=0 (Refer to S2 Equation)
	472.56)							
prevalence model 58	239.85 (130.33-435.46)		1.01 (0.94-1.08)	0.88 (0.55-1.42)			0.12 (0.01-0.50)	0.17 (0.01-0.44)
prevalence model 59	255.43 (134.76-471.50)	0.91 (0.76-1.08)		0.75 (0.42-1.33)			0.08 (0.01-0.38)	0.16 (0.01-0.43)
prevalence model 60	241.39 (130.69-449.39)			0.90 (0.56-1.44)			0.13 (0.01-0.49)	0.17 (0.01-0.44)
prevalence model 61	217.40 (127.68-373.39)	0.95 (0.81-1.12)	1.00 (0.93-1.07)				0.08 (0.00-0.39)	0.17 (0.01-0.45)
prevalence model 62	226.08 (134.78-389.44)		1.01 (0.94-1.08)				0.14 (0.01-0.51)	0.17 (0.01-0.45)
prevalence model 63	221.43 (131.31-373.68)	0.96 (0.82-1.10)					0.09 (0.00-0.47)	0.17 (0.01-0.44)
prevalence model 64	224.66 (133.79-384.09)						0.09 (0.00-0.44)	0.16 (0.01-0.44)

5.S4 Table. Table of all the TB notified neighbourhood level models with spatial random effect. Coefficients (mean rate ratio) were exponentiated and intercepts were multiplied by 100,000 (S1 Equation).

Model names	Intercept	Percentage of adults (≥15y)	Percentage of household heads that did not complete primary school	Distance to nearest TB clinic (km)	Percentage of HIV prevalence	Percentage of male adults	Year: 2015	Year: 2016	Year: 2017	Year: 2018	Random effects SD: CAR
notification model 33	47.05 (39.69-55.57)	0.95 (0.92-1.00)	0.98 (0.96-1.00)	0.87 (0.74-1.02)	1.00 (0.98-1.02)	0.94 (0.87-1.00)	2.89 (2.49-3.36)	2.91 (2.52-3.38)	2.51 (2.16-2.91)	1.22 (1.03-1.45)	0.56 (0.40-0.74)
notification model 34	44.84 (37.88-52.89)		0.99 (0.97-1.00)	0.92 (0.79-1.08)	1.00 (0.98-1.02)	0.96 (0.90-1.03)	2.89 (2.49-3.37)	2.91 (2.52-3.39)	2.51 (2.17-2.92)	1.23 (1.04-1.45)	0.59 (0.43-0.77)
notification model 35	47.90 (40.38-56.69)	0.97 (0.93-1.01)		0.85 (0.72-1.00)	0.99 (0.97-1.01)	0.94 (0.87-1.00)	2.89 (2.48-3.36)	2.91 (2.51-3.38)	2.51 (2.16-2.92)	1.22 (1.03-1.45)	0.59 (0.43-0.77)
notification model 36	46.15 (39.01-54.30)			0.89 (0.77-1.03)	0.99 (0.97-1.01)	0.95 (0.89-1.02)	2.89 (2.49-3.37)	2.91 (2.51-3.38)	2.51 (2.17-2.92)	1.23 (1.04-1.45)	0.60 (0.44-0.78)
notification model 37	42.32 (37.40-47.77)	0.97 (0.93-1.01)	0.98 (0.96-0.99)		1.00 (0.98-1.03)	0.93 (0.86-1.00)	2.89 (2.48-3.35)	2.91 (2.51-3.38)	2.51 (2.16-2.91)	1.22 (1.03-1.45)	0.59 (0.44-0.76)
notification model 38	42.31 (37.31-47.70)		0.98 (0.97-1.00)		1.00 (0.98-1.02)	0.95 (0.89-1.01)	2.89 (2.49-3.37)	2.91 (2.51-3.38)	2.51 (2.17-2.92)	1.22 (1.04-1.45)	0.60 (0.45-0.77)
notification model 39	42.32 (37.35-47.72)	0.99 (0.95-1.03)			0.99 (0.97-1.01)	0.92 (0.86-0.99)	2.89 (2.49-3.37)	2.91 (2.52-3.38)	2.51 (2.16-2.91)	1.23 (1.03-1.45)	0.62 (0.47-0.80)

Model names	Intercept	Percentage of adults (≥15y)	Percentage of household heads that did not complete primary school	Distance to nearest TB clinic (km)	Percentage of HIV prevalence	Percentage of male adults	Year: 2015	Year: 2016	Year: 2017	Year: 2018	Random effects SD: CAR
notification model 40	42.33 (37.31-47.79)				0.99 (0.97-1.01)	0.93 (0.88-1.00)	2.89 (2.49-3.37)	2.91 (2.51-3.39)	2.51 (2.16-2.92)	1.23 (1.04-1.45)	0.62 (0.46-0.79)
notification model 41	47.20 (39.91-55.71)	0.95 (0.92-0.99)	0.98 (0.97-0.99)	0.86 (0.74-1.01)		0.94 (0.88-1.00)	2.89 (2.48-3.36)	2.91 (2.51-3.38)	2.51 (2.16-2.92)	1.22 (1.03-1.45)	0.55 (0.40-0.73)
notification model 42	44.87 (38.02-52.75)		0.99 (0.97-1.00)	0.92 (0.80-1.07)		0.96 (0.90-1.03)	2.89 (2.48-3.36)	2.91 (2.51-3.38)	2.51 (2.17-2.91)	1.22 (1.04-1.45)	0.58 (0.43-0.76)
notification model 43	47.52 (40.04-56.19)	0.97 (0.93-1.01)		0.86 (0.73-1.01)		0.94 (0.88-1.01)	2.89 (2.49-3.35)	2.91 (2.51-3.37)	2.51 (2.16-2.91)	1.22 (1.03-1.45)	0.58 (0.42-0.76)
notification model 44	45.74 (38.85-53.83)			0.90 (0.78-1.04)		0.96 (0.90-1.02)	2.89 (2.49-3.36)	2.91 (2.51-3.38)	2.51 (2.16-2.91)	1.22 (1.04-1.45)	0.59 (0.44-0.77)
notification model 45	42.34 (37.42-47.78)	0.97 (0.93-1.01)	0.98 (0.97-0.99)			0.92 (0.86-0.99)	2.89 (2.49-3.36)	2.91 (2.51-3.38)	2.51 (2.16-2.92)	1.22 (1.03-1.45)	0.58 (0.43-0.75)
notification model 46	42.35 (37.31-47.84)		0.98 (0.97-1.00)			0.95 (0.89-1.01)	2.89 (2.48-3.37)	2.91 (2.51-3.38)	2.51 (2.16-2.92)	1.22 (1.03-1.45)	0.59 (0.44-0.76)
notification model 47	42.35 (37.35-	0.99 (0.95-1.02)				0.93 (0.86-1.00)	2.89 (2.49-3.36)	2.91 (2.51-3.38)	2.51 (2.17-2.92)	1.22 (1.04-1.45)	0.61 (0.45-0.79)

Model names	Intercept	Percentage of adults (≥15y)	Percentage of household heads that did not complete primary school	Distance to nearest TB clinic (km)	Percentage of HIV prevalence	Percentage of male adults	Year: 2015	Year: 2016	Year: 2017	Year: 2018	Random effects SD: CAR
	47.79)										
notification model 48	42.39 (37.43-47.84)					0.94 (0.88-1.00)	2.89 (2.49-3.37)	2.91 (2.52-3.37)	2.51 (2.17-2.92)	1.22 (1.03-1.45)	0.61 (0.46-0.78)
notification model 49	47.94 (40.41-56.77)	0.97 (0.93-1.01)	0.98 (0.96-1.00)	0.85 (0.72-0.99)	1.00 (0.98-1.02)		2.89 (2.49-3.36)	2.91 (2.51-3.38)	2.51 (2.16-2.92)	1.22 (1.03-1.45)	0.57 (0.42-0.75)
notification model 50	45.85 (38.89-53.81)		0.98 (0.97-1.00)	0.90 (0.78-1.04)	1.00 (0.98-1.02)		2.89 (2.48-3.37)	2.91 (2.51-3.38)	2.51 (2.16-2.92)	1.22 (1.04-1.45)	0.59 (0.44-0.77)
notification model 51	48.80 (41.10-57.81)	0.98 (0.94-1.02)		0.83 (0.70-0.97)	0.99 (0.97-1.01)		2.89 (2.49-3.36)	2.91 (2.51-3.38)	2.51 (2.16-2.91)	1.23 (1.04-1.45)	0.61 (0.45-0.79)
notification model 52	47.45 (40.43-55.66)			0.86 (0.75-0.99)	0.99 (0.97-1.01)		2.89 (2.48-3.37)	2.91 (2.51-3.38)	2.51 (2.16-2.92)	1.23 (1.03-1.45)	0.61 (0.45-0.79)
notification model 53	42.27 (37.33-47.71)	0.98 (0.95-1.02)	0.98 (0.96-0.99)		1.01 (0.99-1.03)		2.89 (2.49-3.36)	2.91 (2.51-3.38)	2.51 (2.17-2.92)	1.23 (1.04-1.45)	0.61 (0.46-0.79)
notification model 54	42.29 (37.33-47.75)		0.98 (0.97-1.00)		1.01 (0.99-1.03)		2.89 (2.49-3.37)	2.91 (2.52-3.38)	2.51 (2.17-2.92)	1.22 (1.03-1.45)	0.61 (0.46-0.78)
notification	42.32	1.01 (0.97-			1.00 (0.98-		2.89 (2.48-	2.91 (2.51-	2.51 (2.16-	1.22 (1.03-	0.65 (0.49-

Model names	Intercept	Percentage of adults (≥15y)	Percentage of household heads that did not complete primary school	Distance to nearest TB clinic (km)	Percentage of HIV prevalence	Percentage of male adults	Year: 2015	Year: 2016	Year: 2017	Year: 2018	Random effects SD: CAR
model 55	(37.36-47.74)	1.04)			1.02)		3.36)	3.38)	2.92)	1.45)	0.83)
notification model 56	42.29 (37.29-47.71)				1.00 (0.98-1.02)		2.89 (2.49-3.36)	2.91 (2.52-3.38)	2.51 (2.17-2.92)	1.23 (1.04-1.45)	0.64 (0.49-0.82)
notification model 57	48.29 (40.74-56.87)	0.97 (0.93-1.01)	0.98 (0.97-1.00)	0.84 (0.72-0.98)			2.89 (2.48-3.36)	2.91 (2.51-3.37)	2.51 (2.16-2.92)	1.22 (1.03-1.45)	0.57 (0.42-0.75)
notification model 58	45.99 (39.20-53.76)		0.98 (0.97-1.00)	0.89 (0.78-1.03)			2.89 (2.49-3.37)	2.91 (2.52-3.38)	2.51 (2.16-2.92)	1.23 (1.04-1.45)	0.58 (0.43-0.76)
notification model 59	48.50 (40.78-57.23)	0.98 (0.94-1.02)		0.83 (0.71-0.98)			2.89 (2.49-3.36)	2.91 (2.51-3.38)	2.51 (2.17-2.92)	1.22 (1.04-1.45)	0.60 (0.44-0.78)
notification model 60	47.12 (40.11-55.09)			0.87 (0.76-0.99)			2.89 (2.48-3.37)	2.91 (2.51-3.38)	2.51 (2.16-2.92)	1.23 (1.03-1.45)	0.60 (0.44-0.77)
notification model 61	42.31 (37.32-47.75)	0.99 (0.95-1.03)	0.98 (0.97-1.00)				2.89 (2.49-3.36)	2.91 (2.51-3.39)	2.51 (2.16-2.92)	1.22 (1.03-1.45)	0.61 (0.47-0.79)
notification model 62	42.28 (37.32-47.75)		0.98 (0.97-1.00)				2.89 (2.49-3.37)	2.91 (2.52-3.38)	2.51 (2.17-2.92)	1.23 (1.04-1.45)	0.61 (0.46-0.78)

Model names	Intercept	Percentage of adults (≥15y)	Percentage of household heads that did not complete primary school	Distance to nearest TB clinic (km)	Percentage of HIV prevalence	Percentage of male adults	Year: 2015	Year: 2016	Year: 2017	Year: 2018	Random effects SD: CAR
notification model 63	42.30 (37.29-47.78)	1.01 (0.97-1.04)					2.89 (2.49-3.36)	2.91 (2.51-3.38)	2.51 (2.16-2.92)	1.23 (1.03-1.45)	0.64 (0.49-0.82)
notification model 64	42.33 (37.38-47.84)						2.89 (2.49-3.36)	2.91 (2.51-3.38)	2.51 (2.16-2.92)	1.23 (1.04-1.45)	0.63 (0.48-0.81)

5.S5 Table. Table of top ten prevalence models using the ELPD LOO statistic, comparing the models in S1 and S3 Tables.

Model names	ELPD differences	Standard error of differences
prevalence model 32	0.0000000	0.0000000
prevalence model 31	-0.6439814	0.5953633
prevalence model 64	-0.8100483	4.2533603
prevalence model 30	-0.8401287	0.2957389
prevalence model 24	-0.8426507	0.4155533
prevalence model 16	-0.8837806	0.2603595
prevalence model 28	-1.0206936	0.2818817
prevalence model 15	-1.2537159	0.7767991
prevalence model 23	-1.4261711	0.8171422
prevalence model 63	-1.4481648	4.1704611

5.S6 Table. Table of top ten notification models using the ELPD LOO statistic, comparing the models in S2 and S4 Tables.

Model names	ELPD difference	Standard error of ELPD difference
notification model 62	0.000000000	0.0000000
notification model 25	-0.006516547	3.2225076
notification model 45	-0.164705190	1.6514375
notification model 61	-0.214042393	0.3547003
notification model 9	-0.228767982	3.3141359
notification model 26	-0.267034261	3.1569474
notification model 17	-0.322702170	3.1960427
notification model 18	-0.340474087	3.0894852
notification model 2	-0.433583819	3.0863202
notification model 46	-0.477146663	1.1087664

5.S7 Table. Parameter estimates for final regression models for predicting neighbourhood level TB prevalence and notifications. Analysis based on post stratified TB prevalence and with confirmed TB notifications kept the same as in the primary analysis.

	Notification model		**Prevalence model**	
Fixed effects **Parameters**	**Mean rate ratio**	**95% CrI**	**Mean rate ratio**	**95% CrI**
Percentage of adults (≥15y)	0.96	(0.93, 1.00)	1.00	(0.97, 1.02)
Distance to nearest TB clinic (km)	0.78	(0.69, 0.88)		
Percentage of household heads that did not complete primary school	0.98	(0.96, 0.99)		
Year: 2019	Reference			
Year: 2015	2.89	(2.48, 3.37)		
Year: 2016	2.91	(2.51, 3.38)		
Year: 2017	2.51	(2.16, 2.92)		
Year: 2018	1.23	(1.03, 1.45)		
Intercept	$50.88*10^{-5}$	($42.99*10^{-5}$, $60.00*10^{-5}$)	$234.85*10^{-5}$	($219.71*10^{-5}$, $250.68*10^{-5}$)
Zero inflation intercept			0.01	(0.00, 0.05)
Random effects SD: cluster	0.31	(0.24, 0.39)	0.03	(0.00,0.08)

[a]Percentage of adults was centred by subtracting by its mean (60.90%), Distance to nearest TB clinic (km) was centred by subtracting by 1km, Percentage of household head that did not complete primary school was centred by subtracting by its mean (16.90%).

Abbreviations: km kilometre, sd standard deviation, CrI Credible interval

5.S8 Table. Parameter estimates for final regression models for predicting neighbourhood level TB prevalence and notifications. Analysis based on both microbiologically-confirmed TB and clinically-diagnosed cases and with TB prevalence kept the same as in the primary analysis.

	Notification model		Prevalence model	
Fixed effects **Parameters**	**Mean rate ratio**	**95% CrI**	**Mean rate ratio**	**95% CrI**
Percentage of adults (≥15y)	0.98	(0.95, 1.01)	0.94	(0.80, 1.10)
Distance to nearest TB clinic (km)	0.81	(0.73, 0.90)		
Percentage of household heads that did not complete primary school	0.98	(0.97, 0.99)		
Year: 2019	Reference			
Year: 2015	1.62	(1.47, 1.78)		
Year: 2016	1.82	(1.66, 1.99)		
Year: 2017	1.70	(1.56, 1.86)		
Year: 2018	0.84	(0.76, 0.94)		
Intercept	$163.54*10^{-5}$	($144.68*10^{-5}$, $184.93*10^{-5}$)	$232.08*10^{-5}$	($132.05*10^{-5}$, $404.96*10^{-5}$)
Zero inflation intercept			0.18	(0.01, 0.46)
Random effects SD: cluster	0.30	(0.24, 0.37)	0.33	(0.01, 0.90)

[a]Percentage of adults was centred by subtracting by its mean (60.90%), Distance to nearest TB clinic (km) was centred by subtracting by 1km, Percentage of household head that did not complete primary school was centred by subtracting by its mean (16.90%).

Abbreviations: km kilometre, sd standard deviation, CrI Credible interval

5.S9 Table. Comparing model results of prevalence to microbiologically confirmed notification rations primary analysis, compared to analysis based on post stratified TB prevalence and based on all TB case notifications. Only neighbourhoods that were included in the 4th quartile were included in the table.

Neighbourhood	TB prevalence to confirmed TB notification ratio	Post-stratified TB prevalence to confirmed TB notification ratio	TB prevalence to all TB notification ratio	TB prevalence to confirmed TB notification ratio quartile	TB prevalence to all TB notification ratio quartile	Post-stratified TB prevalence to confirmed TB notification quartile
1	5.882872	6.961638	1.741540	4	3	4
2	5.835493	6.902129	2.107328	3	4	4
14	7.265753	8.613159	1.874598	4	4	4
15	5.891822	6.865971	1.624834	4	3	4
20	5.640371	6.577177	1.707080	3	3	4
22	6.245970	6.274060	1.841799	4	4	3
23	6.763737	7.160686	1.854707	4	4	4
31	6.696159	7.908075	1.823901	4	4	4
34	6.079875	7.172957	1.682113	4	3	4
37	6.293189	7.377623	1.867094	4	4	4
40	5.391903	5.657743	1.791965	3	4	3
41	5.982987	7.113974	2.239345	4	4	4
44	8.829594	9.244351	2.662622	4	4	4
47	7.544831	8.024778	2.375998	4	4	4
48	6.277072	7.441708	2.262311	4	4	4
52	5.640333	6.596341	1.643353	3	3	4
55	6.184062	6.329324	1.854457	4	4	3
56	5.979641	6.155832	1.857542	4	4	3
57	10.369700	12.219568	2.923005	4	4	4
61	6.253442	7.287852	1.870250	4	4	4
63	6.355325	6.691939	2.239623	4	4	4
64	6.889153	8.139249	2.091773	4	4	4
69	5.218804	6.024331	2.056106	3	4	3

References

1. World Health Organisation. Global tuberculosis report 2020. Geneva; 2020. Available: https://www.who.int/publications/i/item/9789240013131

2. World Health Organization (WHO). Global tuberculosis report 2021. Geneva; 2021. Available: https://www.who.int/publications/i/item/9789240037021

3. World Health Organisation. Module 2: Screening WHO operational handbook on tuberculosis Systematic screening for tuberculosis disease. 2021. Available: https://reliefweb.int/report/world/who-consolidated-guidelines-tuberculosis-module-2-screening-systematic-screening

4. Soko RN, Burke RM, Feasey HRA, Sibande W, Nliwasa M, Henrion MYR, et al. Effects of Coronavirus Disease Pandemic on Tuberculosis Notifications, Malawi. Emerg Infect Dis. 2021;27: 1831–1839. doi:10.3201/eid2707.210557

5. World Health Organisation. Global strategy and targets for tuberculosis prevention, care and control after 2015 Report by the Secretariat. Geneva; 2015. Available: http://apps.who.int/gb/ebwha/pdf_files/EB134/B134_12-en.pdf?ua=1

6. Burke RM, Nliwasa M, Feasey HRA, Chaisson LH, Golub JE, Naufal F, et al. Community-based active case-finding interventions for tuberculosis: a systematic review. Lancet Public Heal. 2021;6: e283–e299. doi:10.1016/S2468-2667(21)00033-5

7. Corbett EL, Bandason T, Duong T, Dauya E, Makamure B, Churchyard GJ, et al. Comparison of two active case-finding strategies for community-based diagnosis of symptomatic smear-positive tuberculosis and control of infectious tuberculosis in Harare, Zimbabwe (DETECTB): a cluster-randomised trial. Lancet. 2010;376: 1244–1253. doi:10.1016/S0140-6736(10)61425-0

8. Marks GB, Nguyen N V., Nguyen PTB, Nguyen T-A, Nguyen HB, Tran KH, et al. Community-wide Screening for Tuberculosis in a High-Prevalence Setting. N Engl J Med. 2019;381: 1347–1357. doi:10.1056/nejmoa1902129

9. Shaweno D, Trauer JM, Denholm JT, McBryde ES. A novel Bayesian geospatial method for estimating tuberculosis incidence reveals many missed TB cases in Ethiopia. BMC Infect Dis. 2017;17: 662. doi:10.1186/s12879-017-2759-0

10. Claassens MM, van Schalkwyk C, Dunbar R, Ayles H, Beyers N. Patient diagnostic rate as indicator of tuberculosis case detection, South Africa. Emerg Infect Dis. 2016;22: 535–537. doi:10.3201/eid2203.151618

11. MacPherson P, Khundi M, Nliwasa M, Choko AT, Phiri VK, Webb EL, et al. Disparities in access to diagnosis and care in Blantyre, Malawi, identified through enhanced tuberculosis surveillance and spatial analysis. BMC Med. 2019;17: 21. doi:10.1186/s12916-019-1260-6

12. Cudahy PGT, Andrews JR, Bilinski A, Dowdy DW, Mathema B, Menzies NA, et al. Spatially

targeted screening to reduce tuberculosis transmission in high-incidence settings. Lancet Infect Dis. 2018;0. doi:10.1016/S1473-3099(18)30443-2

13. Dowdy DW, Golub JE, Chaisson RE, Saraceni V. Heterogeneity in tuberculosis transmission and the role of geographic hotspots in propagating epidemics. Proc Natl Acad Sci. 2012;109: 9557–9562. doi:10.1073/pnas.1203517109

14. Horton KC, Sumner T, Houben RMGJ, Corbett EL, White RG. Practice of Epidemiology A Bayesian Approach to Understanding Sex Differences in Tuberculosis Disease Burden. 2018;187. doi:10.1093/aje/kwy131

15. Dowdy DW, Dye C, Cohen T. Data Needs for Evidence-Based Decisions: A TB Modeler's "Wish List" [cited 25 Feb 2019]. doi:10.5588/ijtld.12.0573

16. van Leth F, van der Werf MJ, Borgdorff MW. Prevalence of tuberculous infection and incidence of tuberculosis: a re-assessment of the Styblo rule. Bull World Health Organ. 2008;86: 20–6. doi:10.2471/BLT.06.037804

17. Borgdorff MW. New measurable indicator for tuberculosis case detection. Emerg Infect Dis. 2004;10: 1523–8. doi:10.3201/eid1009.040349

18. Gopi PG, Subramani R, Santha T, Kumaran PP, Kumaraswami V, Narayanan PR. Relationship of ARTI to incidence and prevalence of tuberculosis in a district of south India. INT J TUBERC LUNG DIS. 2006. Available: https://www.ingentaconnect.com/content/iuatld/ijtld/2006/00000010/00000001/art00021?crawler=true&mimetype=application/pdf

19. United Nations Human Settlements Programme. SLUM ALMANAC 2015 2016 Tracking Improvement in the Lives of Slum Dweller. Nairobi; 2015. Available: https://unhabitat.org/slum-almanac-2015-2016/

20. National Statistical Office of Malawi (NSO). Malawi population census preliminary report. Zomba; 2018. Available: www.nsomalawi.mw

21. Malawi Ministry of Health. Malawi Population-Based HIV Impact Assessment (MPHIA) 2015-2016: Final Report. Lilongwe, Ministry of Health. October 2018. [cited 29 Sep 2020]. Available: http://phia.icap.columbia.edu

22. Malawi Ministry of Health. Malawi National Tuberculosis Control Programme 7th edition. Lilongwe; 2012. Available: https://pdf.usaid.gov/pdf_docs/PA00KGDD.pdf

23. Macpherson P, Choko AT, Webb EL, Thindwa D, Squire SB, Sambakunsi R, et al. Development and validation of a global positioning system-based "map book" system for categorizing cluster residency status of community members living in high-density urban slums in Blantyre, Malawi. Am J Epidemiol. 2013;177: 1143–1147. doi:10.1093/aje/kws376

24. Kok MC, Namakhoma I, Nyirenda L, Chikaphupha K, Broerse JEW, Dieleman M, et al. Health surveillance assistants as intermediates between the community and health

sector in Malawi: exploring how relationships influence performance. BMC Health Serv Res. 2016;16: 1–12. doi:10.1186/s12913-016-1402-x

25. Khundi M, MacPherson P, Feasey HRA, Soko RN, Nliwasa M, Corbett EL, et al. Clinical, health systems and neighbourhood determinants of tuberculosis case fatality in urban Blantyre, Malawi: a multilevel epidemiological analysis of enhanced surveillance data. Epidemiol Infect. 2021;149. doi:10.1017/S0950268821001862

26. Payongayong E, Benson T, Ahmed A, Kanyanda S, Chakanza C, Mwanza P, et al. Simple household poverty assessment models for Malawi: proxy means test from the 1997–98 Malawi Integrated Household Survey. Lilongwe, Malawi: National Statistical Office, Government of Malawi, 2006.

27. Malawi National Statistical Office. Intergrated Household Survey (IHS 3) Report 2010-2011. Zomba; 2012.

28. Bürkner P-C. brms: An R Package for Bayesian Multilevel Models Using Stan. J Stat Softw. 2017;80: 1–28. doi:10.18637/jss.v080.i01

29. Vehtari A, Gelman A, Gabry J. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. 2016 [cited 22 Jun 2021]. Available: https://github.com/stan-dev/loo.

30. Tennekes M. tmap: Thematic Maps in R. J Stat Softw. 2018;84: 1–39. doi:10.18637/JSS.V084.I06

31. Stevens FR, Gaughan AE, Linard C, Tatem AJ. Disaggregating census data for population mapping using Random forests with remotely-sensed and ancillary data. PLoS One. 2015;10. doi:10.1371/JOURNAL.PONE.0107042

32. Brooks MB, Jenkins HE, Puma D, Tzelios C, Millones AK, Jimenez J, et al. A role for community-level socioeconomic indicators in targeting tuberculosis screening interventions. Sci Rep. 2022;12: 781. doi:10.1038/S41598-022-04834-7

33. Alba S, Rood E, Mecatti F, Ross JM, Dodd PJ, Chang S, et al. TB Hackathon: Development and Comparison of Five Models to Predict Subnational Tuberculosis Prevalence in Pakistan. Tropical Medicine and Infectious Disease. 2022. doi:10.3390/tropicalmed7010013

34. Philippe G, Charalambos S, Carel P, Katherine F. Methods used by WHO to estimate the Global burden of TB disease. 2016. Available: http://arxiv.org/abs/1603.00278

35. Shaweno D, Karmakar M, Alene KA, Ragonnet R, Clements AC, Trauer JM, et al. Methods used in the spatial analysis of tuberculosis epidemiology: a systematic review. BMC Med. 2018;16: 193. doi:10.1186/s12916-018-1178-4

36. Khundi M, Carpenter JR, Nliwasa M, Cohen T, Corbett EL, MacPherson P. Effectiveness of spatially targeted interventions for control of HIV, tuberculosis, leprosy and malaria: a systematic review. BMJ Open. 2021;11: e044715. doi:10.1136/BMJOPEN-2020-044715

37. Thindwa D, Macpherson P, Choko AT, Khundi M, Sambakunsi R, Ngwira LG, et al. Completion of isoniazid preventive therapy among human immunodeficiency virus positive adults in urban Malawi. Int J Tuberc Lung Dis. 2018;22: 273–279. doi:10.5588/IJTLD.17.0370

38. Ku C-C, MacPherson P, Khundi M, Nzawa R, Feasey HR, Nliwasa M, et al. Estimated durations of asymptomatic, symptomatic, and care-seeking phases of tuberculosis disease. medRxiv. 2021; 2021.03.17.21253823. doi:10.1101/2021.03.17.21253823

39. Wold Health Organisation. Systematic screening for active tuberculosis. Geneva; 2013. doi:ISBN 978 92 4 154860 1

40. Shaweno D, Karmakar M, Alene KA, Ragonnet R, Clements AC, Trauer JM, et al. Methods used in the spatial analysis of tuberculosis epidemiology: a systematic review. BMC Med. 2018;16: 193. doi:10.1186/s12916-018-1178-4

41. MacPherson P, Khundi M, Nliwasa M, Choko AT, Phiri VK, Webb EL, et al. Disparities in access to diagnosis and care in Blantyre, Malawi, identified through enhanced tuberculosis surveillance and spatial analysis. BMC Med. 2019;17: 21. doi:10.1186/s12916-019-1260-6

42. World Health Organization. Tuberculosis PREVALENCE SURVEYS: a handbook. 2011. Available: www.who.int/tb/advisory_bodies/impact_measurement_taskforce/resources_documents/thelimebook/en/index.html

43. Floyd K, Glaziou P, Houben RMGJ, Sumner T, White RG, Raviglione M. Global tuberculosis targets and milestones set for 2016-2035: Definition and rationale. Int J Tuberc Lung Dis. 2018;22: 723–730. doi:10.5588/ijtld.17.0835

6 Summary discussion, recommendations and conclusions

6.1 Summary of rationale, objectives and key findings

While African countries are making progress in combating the TB pandemic, there is still a significant gap between what has been accomplished and what is required to attain the EndTB targets (1). To make additional progress, current approaches to TB case finding must be improved, and new effective approaches developed and implemented (2). Facility-based TB case-finding is still an important part of case-finding and will be for the foreseeable future (3). But due to access to care barriers that exist at the individual-, family- and community-levels, access to TB screening, diagnosis, and treatment remains challenging for many, exacerbating treatment delays and resulting in continued high-levels of community transmission (4). Even if implemented with optimal operating capacity, strategies wholly reliant upon facility-based case-finding will result in considerable numbers of missed and delayed TB diagnoses (5).

Systematically screening people for TB (for example through community-based active case finding (ACF) interventions delivered to priority populations) identifies people with symptomatic and asymptomatic TB earlier (6), and so identifies cases that would have been missed by passive case finding (PCF) (5,7). WHO recognises the importance of community ACF to tackling TB epidemics, and has developed recommendations to guide the implementation of community-based ACF (5). However, the current WHO recommendations on selection of populations for community ACF are still not specific enough for most cities, especially in Africa where rapid urbanisation means that TB risk profiles and epidemiology are likely to be highly heterogenous. WHO guidelines have two broad recommendations for prioritising communities

for ACF: (1) populations with a TB prevalence of 0.5% prevalence or more, and (2) populations with structural risk factors for TB, these include urban poor communities, homeless communities, communities in remote or isolated areas, migrants, refugees, internally displaced persons and other vulnerable or disadvantaged groups with limited access to health care (5). This thesis proposes that leveraging surveillance data and spatial modelling techniques to identify hotspots of underdiagnosis of TB has substantial potential to provide further guidance for prioritisation of communities for implementation of ACF interventions (5).

However, prior to commencing this thesis, there were critical knowledge gaps about how to best investigate the spatial epidemiology of TB, and how these data may be used to improve community ACF interventions and reduce the mortality of TB patients. This led to our research aims, hypothesis and objectives of the thesis that were first stated in the introduction (Section 1.6). To recap:

We **hypothesised** that we could find epidemiologically-important spatial variability in TB epidemiology in urban Malawi, which may be used to guide better-targeted and perhaps more successful TB care and prevention activities, based on recent research and TB notification and prevalence data.

The specific **objectives** were to**:**

1. Systematically review the available literature on the effectiveness of spatially targeted interventions for HIV, TB, leprosy, and malaria. Note that, since spatially-targeted interventions of TB are still relatively new, I included the other infectious diseases to broaden the scope of our understanding of spatially targeted interventions.

2. Identify spatial and non-spatial risk factors for case fatality rates of notified TB cases in urban Blantyre Malawi.
3. Estimate neighbourhood level TB case prevalence to notification ratios (P:N ratios) using data from systematic surveillance of notified TB cases and a citywide TB prevalence survey form Urban Blantyre, Malawi. To help identify neighbourhoods that have a higher burden of undiagnosed TB that may be being missed by the routine TB notification surveillance system.
4. Discuss the findings of Objectives 1 to 3 to inform design of effective community ACF interventions.

6.2 Overview of the main findings

The systematic review of the effectiveness of spatially targeted community interventions showed that these interventions are feasible and that they have potential as alternative strategies in targeting communities for ACF (8). Through a systematic literature search, we identified ten studies from six countries between 1 January 1993 and 22 March 2021. The spatially targeted interventions that were evaluated were three against TB, three against leprosy, three against malaria and one against HIV. Synthesising these data, we concluded that, although data were limited and understanding of effectiveness also limited by high risk of bias (particularly in classification of hotspots and ascertainment of outcomes), the spatially-targeted interventions demonstrated potential to identify communities with a higher yield of identified cases, communities with a high prevalence of cases, and also for reducing the TB case notifications based on before-and-after analyses.

There were key weaknesses when it comes to evaluating the utility of hotspot-targeted interventions. Only one study used random allocation of communities in evaluation of the outcome of interest, with the remainder being observational (8). The hotspot communities in most of the studies were selected based on notified disease cases (8,9). Since disease notifications can be affected by barriers to access to care (1,10), hotspots based on notification only can end up selecting communities that have greater access to care. Therefore, better methods for defining hotspots are needed to address bias (11).

In multi-level regression modelling analysis of TB case fatality rates across the city of Blantyre, (in which we evaluated 4397 newly diagnosed TB cases, 10.9% (479) of whom died), we found that older age, being HIV positive, and distance to TB treatment clinic were associated with an increased odds of death on TB treatment (12). Distance to health facility is a proxy indicator of ease of access to care: populations that have to travel longer distances to get health care are at an increased risk of adverse health outcomes which include delayed diagnosis, treatment default and death (13–16). In this population distance increased the odds of death only for patients that were registered at the referral clinic but not a primary health care clinic (12).

We have already commented that the majority of studies in the systematic reviews defined hotspots based on routine notifications data, which might be affected by barriers to access to care (8,17,18). The systematic reviews highlighted the need for using other methods in areas where access to health care may be uneven across neighbourhoods (8,17,18). Most African cities such as Blantyre have populations that are unevenly served by routine health care services because of geographical distance to health care facilities and because individuals in some communities may lack resource to access care even when the services are free (19).

Using data from prevalence survey and a routine TB surveillance system from Blantyre Malawi, we developed a statistical modelling approach to enable identification of TB hotspot neighbourhoods based on ranking of neighbourhood prevalence to notification ratios (11). Adult bacteriologically-confirmed TB case notification rates were decreasing between 2015 to 2019; 131, 134, 114, 56 and 46, per 100, 000 population respectively. In 2019, the prevalence of TB in Blantyre was 215 per 100, 000 population. Model derived mean neighbourhood prevalence to notification ratios were 4.49 (95% credible interval (CrI): 0.98–11.91) and (range: 1.70–10.40, standard deviation: 1.79). The neighbourhoods were ranked according to the size of the prevalence to notification ratio from lowest to highest number. Neighbourhoods with the highest prevalence to notification ratios were likely to have a higher burden of undiagnosed TB.

Our model should enable other places with similar characteristics to urban Blantyre to highlight potential hotspot neighbourhoods without the need for carrying out a full prevalence survey. This would provide a cost saving strategy for prioritising neighbourhoods because most TB programmes cannot afford a prevalence survey (20). Once individuals are diagnosed and are put on TB treatment, generally they have a good prognosis if they adhere to the treatment regimen and can again be productive members of society (1).

Taken together, this body of research has provided novel insights into the epidemiology of TB in a densely-populated African city, typical of many urban centres on the continent. We have, for the first time, identified a modelling approach that can assist policymakers with targeting critically-needed community-case finding interventions towards neighbourhoods with the highest prevalence of undiagnosed TB. Moreover, we have shown that spatially-targeted interventions have potential to be effective in improving community TB epidemiology (11), but

that access to clinic-based TB diagnosis and care, and quality of care, remain critically important components of a comprehensive TB control strategy (11,21).

6.3 Findings in context

The systematic literature review that was done as part of this thesis in Chapter 3 (8), should be seen in the context of other systematic literature reviews that have been done in this area, covering separate but relevant topics on the subject of community spatially targeted ACF (17,18). In particular, Shaweno et.al 2018 (17), present a systematic review of methods used for identification of TB hotspots, providing collective evidence that TB is a disease that is spatially heterogenous in the population, as well as the common approaches for identifying these hotspots. Cudahy et al. 2018 (18), undertook a systematic review of spatially targeted ACF interventions of TB, identifying TB studies that used spatially targeted community ACF approaches.

The systematic review in Chapter 3 (8), was aimed at investigating the effectiveness of spatially targeted interventions of TB. The focus of the systematic review was to review the methods that were used to identify hotspots in spatially targeted interventions and to evaluate the outcome of the studies. Unlike the two other systematic reviews, this systematic review also included spatially targeted interventions of HIV, leprosy, and Malaria. The other diseases were included because spatially targeted interventions are a new study design so we wanted to learn what has been accomplished in these other infectious diseases.

The systematic reviews (8,17,18) recognise that for TB control efforts to advance further, they must begin to take advantage of the fact that TB cases tend to cluster more in particular sectors

of the population than in others. Our systematic review gathered all relevant information for assessing the level of evidence available to aid in determining the feasibility and effectiveness of spatially targeted ACF interventions (8).

Since we completed the systematic review of the effectiveness of spatially-targeted interventions for infectious diseases, three additional studies have been published.

Robsky et al (2020) (22) developed a method for ranking neighbourhoods using TB notification rates. The study was done in Kampala Uganda in an area with 33 contiguous zones (neighbourhoods), each zone had at least a population of 500 and a median size of 0.005 km^2. Zonal level notification rates were calculated for the period May 2018 to January 2019. Neighbourhoods that were selected as TB hotspots were those that had had the highest notification rates and whose total population comprised at least 20% of the 33 neighbourhood populations. Five zones (neighbourhoods), which made up 22 % of the total population and were responsible for an estimated 62% (95% CI:47–75%) of the TB case notifications, were defined as hotstpots of TB in this analysis. A door-to-door ACF intervention was subsequently done in all the neighbourhoods to verify if the five identified hotspot neighbourhoods would be the source of most undiagnosed TB cases. In total, the five "hotspot" neighbourhoods accounted for 42% (95% CI:34–51%) of the total 128 cases that were identified during the community ACF intervention; this was double compared to what could have been expected by chance ($P < 0.001$). While this shows the potential for the use of TB notifications in defining hotspots, it does not readily generalise to some other areas. The main reason for this was because the study setting was small (about 2.2 km^2) with a population of 49,527 and hence access to care might have been similar across neighbourhoods because of the small size of the

area and population. Access to care barriers are likely to affect TB notification from populations distributed across a greater area, like in Blantyre.

In Alba et. al 2022 (23), the Pakistan National TB Control Programme organised a data modelling competition ("hackathon") for groups of independent data modelling teams to develop models for predicting the subnational district-level TB burden estimates for Pakistan to inform disease control efforts. Investigator were provided with individual-level data from the 2010-2011 national TB prevalence survey, and additional data include census denominators and TB notification data. Each of the five teams employed different modelling strategies and found a different set of predictor variables and predicted district level notification to prevalence ratios. Nevertheless, several districts were consistently identified as areas with low notification to prevalence ratios by all the competing teams. The statistic used here notification to prevalence ratio is an inverse of our calculated statistic in Ch 5. Low values of notification to prevalence ratio were indicative of higher likelihood of having undiagnosed TB cases in the district. This work also raises the important aspect that predictors of notification to prevalence ratios might not only be setting dependent but also are dependent on the modelling strategy used. This highlights the need for principled model building approaches and better methods for comparing the validity of models. Also, the models need to be evaluated on a new data source because there is a limit to what can be learnt from the data source that was used to define the model. A major limitation of the analysis was sparsity of data, with small number of TB prevalent cases in most districts. District level hotspots are still at a very top level, while in this thesis we argue that in cities or districts that are considered to have a high burden of TB, there

is still heterogeneity in the burden of TB and further stratification is necessary to exclude areas with low risk of TB out of ACF interventions (11).

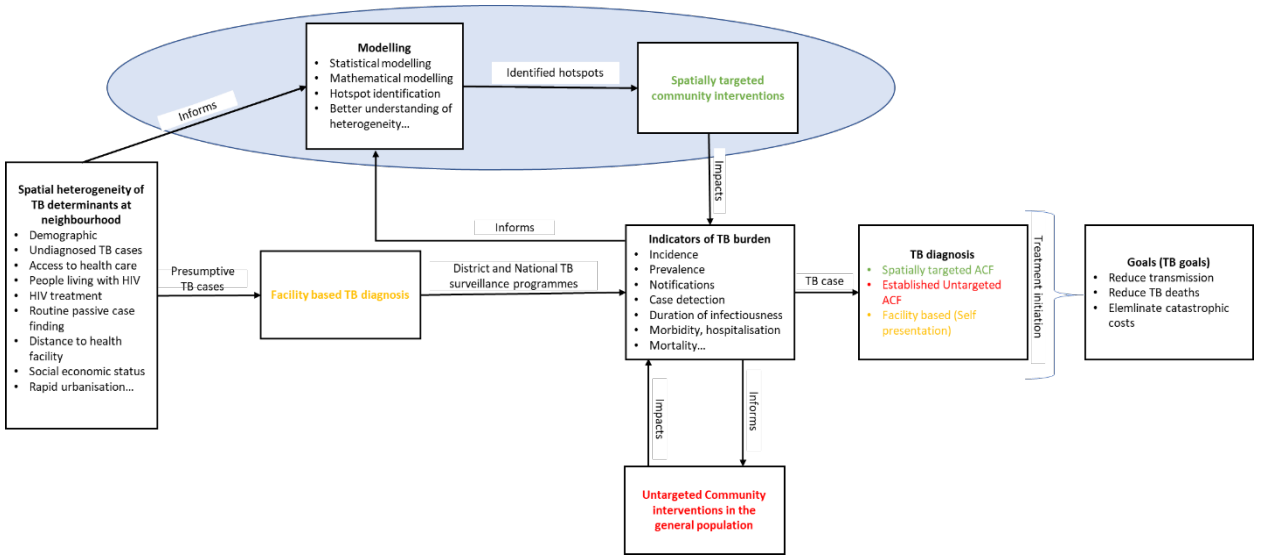
Brooks et. al 2022 (24) undertook a study in 74 contiguous neighbourhoods in Lima, Peru with a population of approximately 212,000. A community screening ACF intervention was implemented in the neighbourhoods using mobile screening units. In exploratory analysis, economic indicators were identified as better predictors of screening yield (number of individuals tested to identify one positive case). However, previous year's TB case notification rates were not associated with the screening yield. The finding by Brook et.al that notification rates of previous years was not associated with screening yield adds further evidence to why hotspots defined based on notifications only, can be biased. The analysis identified percentage of vehicle ownership, percentage of blender ownership and percentage of TB patients with prior TB episode as important univariate predictors of screening yield.

6.4 A framework for including targeted public health interventions within TB control programmes

We now consider how the findings of this thesis can aid public health policymakers and researchers in improving case detection of TB and other infectious diseases. The conceptual diagram below Figure 6.1, depicts where spatially targeted interventions fit in, in the cascade of community interventions of TB. Spatial heterogeneity in the burden of TB is associated with neighbourhood TB risk determinants. Data on these determinants when combined with data on indicators of TB burden can be used through modelling to identify hotspots of undiagnosed TB. These identified hotspots can be targeted with spatially targeted ACF interventions. Availability

of quality data for TB surveillance underpins the extent to which data on determinants of TB spatial heterogeneity and TB burden are available for use for identification of hotspots. Other interventions can be untargeted, these are interventions that do not take into consideration the variation in neighbourhood risk of undiagnosed TB at subdistrict level. Newly diagnosed TB cases are referred to their closest health facility that offers TB treatment. The figure was created using the available literature on TB and the findings of our systematic review in Chapter 3 (4,8,17,18,25).

Figure 6.1: Conceptual diagram to describe the mechanisms by which spatial epidemiology of TB vary by neighbourhood



6.4.1 Contributions to public health (Recommendations for future directions)

The findings of the systematic review lend support to the potential and feasibility of spatially targeted community case finding interventions of TB and infectious diseases (8). The current WHO recommendation for community ACF in generalised populations is to do community ACF in populations with a prevalence of 0.5% (5). This guidance is difficult to follow in a resource limited setting that cannot afford to carry out a prevalence survey to ascertain it. In Africa, when TB prevalence surveys are done, they are mostly done at the country level resolution and have inefficient sample size to provide meaningful direction for subnational level prevalence estimates (26,27). With such a lack of subnational prevalence level data, it is challenging to prioritise district, or sub-district level areas for ACF interventions.

The other WHO recommendation for generalised populations is directed at subpopulations with structural risk factors for TB (5). These include urban poor communities and people living in slums. A category such as “people living in slums” is a wide category, without a clear cut off point, for instance Malawi had about 1.8 million urban slum dwellers in 2014 (28). Even within this category it is possible that the risk of TB would be heterogenous to warrant further stratification of the group according to risk of undiagnosed TB (19). By understanding community risk factors for concentration of high burden of undiagnosed TB cases, these communities can further be categorised to identify those that are at most need of additional intervention (17).

The work on TB case-fatality identified familiar risk factors of death on TB treatment that have been previously reported (12). But our finding that straight line distance estimate had the same

association with case-fatality as road network distance estimate adds to the evidence that in this population straight line distance estimate can roughly be as good as shortest road network. This conclusion is similar to what others found in a Kampala Uganda (29), using the centroid of the residential neighbourhood area of a patient as the approximate household location of the TB patient, in a retrospectively enrolled cohort of TB patients. Unlike the Kampala study, ours (Chapter 4) was based on prospectively enrolment TB patients who had their household captured at enrolled (12). Hence distance was estimated between the TB treatment facility and the patient's household (12). This is an important finding since a straight-line distance estimate is an easier estimate to measure and requires less technical capability compared to a road network distance estimate. With this evidence, TB programmes and public health officers can calculate and use the simpler straight-line distance estimate for use in public health modelling analysis of TB case fatality risk.

This PhD has developed an innovative modelling approach that can be used by a new district that is similar to urban Blantyre to prioritise its neighbourhoods for community ACF (11). This can be achieved by collecting relatively easy to measure neighbourhood variables, namely the percentage of adult residents, distance to the nearest TB clinic and the percentage of household heads who did not finish primary school, and then using the model to produce prevalence to notification ratios for neighbourhoods (Figure 6.1). These predicted prevalence to notification ratios can then be used to rank the neighbourhoods, where high rank order would indicate a high likelihood of presence of undiagnosed TB cases in that neighbourhood. This would be much less demanding compared to carrying out a TB prevalence survey.

While data on TB notifications are a readily available estimate of TB burden, in most countries –including those in the African region – they provide a suboptimal measure of the true underlying TB incidence because of missed incident cases that never get notified (10). TB hotspots based on TB notifications are bound to inherit the same bias that TB notifications have (17). TB prevalence provides the next best level evidence but on their own are also not enough to properly stratify neighbourhoods based on barriers to access to TB care. The areas with the most limited access to care will be those with a high TB prevalence rate and low TB notification rates, indicating that routine TB care in those areas is not performing optimally. (1,30). By calculating prevalence to notification ratios, this enables the identification of areas likely to have a higher burden of prevalent TB cases, especially areas where barriers to access to routine passive care surveillance system are greater (11).

Even in districts that meet the 0.5% prevalence cut-off level for recommending community ACF, the burden of undiagnosed cases in such a population would be unevenly distributed (11,17). For instance, low density high social economic neighbourhoods are likely to have a low burden of undiagnosed TB cases to warrant an ACF intervention in such communities (11,31). Since community ACF interventions are logistically challenging and resource intensive further stratification at subdistrict level to reduce the number of individuals that need to be screened is an important step (11,31). Overall, therefore, based on the findings from this thesis, we suggest that the WHO-recommended prevalence threshold for community ACF should be revised to consider the use of surveillance data and models to guide targeted implementation (Figure 1.6) (5). This is to enable achievement of the EndTB goal of equitable access to TB care, by making

sure that those that are at the most risk of undiagnosed TB and have access to care barriers are reached (30).

6.4.2 Measuring impact of spatially targeted community case finding interventions

There are a variety of outcomes that can be used to assess the impact of community interventions that are geographically targeted (8). This is because the burden of TB can be estimated using various methods as illustrated in Figure 6.1. These outcome measures are similar to those that would be utilised in a non-spatially targeted community-based ACF programme (7). A prevalence survey, can be used to compare the prevalence of active TB disease in the intervention and control arms, this is expensive and resource intensive (7). The yield of prevalent TB cases diagnosed can also be used, this is the number of individuals that have to be tested to diagnose one individual with active disease (8).

Before and after analysis of notification of TB cases from the TB hotspots can be used to assess the impact of the intervention on notifications (8). The ideal impact assessment for ACF interventions is to assess the effect on new incident cases, unfortunately incident cases are impossible to directly estimate (7). Instead, incidence is inferred by conducting prevalence or incidence surveys of latent TB infection in young children to estimate the annual risk of TB infection (7). This is also challenging because the sensitivity of latent tests can be affected by the BCG vaccine and they are expensive and require complicated logistics (32).

Community ACF interventions have both benefits at the individual level and at the community level. At the individual level they help individuals with undiagnosed active TB, some of which who would not have had access to diagnosis and treatment with passive ACF interventions to

get diagnosis and treatment (5). Hence reducing their risk of mortality, morbidity and improving their quality of life (7). At the community level screening programmes in the short term usually increase notifications and the uptake of facility-based TB care (33). This is due to the cases that get directly diagnosed by the ACF and because the general public is becomes more aware of TB symptoms and where they should go to get a TB test (33).

Community ACF interventions are required to be done in corroboration with routine facility-based health care to ensure that the respective health facilities are equipped to accommodate the demand generated by the community ACF (5,33). This preparedness includes the availability of WHO-recommended tests, treatment, and competent medical professionals at the health facilities (5). But in the long term the notifications of a disease in the ACF intervention area are expected to decrease as most cases get diagnosed and undiagnosed cases become fewer (33).

The spatially targeted intervention can also have positive effect on the general population by reducing the transmission of infection but this depends on factors such as the rate of effective contact (34) between the members of the hotspot and the members of the general population, immigration of residents between the general population and the general public and others (figure 6.1) (35). Mathematical modelling can be used to assess the impact of a spatially targeted intervention on the surrounding communities (20,35).

Spatially targeted community ACF interventions are complimentary to routine facility-based TB care just as untargeted community ACF interventions (5). Spatially targeted interventions are meant to aid in reaching individuals that have the most difficulty in accessing facility-based TB

care (5). In terms of having a sustainable solution, areas that benefit from spatially targeted ACF should be supported to have increased access to routine TB care (36). Health facilities that offer the recommended TB care need to be available at a walkable distance and be financially accessible. Other initiatives to support individuals affected by TB should be considered this includes financial support to avoid catastrophic costs in accessing TB diagnosis and treatment (37).

6.4.3 Cost-effectiveness considerations

Cost effectiveness analysis is used to inform decision makers in deciding resource allocation where there are competing interventions (38). We were only able to identify one study for cost effectiveness of spatially targeted interventions of TB.

A spatially explicit transmission model was used to evaluate the cost effectiveness of spatially targeted ACF compared to passive surveillance and untargeted community ACF, ten years from intervention commencement (20). The intervention was door to door screening for chronic cough (greater than 2 weeks) followed by 2 sputum sample submission for microscopy. The setting was rural Ethiopia (39).

The hotspot region comprised 20% of the overall population. The spatially targeted ACF intervention at a community screening coverage of 95% in the hotspot communities reduced the overall incidence of TB by 52% compared to the baseline (20). To achieve a similar decline in incidence an untargeted community ACF intervention would have to screen 80% of the overall population (20). Compared to the passive case surveillance the spatially targeted ACF intervention was expected to avert 1,023 new TB cases over ten years saving $170 per averted

case. While the untargeted ACF intervention was expected to avert 1, 316 new TB cases over ten years saving $3 per averted TB case.

To compare the cost effectiveness of the spatially targeted intervention versus the untargeted intervention, an incremental-cost effectiveness-ratio was done and the untargeted intervention averted an additional 293 more TB cases at a cost of $528 per averted cost, corresponding to an additional cost of about $170,700 for these additional averted cases. This analysis suggested that TB spatially targeted interventions would be cost effective in this population and would also significantly reduce the overall TB incidence.

6.5 Limitations

Here we acknowledge, describe and discuss the limitations of the PhD thesis. The study setting was Blantyre Malawi, a country that is classified as a low-income country by the World Bank (40). This might limit the generalisability of our models in other countries. This is especially because the highest burden of TB is currently in middle to upper-middle-income countries in Africa, Asia, South America and Eastern Europe according to the WHO 2021 report (1). Characteristics such as access to health care, determinants of TB, social-economic support systems and available health systems are likely to be different in these settings compared to Blantyre. Therefore, there is a need to carry out similar studies and analysis in other areas to which we cannot be confident our findings generalise, to identify predictors of prevalence to notification ratios for these settings and the risk factors of death on TB treatment (11).

Further, the population in the study setting have been involved in various research studies overtime (41), this might have affected the way the population participated in the research

studies that contributed to the data of this PhD (the Harthorne effect) (42). It is possible that the population could have been fatigued by research from prior participation in research and could therefore have tried to avoid participation in the research studies whose data are analysed in this PhD thesis (42). This might also affect different groups differently, for instance people living with HIV, who have frequent appointments with medical practitioners, might be more willing to get involved in research trials (42). Indeed, a qualitative study in the study area, identified that community members retained high awareness and positive opinions of community wide case finding two years after the intervention, suggesting that the Hawthorne effect made this community more positive towards participation in research (43). Other settings might have a different experience.

Malawi has also achieved excellent results in meeting the 90:90:90 targets of HIV care cascade, according to a 2020 to 2021 population HIV impact assessment (MPHIA) survey: 88.3% of people living with HIV already knew their status, 97.9% of HIV positive individuals who were aware of their positive HIV status were on HIV treatment and 96.9% of people who were on HIV treatment were virally suppressed (44). This is in contrast to other high TB and HIV burden settings that are struggling to meet their HIV care cascade goals like South Africa: there only 72.8% of people living with HIV were aware of their status, 91.5% individuals who were aware of their HIV positive status were on ART treatment and 81.4% of HIV positive individuals who were on ART treatment were virally suppressed (45). In general, other settings may have higher heterogeneity in their coverage of HIV diagnosis and treatment than Blantyre (46). Since people living with HIV who have untreated TB tend to develop symptoms earlier and potentially die sooner than those who do not have the disease, they contribute less to the transmission of TB

compared to HIV-negative individuals (47,48). In addition, since having symptoms is a precursor to seeking care, people living with HIV are likely to seek care through facility-based TB care (47,49). Also, when on HIV treatment, scheduled visits at health facilities for HIV treatment refill offers an opportunity for TB screening (50). Thus, our models would have a limited generalisation to areas who have an uncontrolled HIV epidemic since they were based on data from Blantyre. Settings that have high TB and high HIV undiagnosed cases the type of targeted interventions might also have to be different than those that were done in low TB and low HIV settings as reported in the systematic review in Chapter 3 (8,18).

While the focus of my PhD thesis is improving community ACF methods and TB care in health facilities, it is important to keep in mind that TB is exacerbated by low social economic status factors such as poor nutrition and poor housing conditions (4). Therefore, interventions to reduce incidence should be multisectoral, and not only focus on medical interventions (4). Historical data from Europe show that TB incidence was falling as the economic conditions were improving even before improvements in TB case-finding and treatment were implemented (4). Other Asian countries experienced similar trends (4). The WHO guideline recognises this and has highlighted all the sustainable goals that are related towards achieving TB eradication goals (51). Mathematical modelling estimated that ending extreme poverty and expanding social protection coverage could result in about 11% annual reduction in TB incidence (25).

I have written the thesis with emphasis on improving community ACF and facility-based care of TB patients and have given little attention to how social economic interventions could be implemented to reduce inequalities that are responsible for TB risk heterogeneity, this was out of the scope of my PhD thesis. Disease control programmes have limited resources that they

can use for interventions, hence they allocate resources to activities that they can afford to, improving housing conditions and nutrition of individuals may improve populations resilience to TB disease but these interventions would be out of the budget of disease control programmes (38). Social economic status was included as a confounder in the models of the PhD thesis but measuring social economic status in a setting with a lot of informal employment like Blantyre is hard. I used completion of primary school education of head of household in the model of Chapter 5, but it is possible that this indicator misclassified households (11), although a Malawi Integrated Housing Survey found that the household head education level was associated with poverty (52). In the case-fatality analysis we used household ownership of items to develop an indicator of social economic status (12), again this can be biased, as individuals can end up being classified as belonging to upper social economic status because they have the specified household items but when they are poor in terms of income available for food and health care (53).

The prevalence survey that contributed to the hotspot identification analysis was a one-off cross-sectional survey (9). It is possible that a different prevalence survey in the same setting would have resulted into the analysis identifying a different set of hotspots (9). Nevertheless, the prevalence survey was based on random sampled households in the setting and was less likely to be affected by selection bias of the households who took part in the survey (11). We were also unable to verify if the hotspots would remain with time because we only had one prevalence survey. Ideally, we were supposed to have data from sequential prevalence surveys to assess the permanency of the hotspots.

In routine care in Blantyre, symptom screening is used and when available a chest X-ray can also be done at the recommendation of a physician (54). By contrast in the prevalence survey screening used a combination of symptom screening and chest X-ray, this increased the sensitivity of the screening algorithm (11). The accuracy of the TB diagnostic tests that were used in the prevalence survey and the routine notification surveillance are important, this helps ensure that only people with active TB cases are diagnosed as such (55). In routine care, a diagnosis can also be done based on a clinical judgement by the physician regardless of the absence of a confirmatory positive result (54). While in the prevalence survey a TB case had to have a positive confirmatory TB test result (11). The gold standard test for TB diagnosis is the TB culture test (47). In Blantyre enhanced laboratory support (11) was able to provide MTB culture tests to routinely collected samples and the prevalence survey participants. Both modelling analyses have a primary analysis and a sensitivity analysis with one based on all the cases and another based on only confirmed MTB diagnosed cases, the results of the primary analysis and the sensitivity analysis agreed (11,12).

The prevalence survey identified sparse prevalent cases (11). This meant that the predictive models were unlikely to have enough power to identify more than one or two predictive covariates for prevalence. However, prevalence surveys are extremely costly and resource intensive, and only a few ever get conducted at the city scale (11,23). In addition, to the best of our knowledge, there is no established model based on covariates that predicts prevalence. Indeed, in a recent publication from Pakistan using subnational prevalence survey data from Pakistan, five separate research groups were unable to identify a unifying set of predictors for prevalence (23).

The models that were run in our analysis lack external validation. A natural next step is therefore to repeat this analysis in a different setting to see if the conclusions arrived at are like the ones in this analysis. The prevalence to notification analysis could be done in this different setting and we could check the predicators that would be identified, and the hotspots identified. The same applies to our analysis of case fatality.

6.6. Future directions of research

As noted above, model validation using data from a different setting would help improve our understanding of how the models perform in other settings. This could be achieved by collecting data from another setting that is similar to Blantyre and re-producing the models of Chapter 4 and 5 using the data from that setting. This would enable us to redo our analysis to check if the predictors identified in the validation setting are similar to those that we identified in Blantyre, and also to check the characteristics of the identified hotspots. Depending on the results of the model validation exercise, the models could remain the same or be changed to accommodate the newly found knowledge.

The identification of hotspots in Chapter 5 was based on prevalence of active TB cases and TB notifications. These identified prevalent active TB cases might have contracted the infection elsewhere other than their residential neighbourhoods (18). Therefore, it is not possible to know whether these identified hotspots are also areas of high transmission of TB (18). To address this, we might focus future work on prevalence of LTBI in under 5-year-olds, in order to better understand TB transmission. This is because LTBI in under 5-year-olds is a recent event and can be assumed to have occurred while the child was a resident of the neighbourhood.

Therefore, it can be used in identifying areas of high TB transmission (7). In addition, genetic analysis to find transmission links between TB isolates would enable to identify areas where transmission is taking place (18). Combining this information with the approach developed in this thesis should further improve identification of areas where transmission is occurring. This in turn will allow development of appropriate interventions for these areas to interrupt transmission, and to protect vulnerable populations that are at risk of progressing to active TB disease such as people living with HIV (18).

Calculation of prevalence and notification rates relies on denominators that accurately represent the population. National census data are a common source of denominators but they are done periodically and can get out of date. Linear interpolation of the census data to project the population in the future, based on observed population growth rates, can be used to update population estimates (11). This assumes that the population growth that was observed in the past will be the same in the future, but this assumption can be broken if the growth rates changes. Census surveys are also done at census tracts level, these differ from the geographical demarcations that are used as the catchment area for health facilities. Alternatively, the worldPop project has produced 100 by 100 m gridded population estimates based on census data using machine learning (19,56). These do allow for arbitrary neighbourhood demarcations to have census-based estimates once a defined shapefile boundary is submitted to the WorldPop API to obtain the population estimates (56).

Research is only good if it gets adopted into policy and is implemented. For this to be possible there is a need to engage with the technical working groups within the National TB programmes and also the WHO. Since resource allocation is always competitive, to justify

allocation of the resources to spatially targeted interventions, cost effective analysis work needs to be done to prove that this would be cost effective (20). This can be achieved by using transmission models parameterised with data from the setting of interest and also based on empirically collected data from trials (20,38).

6.7. Conclusion

In summary, by using high-resolution surveillance data at the neighbourhood-level to better understand the epidemiology of TB in urban Africa, this thesis has confirmed that there are high-levels of spatial heterogeneity in the risk factors for TB, burden of TB disease, and access to high quality TB diagnosis and care. The approaches advanced here could be used in other similar urban settings to prioritise neighbourhoods by burden of undiagnosed TB, and efficiently direct community-based ACF interventions. WHO supports countries in the collection of notification and prevalence data; spatially explicit models that capture the spatial heterogeneity of the TB epidemic can help countries to stratify communities into areas of low risk to high risk of undiagnosed TB. As the TB epidemic declines and undiagnosed TB cases concentrate among high-risk and vulnerable populations to form hotspots of undiagnosed TB, spatially targeted ACF interventions are going to come to the forefront.

This thesis suggests that targeting these neighbourhood hotspots could contribute to accelerated improvements in TB case detection and reduction in transmission. This thesis opens promising avenues of research that that involves the use of spatially explicit models to direct TB programmes where to do community ACF interventions. It also highlights the need for additional evidence for the impact assessment of spatially targeted interventions involving

random allocation of hotspots in evaluation of outcome of interest and cost effectiveness analysis. This evidence could be informed by empirical studies or spatially explicit transmission models.

6.8 References

1. World Health Organization (WHO). Global tuberculosis report 2021 [Internet]. Geneva; 2021. Available from: https://www.who.int/publications/i/item/9789240037021

2. Ho J, Fox GJ, Marais BJ. Passive case finding for tuberculosis is not enough. Int J Mycobacteriology [Internet]. 2016 Dec 1 [cited 2019 Mar 7];5(4):374–8. Available from: https://www.sciencedirect.com/science/article/pii/S2212553116301595?via%3Dihub

3. Golub JE, Mohan CI, Comstock GW, Chaisson RE. Active case finding of tuberculosis: historical perspective and future prospects. Int J Tuberc Lung Dis [Internet]. 2005 Nov [cited 2022 May 12];9(11):1183. Available from: /pmc/articles/PMC4472641/

4. Lönnroth K, Jaramillo E, Williams BG, Dye C, Raviglione M. Drivers of tuberculosis epidemics: The role of risk factors and social determinants. Soc Sci Med. 2009;68(12):2240–6.

5. World Health Organisation. Module 2: Screening WHO operational handbook on tuberculosis Systematic screening for tuberculosis disease [Internet]. Geneva; 2021 [cited 2021 Apr 6]. Available from: https://reliefweb.int/report/world/who-consolidated-guidelines-tuberculosis-module-2-screening-systematic-screening

6. Kendall EA, Kitonsa PJ, Nalutaaya A, Erisa KC, Mukiibi J, Nakasolya O, et al. The Spectrum of Tuberculosis Disease in an Urban Ugandan Community and Its Health Facilities. Clin Infect Dis [Internet]. 2021 Jun 15 [cited 2022 Jun 11];72(12):e1035–43. Available from: https://academic.oup.com/cid/article/72/12/e1035/6024999

7. Burke RM, Nliwasa M, Feasey HRA, Chaisson LH, Golub JE, Naufal F, et al. Community-based active case-finding interventions for tuberculosis: a systematic review. Lancet Public Heal [Internet]. 2021;6(5):e283–99. Available from: http://dx.doi.org/10.1016/S2468-2667(21)00033-5

8. Khundi M, Carpenter JR, Nliwasa M, Cohen T, Corbett EL, MacPherson P. Effectiveness of spatially targeted interventions for control of HIV, tuberculosis, leprosy and malaria: a systematic review. BMJ Open [Internet]. 2021 Jul 1 [cited 2021 Jul 20];11(7):e044715. Available from: https://bmjopen.bmj.com/content/11/7/e044715

9. Shaweno D, Karmakar M, Alene KA, Ragonnet R, Clements AC, Trauer JM, et al. Methods used in the spatial analysis of tuberculosis epidemiology: a systematic review. BMC Med [Internet]. 2018 [cited 2019 Feb 20];16(1):193. Available from: https://bmcmedicine.biomedcentral.com/articles/10.1186/s12916-018-1178-4

10. Dodd PJ, Sismanidis C, Glaziou P. Methods for estimating tuberculosis incidence and mortality by age and sex. Int J Epidemiol [Internet]. 2021 May 17 [cited 2021 Jun 28];50(2):570–7. Available from: https://academic.oup.com/ije/article/50/2/570/6149009

11. Khundi M, Carpenter JR, Corbett EL, Feasey Id HRA, Soko RN, Nliwasa M, et al. Neighbourhood prevalence-to-notification ratios for adult bacteriologically-confirmed tuberculosis reveals hotspots of underdiagnosis in Blantyre, Malawi. Fenner L, editor. PLoS One [Internet]. 2022 May 23 [cited 2022 May 25];17(5):e0268749. Available from: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0268749

12. Khundi M, MacPherson P, Feasey HRA, Soko RN, Nliwasa M, Corbett EL, et al. Clinical, health systems and neighbourhood determinants of tuberculosis case fatality in urban Blantyre, Malawi: a multilevel epidemiological analysis of enhanced surveillance data. Epidemiol Infect [Internet]. 2021 [cited 2021 Nov 19];149. Available from: https://www.cambridge.org/core/journals/epidemiology-and-infection/article/clinical-health-systems-and-neighbourhood-determinants-of-tuberculosis-case-fatality-in-urban-blantyre-malawi-a-multilevel-epidemiological-analysis-of-enhanced-surveillance-data/F1

13. Mayer CM, Owaraganise A, Kabami J, Kwarisiima D, Koss CA, Charlebois ED, et al. Distance to clinic is a barrier to PrEP uptake and visit attendance in a community in rural Uganda. J Int AIDS Soc [Internet]. 2019 Apr 1 [cited 2020 Sep 23];22(4):e25276. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=prem&NEWS=N&AN=31037845

14. Wold Health Organisation. Systematic screening for active tuberculosis. Geneva; 2013.

15. Momenyan S, Kavousi A, Poorolajal J, Momenyan N. Spatial inequalities and predictors of HIV/AIDS mortality risk in Hamadan, Iran: a retrospective cohort study. Epidemiol Health [Internet]. 2018;40(101519472):e2018038. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med13&NEWS=N&AN=30081619

16. O'Donnell O. Access to health care in developing countries: Breaking down demand side barriers [Internet]. Vol. 23, Cadernos de Saude Publica. Fundacao Oswaldo Cruz; 2007 [cited 2020 Oct 12]. p. 2820–34. Available from: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0102-311X2007001200003&lng=en&nrm=iso&tlng=en

17. Shaweno D, Karmakar M, Alene KA, Ragonnet R, Clements AC, Trauer JM, et al. Methods used in the spatial analysis of tuberculosis epidemiology: a systematic review. BMC Med. 2018;16(1):193.

18. Cudahy PGT, Andrews JR, Bilinski A, Dowdy DW, Mathema B, Menzies NA, et al. Spatially targeted screening to reduce tuberculosis transmission in high-incidence settings. Lancet Infect Dis [Internet]. 2018 Dec 13 [cited 2019 Feb 16];0(0). Available from: http://www.ncbi.nlm.nih.gov/pubmed/30554997

19. MacPherson P, Khundi M, Nliwasa M, Choko AT, Phiri VK, Webb EL, et al. Disparities in access to diagnosis and care in Blantyre, Malawi, identified through enhanced tuberculosis surveillance and spatial analysis. BMC Med [Internet]. 2019 Dec 29 [cited 2019 Feb 6];17(1):21. Available from: https://bmcmedicine.biomedcentral.com/articles/10.1186/s12916-019-1260-6

20. Shaweno D, Trauer JM, Doan TN, Denholm JT, McBryde ES. Geospatial clustering and modelling provide policy guidance to distribute funding for active TB case finding in Ethiopia. Epidemics. 2021 Sep 1;36.

21. MacPherson P, Khundi M, Nliwasa M, Choko AT, Phiri VK, Webb EL, et al. Disparities in access to diagnosis and care in Blantyre, Malawi, identified through enhanced tuberculosis surveillance and spatial analysis. BMC Med. 2019;17(1):21.

22. Robsky KO, Robsky KO, Kitonsa PJ, Mukiibi J, Nakasolya O, Isooba D, et al. Spatial distribution of people diagnosed with tuberculosis through routine and active case finding: a community-based study in Kampala, Uganda. Infect Dis Poverty [Internet]. 2020 Jun 22 [cited 2022 May 26];9(1). Available from: /pmc/articles/PMC7310105/

23. Alba S, Rood E, Mecatti F, Ross JM, Dodd PJ, Chang S, et al. TB Hackathon: Development and

Comparison of Five Models to Predict Subnational Tuberculosis Prevalence in Pakistan. Vol. 7, Tropical Medicine and Infectious Disease. 2022.

24. Brooks MB, Millones AK, Puma D, Contreras C, Jimenez J, Tzelios C, et al. Mapping local hot spots with routine tuberculosis data: A pragmatic approach to identify spatial variability. Mollalo A, editor. PLoS One [Internet]. 2022 Mar 24 [cited 2022 Mar 27];17(3):e0265826. Available from: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0265826

25. Carter DJ, Glaziou P, Lönnroth K, Siroka A, Floyd K, Weil D, et al. The impact of social protection and poverty elimination on global tuberculosis incidence: a statistical modelling analysis of Sustainable Development Goal 1. Lancet Glob Heal [Internet]. 2018 May 1 [cited 2022 May 30];6(5):e514–22. Available from: https://pubmed.ncbi.nlm.nih.gov/29580761/

26. Law I, Floyd K, Abukaraig EAB, Addo KK, Adetifa I, Alebachew Z, et al. National tuberculosis prevalence surveys in Africa, 2008–2016: an overview of results and lessons learned. Trop Med Int Heal [Internet]. 2020 Nov 12 [cited 2020 Nov 12];25(11):1308–27. Available from: https://onlinelibrary.wiley.com/doi/10.1111/tmi.13485

27. World Health Organization (WHO). National tuberculosis prevalence surveys 2007-2016 [Internet]. Geneva; 2021 [cited 2022 Mar 1]. Available from: https://www.who.int/publications/i/item/9789240022430

28. United Nations Human Settlements Programme. SLUM ALMANAC 2015 2016 Tracking Improvement in the Lives of Slum Dweller [Internet]. Nairobi; 2015. Available from: https://unhabitat.org/slum-almanac-2015-2016/

29. Robsky KO, Robsky KO, Hughes S, Kityamuwesi A, Kendall EA, Kitonsa PJ, et al. Is distance associated with tuberculosis treatment outcomes? A retrospective cohort study in Kampala, Uganda. BMC Infect Dis [Internet]. 2020 Jun 11 [cited 2020 Sep 23];20(1):406. Available from: https://bmcinfectdis.biomedcentral.com/articles/10.1186/s12879-020-05099-z

30. Borgdorff MW. New measurable indicator for tuberculosis case detection. Emerg Infect Dis [Internet]. 2004 Sep [cited 2019 Feb 6];10(9):1523–8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15498151

31. Brooks MB, Jenkins HE, Puma D, Tzelios C, Millones AK, Jimenez J, et al. A role for community-level socioeconomic indicators in targeting tuberculosis screening interventions. Sci Rep [Internet]. 2022 Dec 17 [cited 2022 Jan 20];12(1):781. Available from: https://pubmed.ncbi.nlm.nih.gov/35039612/

32. World Health Organization. Latent tuberculosis infection Updated and consolidated guidelines for programmatic management. Geneva; 2018.

33. Feasey Id HRA, Burke Id RM, Id MN, Chaisson Id LH, Golub Id JE, Id N, et al. Do community-based active case-finding interventions have indirect impacts on wider TB case detection and determinants of subsequent TB testing behaviour? A systematic review. PLOS Glob Public Heal [Internet]. 2021 Dec 8 [cited 2022 Mar 23];1(12):e0000088. Available from: https://journals.plos.org/globalpublichealth/article?id=10.1371/journal.pgph.0000088

34. Thindwa D, Jambo KC, Ojal J, MacPherson P, Dennis Phiri M, Pinsent A, et al. Social mixing patterns relevant to infectious diseases spread by close contact in urban Blantyre, Malawi. Epidemics [Internet]. 2022 Sep [cited 2022 Jun 16];40:100590. Available from:

https://pubmed.ncbi.nlm.nih.gov/35691100/

35. Dowdy DW, Golub JE, Chaisson RE, Saraceni V. Heterogeneity in tuberculosis transmission and the role of geographic hotspots in propagating epidemics. Proc Natl Acad Sci [Internet]. 2012;109(24):9557–62. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med8&NEWS=N&AN=22645356

36. Trauer JM, Dodd PJ, Gomes MGM, Gomez GB, Houben RMGJ, McBryde ES, et al. The Importance of Heterogeneity to the Epidemiology of Tuberculosis. Clin Infect Dis [Internet]. 2019 Jun 18 [cited 2022 Jun 14];69(1):159–66. Available from: https://academic.oup.com/cid/article/69/1/159/5154892

37. Ukwaja KN. Social protection interventions could improve tuberculosis treatment outcomes. Lancet Glob Heal [Internet]. 2019 Feb 1 [cited 2022 May 28];7(2):e167–8. Available from: http://www.thelancet.com/article/S2214109X18305230/fulltext

38. Caroline Dobler C, Dobler C. Screening strategies for active tuberculosis: focus on cost-effectiveness. Clin Outcomes Res [Internet]. 2016 Jun 30 [cited 2022 Jun 14];8:335–47. Available from: https://www.dovepress.com/screening-strategies-for-active-tuberculosis-focus-on-cost-effectivene-peer-reviewed-fulltext-article-CEOR

39. Shaweno D, Trauer JM, Denholm JT, McBryde ES. A novel Bayesian geospatial method for estimating tuberculosis incidence reveals many missed TB cases in Ethiopia. BMC Infect Dis [Internet]. 2017 Dec 2 [cited 2019 Feb 11];17(1):662. Available from: http://www.ncbi.nlm.nih.gov/pubmed/28969585

40. World Bank. The world bank data [Internet]. [cited 2022 Jun 8]. Available from: https://data.worldbank.org/country/XM

41. Choko AT, Corbett EL, Stallard N, Maheswaran H, Lepine A, Johnson CC, et al. HIV self-testing alone or with additional interventions, including financial incentives, and linkage to care or prevention among male partners of antenatal care clinic attendees in Malawi: An adaptive multi-arm, multi-stage cluster randomised trial. PLOS Med [Internet]. 2019 [cited 2022 Jun 20];16(1):e1002719. Available from: https://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.1002719

42. McCambridge J, Witton J, Elbourne DR. Systematic review of the Hawthorne effect: New concepts are needed to study research participation effects. J Clin Epidemiol [Internet]. 2014 [cited 2022 Jun 20];67(3):267. Available from: /pmc/articles/PMC3969247/

43. Monk EJM, Kumwenda M, Nliwasa M, Mpunga J, Corbett EL. Factors affecting tuberculosis health message recall 2 years after active case finding in Blantyre, Malawi. Int J Tuberc Lung Dis [Internet]. 2018 Sep 1 [cited 2020 Oct 29];22(9):1007–15. Available from: /pmc/articles/PMC6086287/?report=abstract

44. The Elizabeth Glaser Pediatric AIDS Foundation (EGPAF). MALAWI EXCELS ON UNAIDS 95-95-95 TREATMENT AND VIRAL SUPPRESSION TARGETS The Elizabeth Glaser Pediatric AIDS Foundation (EGPAF) [Internet]. Lilongwe; 2021 [cited 2022 Jun 7]. Available from: https://www.pedaids.org/wp-content/uploads/2021/12/2021MalawiInfographic_1215.pdf

45. Lippman SA, El Ayadi AM, Grignon JS, Puren A, Liegler T, Venter WDF, et al. Improvements in the South African HIV care cascade: findings on 90-90-90 targets from successive population-

representative surveys in North West Province. J Int AIDS Soc [Internet]. 2019 Jun 1 [cited 2022 Jun 7];22(6). Available from: /pmc/articles/PMC6562149/

46. Dwyer-Lindgren L, Cork MA, Sligar A, Steuben KM, Wilson KF, Provost NR, et al. Mapping HIV prevalence in sub-Saharan Africa between 2000 and 2017. Nat 2019 5707760 [Internet]. 2019 May 15 [cited 2022 Jun 21];570(7760):189–93. Available from: https://www.nature.com/articles/s41586-019-1200-9

47. Pai M, Behr MA, Dowdy D, Dheda K, Divangahi M, Boehme CC, et al. Tuberculosis. Nat Rev Dis Prim 2016 21 [Internet]. 2016 Oct 27 [cited 2021 Sep 28];2(1):1–23. Available from: https://www.nature.com/articles/nrdp201676

48. Ku C-C, MacPherson P, Khundi M, Nzawa Soko RH, Feasey HRA, Nliwasa M, et al. Durations of asymptomatic, symptomatic, and care-seeking phases of tuberculosis disease with a Bayesian analysis of prevalence survey and notification data. BMC Med 2021 191 [Internet]. 2021 Nov 10 [cited 2021 Nov 16];19(1):1–13. Available from: https://bmcmedicine.biomedcentral.com/articles/10.1186/s12916-021-02128-9

49. Ku C-C, MacPherson P, Khundi M, Nzawa R, Feasey HR, Nliwasa M, et al. Estimated durations of asymptomatic, symptomatic, and care-seeking phases of tuberculosis disease. medRxiv [Internet]. 2021 Apr 4 [cited 2021 Aug 26];2021.03.17.21253823. Available from: https://www.medrxiv.org/content/10.1101/2021.03.17.21253823v1

50. Getahun H, Kittikraisak W, Heilig CM, Corbett EL, Ayles H, Cain KP, et al. Development of a Standardized Screening Rule for Tuberculosis in People Living with HIV in Resource-Constrained Settings: Individual Participant Data Meta-analysis of Observational Studies. PLOS Med [Internet]. 2011 [cited 2022 Jun 20];8(1):e1000391. Available from: https://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.1000391

51. Lö K, Raviglione M. The WHO's new End TB Strategy in the post-2015 era of the Sustainable Development Goals. Trans R Soc Trop Med Hyg. 2016;110:148–50.

52. Malawi National Statistical Office. Intergrated Household Survey (IHS 3) Report 2010-2011. Zomba; 2012.

53. Payongayong E, Benson T, Ahmed A, Kanyanda S, Chakanza C, Mwanza P, et al. Simple household poverty assessment models for Malawi: proxy means test from the 1997–98 Malawi Integrated Household Survey. Lilongwe, Malawi: National Statistical Office, Government of Malawi, 2006.

54. Nyirenda T. Epidemiology of Tuberculosis in Malawi. Malawi Med J [Internet]. 2006 Oct 16 [cited 2022 Jun 20];18(3):147. Available from: /pmc/articles/PMC3345620/

55. World Health Organization (WHO). WHO consolidated guidelines on tuberculosis rapid diagnostics for tuberculosis detection; Module 3: Diagnosis. [Internet]. Geneva; [cited 2022 Mar 1]. Available from: https://www.who.int/publications/i/item/9789240029415

56. Linard C, Gilbert M, Snow RW, Noor AM, Tatem AJ. Population distribution, settlement patterns and accessibility across Africa in 2010. PLoS One [Internet]. 2012 Feb 21 [cited 2022 Jun 20];7(2). Available from: https://pubmed.ncbi.nlm.nih.gov/22363717/

Appendix

Appendix: 1 Ethics approval

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT
United Kingdom
Switchboard: +44 (0)20 7636 8636

www.lshtm.ac.uk

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

Observational / Interventions Research Ethics Committee

Prof Liz Corbett
LSHTM

21 May 2019

Dear Prof Liz

Study Title: Sustainable Community-wide Active case finding for Lung hEalth (SCALE)

LSHTM Ethics Ref: 16228 - 1

Thank you for your application for the above amendment to the existing ethically approved study and submitting revised documentation. The amendment application has been considered by the Interventions Committee.

Confirmation of ethical opinion

behalf of the Committee, I am pleased to confirm a favourable ethical opinion for the above amendment to research on the basis described in the application form, protocol and supporting umentation as revised, subject to the conditions specified below.

Conditions of the favourable opinion

Approval is dependent on local ethical approval for the amendment having been received, where relevant.

Approved documents

The final list of documents reviewed and approved by the Committee is as follows:

Document Type	File Name	Date	Version
Local Approval	20181218 SCALE study response from COMREC (1)	14/12/2018	1
Other	TS05FMa Consent form for TB Prevalance Survey_English v2.0 Highlighted changes	07/01/2019	2
Other	TS05FMb Consent form for TB Prevalance Survey_Chichewa v2.0, 7 Jan 2019 Highlighted Changes	07/01/2019	2
Other	PS05FMa Consent form for TB Post Prevalance Survey v2.0 Highlighted Changes	07/01/2019	2
Local Approval	SCALE Corbett 2556 expedited review response (2)	08/02/2019	1
Local Approval	SCALE Corbett 2556 expedited review response (3)	28/03/2019	1
ier	CF01FLa_Information Sheet for TB ithe_ACFTB_SCALE Study_Eng_v4.0_04.04.2019 Changes highlighted	04/04/2019	4
Other	CF01FLb_Information Sheet for TB ithe_ACFTB_SCALE Study_chi_v4.0_04.04.2019 Changes highlighted	04/04/2019	4
Other	TS02a Pre-Intervention Prevalence Survey Participant Information Sheet_v3.0_English_changes highlighted	04/04/2019	3
Other	TS02b Participant Information Sheet for Individual Interviews_v3.0_Chichewa Changes highlighted	04/04/2019	3
Other	TS65FM X-RAY Referral Card v1.0	04/04/2019	1
Other	SCALE Trial LSHTM protocol v4.0 05.04.19	05/04/2019	4
Other	TS01a General Flyer TB Prevalence Survey English_v3.0_Changes highlighted	05/04/2019	3
Other	TS01b General Flyer TB Prevalence Survey Chichewa v3.0 changes highlighted	05/04/2019	3
Other	TS09b X-ray Consent Form Chichewa 3.0, changes highlighted	05/04/2019	3
Other	TS09b X-ray Consent Form Chichewa 3.0, changes highlighted	05/04/2019	3
Other	PS05FMb Consent form for TB Post Prevalance Survey_7 January 2019 v2.0	07/04/2019	2

After ethical review

The Chief Investigator (CI) or delegate is responsible for informing the ethics committee of any subsequent changes to the application. These must be submitted to the Committee for review

using an Amendment form. Amendments must not be initiated before receipt of written favourable opinion from the committee.

The CI or delegate is also required to notify the ethics committee of any protocol violations and/or Suspected Unexpected Serious Adverse Reactions (SUSARs) which occur during the project by submitting a Serious Adverse Event form.

An annual report should be submitted to the committee using an Annual Report form on the anniversary of the approval of the study during the lifetime of the study.

At the end of the study, the CI or delegate must notify the committee using an End of Study form.

All aforementioned forms are available on the ethics online applications website and can only be submitted to the committee via the website at: http://leo.lshtm.ac.uk

Additional information is available at: www.lshtm.ac.uk/ethics

Yours sincerely,

[redacted]

Professor John DH Porter
Chair

ethics@lshtm.ac.uk
http://www.lshtm.ac.uk/ethics/

Improving health worldwide

CERTIFICATE OF ETHICS APPROVAL

This is to certify that the College of Medicine Research and Ethics Committee (COMREC) has reviewed and approved a study entitled:

P.12/18/2556 - Sustainable Community-wide Active Case Finding for Lung hEalth (SCALE) Version 4.0 dated 5 April 2019 by Prof. Liz Corbett

On 05-May-19

As you proceed with the implementation of your study, we would like you to adhere to international ethical guidelines, national guidelines and all requirements by COMREC some of which are indicated on the next page for your study

Approved by College of Medicine

[redacted]

Dr. YB. Mlombe - Chairperson (COMREC)

05-May-19

06-May-2019

Date

(COMREC) Research and Ethics Committee