INVESTIGATING THE SOCIO-ECONOMIC AND EPIDEMIOLOGICAL RISK FACTORS ASSOCIATED WITH TB TRANSMISSION IN A HIGH TB AND HIV BURDENED COMMUNITY IN CAPE TOWN, SOUTH AFRICA

By
RABECCA TADOKERA
STUDENT NUMBER: TDKRAB001

Submitted to The University Of Cape Town
In Partial Fulfilment Of The Requirements For The Degree of
MASTERS IN PUBLIC HEALTH
in the School of Public Health and Family Medicine, Faculty of Health Sciences
at the
UNIVERSITY OF CAPE TOWN

FEBRUARY 2016

Supervisor: Keren Middelkoop, Desmond Tutu HIV Centre, Department of Medicine, University of Cape Town
Co-Supervisor: Linda-Gail Bekker, Desmond Tutu HIV Centre, Department of Medicine, University of Cape Town
The copyright of this thesis vests in the author. No quotation from it or information derived from it is to be published without full acknowledgement of the source. The thesis is to be used for private study or non-commercial research purposes only.

Published by the University of Cape Town (UCT) in terms of the non-exclusive license granted to UCT by the author.
DECLARATION

I, RABECCA TADOKERA, hereby declare that the work on which this dissertation/thesis is based is my original work (except where acknowledgements indicate otherwise) and that neither the whole work nor any part of it has been, is being, or is to be submitted for another degree in this or any other university.

I empower the university to reproduce for the purpose of research either the whole or any portion of the contents in any manner whatsoever.

Signed

Signature: .......... ........................................

Date: February 2016
Acknowledgements

I am very grateful to my supervisors, mentors and peers for their support and help throughout this thesis journey. Keren, thank you for all your advice, support and help throughout this process. Thank for your attention to detail! Linda-Gail, I am grateful for all the mentorship and guidance and for always making time despite your very busy schedule.

I am thankful to Monika Kamkuenah for all the statistical advice and help. I am also grateful to my friends for moral support throughout this journey whenever I needed it. I am especially grateful to my family for all the love, support and their patient endurance. Ebenezer!
ABSTRACT

Background

While several studies have studied the associations between biological factors such as HIV-status with TB transmission or clustering, our understanding of the associations between TB transmission and socio-economic risk factors for TB remains incomplete. More studies are required to enhance our understanding, and hence inform targeted interventions to curb TB transmission, particularly in high burden communities. This study aimed to explore the associations between TB transmission and socio-economic risk factors in one such high TB and HIV burdened community.

Methods

A cross-sectional molecular epidemiology study was conducted among adult TB patients resident in a geographically well-defined peri-urban township of Cape Town between 2001 and 2010. Following informed consent, clinical and demographic data were extracted from TB registers and clinical folders. Additional socio-economic data were collected using interviewer-administered questionnaires that were designed to capture data on TB history, TB contacts, socio-economic conditions such as occupation, income level, educational level, sexual behaviour, sexual history in addition to other social and demographic data.
*M.tb* isolates from TB patients were previously analysed using *IS6110*-based RFLP. Strains with <6 copies of *IS6110* (low bandwidth strains) are known to be poorly differentiated and so were excluded from analysis. Composite variables were generated for the social and economic factors using a scoring algorithm to create a “social score” and an “economic score”. Data was analysed using StataCorp version 12 software. Bivariate associations and adjusted binary logistic regression analyses were performed to determine associations between TB transmission and the social/economic score in addition to other risk factors that were studied.

**Results:**

Of the 509 participants who had complete data available, 352 (69%) were classified as clustered while the remaining 157 participants (31%) were classified as non-clustered. Our analysis showed that clustered cases were more likely to have stayed for a longer period in the study community, (OR=1.06, C.I: 1.02 to 1.10, p=0.006). Clustered cases were also more likely to have stayed in the same house for longer, (median=3 years vs. 2 years, p=0.06) and to live in more crowded conditions as shown by the size of the house and number of rooms used for sleeping (p=0.038). While the evidence was weak, there was a tendency towards a positive association between a high social score and clustering (OR=1.39, C.I: 0.94; 2.03, p=0.08). Conversely, there was a moderate negative association between a high economic score and clustering (OR=0.69, C.I: 0.45; 1.06, p=0.09).

**Conclusions**
While the association between poverty (poor socio-economic status) and TB transmission is not new, the association between TB transmission and prolonged stay within a high burdened community that we report in this study is novel. Our findings further suggest that even in poorer communities there is a “sliding-scale of poverty”, with individuals at the lower end of the economic scale being at greater risk for acquiring TB infection and that targeted interventions to address TB transmission in such high burdened communities may be required.
5.2 Study-design

5.3 Inclusion and Exclusion criteria

5.4 Patient Enrolment Procedures

5.4 Data Collection

5.5 Validity and variability of questionnaires

5.6 Laboratory Procedures

5.7 Sample Size and power calculations

6. Data Analysis Plan

6.1 Data Management

6.2 Measurements and variables

7. Ethical Considerations

8. Communication

9. Logistics

10. References

PART B: LITERATURE REVIEW

1. Molecular Epidemiology of Tuberculosis: Overview

2. Search Strategy and Selection Criteria

3. Global Epidemiology of Tuberculosis

4. Molecular epidemiology of TB

4.1 Molecular Typing Techniques

4.2 Socio-economic risk factors for TB

4.3 Public Health Applications of Molecular Epidemiology

4.4 Molecular epidemiology of TB in High Income countries

4.5 Molecular Epidemiology of TB in LMIC

4.6 Molecular Epidemiology of TB in South Africa

5. Summary

PART C: MANUSCRIPT

Title: Investigating the association between TB transmission and socio-economic risk factors in a high TB and HIV burdened community in Cape Town, South Africa

Abstract

Introduction
### Abbreviations and Acronyms

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AIDS</td>
<td>Acquired Immune Deficiency Syndrome</td>
</tr>
<tr>
<td>ART</td>
<td>Antiretroviral Therapy</td>
</tr>
<tr>
<td>CDC</td>
<td>Centre for Disease Control</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic acid</td>
</tr>
<tr>
<td>DTHC</td>
<td>Desmond Tutu HIV Foundation</td>
</tr>
<tr>
<td>GCP</td>
<td>Good Clinical Practise</td>
</tr>
<tr>
<td>HIV</td>
<td>Human Immunodeficiency Virus</td>
</tr>
<tr>
<td>IS6110</td>
<td>Insertion Sequence 6110</td>
</tr>
<tr>
<td>HREC</td>
<td>Human Research Ethics Committee</td>
</tr>
<tr>
<td>MIRU</td>
<td>Mycobacterial Interspersed Repetitive Units</td>
</tr>
<tr>
<td>MDG</td>
<td>Millennium Development Goals</td>
</tr>
<tr>
<td>M.t.b</td>
<td>Mycobacterium tuberculosis</td>
</tr>
<tr>
<td>PCR</td>
<td>Polymerase Chain Reaction</td>
</tr>
<tr>
<td>RFLP</td>
<td>Restriction Fragment Length Polymorphism</td>
</tr>
<tr>
<td>RSA</td>
<td>Republic of South Africa</td>
</tr>
<tr>
<td>TB</td>
<td>Tuberculosis</td>
</tr>
<tr>
<td>UCT</td>
<td>University of Cape Town</td>
</tr>
<tr>
<td>VNTR</td>
<td>Variable Number Tandem Repeat</td>
</tr>
<tr>
<td>WHO</td>
<td>World Health Organisation</td>
</tr>
</tbody>
</table>
Definitions

**Discriminatory power** - ability of a typing method to distinguish between different strains

**Clone** - isolates derived from the same parent strain, or isolates/strains that are indistinguishable from each other

**Cluster** - two or more TB isolates in a given geographical area during a specified time period with the same genotype (identical DNA hybridization banding pattern)

**Family** – a collection of related strains; strains that exhibited similar, but not identical IS6110 hybridization profiles

**Genotyping** - a method of identifying strains by analysing their nucleic acid/genetic material

**Isolate** - pathogen recovered from a single specimen

**Nonclustered/unique strains** - isolates with genotypes that do not match another isolate in the geographical area and time frame

**Phylogeography** - geographical distribution of molecular strains

**Specimen** - an individual sample derived from a particular patient

**Strain** - genetic variant/genotype of an isolate
1. Synopsis

**Study Design:** Cross sectional survey

**Study Population:** Adolescent and adult patients diagnosed with TB, resident in Masiphumelele in the Fish Hoek Area of Cape Town

**Sample Size:** 509 tuberculosis patients with IS6110-based RFLP and socio-economic data available

**Study Objective**
To explore the socio-economic determinants/risk factors of TB transmission/clustering in a high TB and HIV burdened community.

**Secondary objective 1:** To investigate possible associations between TB transmission/clustering and “social” risk factors.

**Secondary objective 2:** To investigate possible associations between TB transmission/clustering and the “economic” risk factors.
2. Background

Tuberculosis (TB) remains a significant global public health concern, ranking as the world’s second leading cause of death from an infectious agent after the Human Immune Virus (HIV) (1). While established tools and drugs to fight TB have been in existence since the 1940s, issues such as the evolution of multi-drug resistant TB, increasing evidence of strain diversity in *Mycobacterium tuberculosis* (*M.tb*) and the advent of HIV/AIDS have posed new challenges in the fight against TB (2,3). These have necessitated the development of new tools to improve and enhance our understanding of TB epidemiology and subsequently, improve TB control.

2.1 Global Epidemiology of Tuberculosis

While significant progress has been made to halt and reverse TB cases and deaths, the burden of TB remains enormous, particularly in Africa and parts of Europe (1). The Millennium Development Goals (MDGs) set for 2015, which aimed to halve the prevalence and mortality rates compared with those in 1990 have been partially met in some regions of the world (1,5). Despite this progress, the burden of TB on a global scale remains very high, with an estimated 9.0 million incident cases in 2013 (equivalent to 126 cases per 100 000 population) and 1.5 million TB deaths (1.1 million among the HIV-negative and 0.4 million among HIV-positive persons) in the same year (1). According to WHO, the African region has more than 24% of the world’s TB cases and deaths, while approximately 78% of total TB deaths and 73% of TB deaths among HIV-
negative people occurred in the African and the South-East Asia regions combined. In sub-Saharan Africa, 60-80% of all reported TB patients are dually infected with HIV (1,5). On a global scale, the African Region alone accounts for at least 80% of HIV-positive TB cases worldwide (1). This is because in many African countries, the tuberculosis and HIV epidemics are fueling each other (6).

South Africa remains one of the world’s top 6 high burden countries and has the highest TB prevalence rate of over 860/100 000 (ranging from 396-1130) accompanied by an HIV prevalence of more than 60% in incident TB cases (1). The emergence and expansion of drug-resistant TB in the country has further complicated the control of TB with recent studies reporting that at least 20% of HIV-associated TB cases have multi-drug resistant strains (7). Furthermore, at least 13 out of 100 deaths are reported to be from tuberculosis, hence making TB the leading cause of death in South Africa (6). This calls for novel approaches and strategies in understanding and fighting the TB epidemic in this country.

2.2 Molecular Epidemiology of Tuberculosis

In the last decade, the increasing availability of genotyping tools, which are based on analysing the degree of genetic similarity and the distribution of variable elements throughout the genome, have been instrumental in our understanding of the molecular epidemiology of TB, enhancing our ability to investigate the epidemiology and transmission of TB (3,8).
Innovations in molecular epidemiology have helped define the global distribution of *M. tb* lineages; monitor the international spread of highly transmissible or hyper virulent strains; explore the evolutionary features of the bacterium and to discriminate between events of recent TB transmission from those due to reactivation of latent infection (9).

Many TB molecular epidemiology studies aim to distinguish between new infections or exogenous re-infection and the reactivation of previously acquired or latent infection and to identify risk factors in cases of recent transmission (10). The combination of molecular and conventional epidemiological approaches such as contact-tracing is more useful than either method alone for investigating transmission chains and geographic patterns of TB transmission (11,12). This is important in our efforts to understand the transmission of TB and hence, in potentially mitigating the spread of TB in populations.

### 2.3 TB transmission dynamics

In epidemiological investigations, two or more cases occurring within a given time in a given place with the same genotype are considered to be clustered (13) and are assumed to be denote a transmission link. Clustered cases may be linked either through a direct recent transmission event between persons with the same genotype or as part of a common transmission chain. Clustering is thus often used as a proxy for recent or on going transmission occurring within a population (13). In contrast to this, isolates with unique DNA fingerprint patterns, i.e. that do not match or resemble another case within the study population, are generally
considered to reflect reactivation of remotely acquired infection (14). Such cases are considered to be non-clustered or unique (15) and this is often the case in low incidence settings.

Molecular epidemiological studies of TB in our settings and in other southern African countries have shown that recent infection is responsible for as much as two-thirds of the TB cases that are reported (16-19). These high rates of ongoing transmission are likely to be undermining the TB control programs, hence our need to understand the factors contributing to this ongoing transmission.

Studies in different parts of the world have reported varying findings on risk factors for clustering. Factors such as age, immigrant status, HIV infection, homelessness, alcoholism and intravenous drug use have been reported to be risk factors for clustering although there are discrepancies across studies (10,20). Thus, molecular epidemiology has been very useful in identifying groups at high risk such as prisoners and migrant populations, hence enabling public health TB control interventions to be targeted at these populations.

In our study population, moderate associations between biological factors such as HIV-status and clustering have been shown (16). However, gaps still remain in our knowledge and specifically we need to further understand the associations between transmission and socio-economic risk factors for TB. Identification of the socio-economic risk factors would
be important for improving our understanding of M.tb transmission in this and similar settings (3). This knowledge, in turn may contribute to the development of targeted interventions and control measures aimed at interrupting transmission chains (21) and thus controlling TB, particularly in high burden settings.

3. **Problem Statement And Justification**

Despite the advances in our knowledge of TB epidemiology, there are still gaps in our understanding of the risk factors affecting transmission of TB, particularly in high TB incidence settings. Much of the data pertaining to these associations has been reported from developed country settings, with consistency being reported across studies for some risk factors while discrepancies have been reported for others (15,22). Although a number of studies have reported associations between biological factors such as age, HIV-status, ART use, drug resistant TB, and retreatment TB among others, knowledge on the association between TB transmission and socio-economic factors remains incomplete (17,18). This is particularly so in high TB and HIV burdened settings such as our current study setting.

The proposed study, therefore aims to investigate the risk factors of TB transmission that can be identified by combining molecular genotyping data with socio-economic data from the study community. The risk factors to be explored will include economic (e.g educational level; employment status; income levels and housing conditions) and other social risk factors.
4. Study Overview

4.1 Study Hypothesis

We hypothesize that TB transmission (as defined by \textit{M.\,tb} strain clustering) in the study population is associated with many of the socio-economic risk factors.

4.2 Study Objectives

\textbf{Primary Objective:} To explore the associations between transmission (clustering and non-clustering) and the socio-economic factors in a high TB and HIV burdened community.

\textbf{Secondary objectives:} To investigate the associations between TB transmission/clustering and;

- a) Social risk factors
- b) Economic risk factors

5. Methods:

5.1 Study population

The study population is a geographically well-defined peri-urban township in Cape Town, South Africa. In 2006 (when the primary study was commenced) this community had a population of approximately 18 000 people, with nearly 1 in every 4 people in this community being HIV-infected (23). Despite a functional primary care TB clinic and high anti-retroviral therapy (ART) coverage in this community, TB case notifications were as high as 2000/100 000 in this community at the time.
5.2 Study-design

The proposed study will be a cross-sectional study. This study will entail secondary analysis of data obtained from participants who were enrolled into the primary “TB Epidemiology” study conducted in Masiphumelele between 2006 and 2010 (HREC: 321/2005).

5.3 Inclusion and Exclusion criteria

Below is a summary of the inclusion and exclusion criteria used as the basis for enrolment into this study.

Inclusion criteria:

Participants should meet the following criteria to be included into the study:

- Culture/smear confirmed TB disease, notified from 2006 to end 2010
- Resident in Masiphumelele
- Adults ≥ 15 years of age
- Provided consent for the parent study (HREC: 321/2005)
- Completion of the social data questionnaire

Exclusion Criteria:

Participants will be ineligible for the study in the case of the following:

- Missing IS6110 RFLP data
- Strains with <6 bands or low bandwidth strains (strains with <6 copies of IS6110 are not very well differentiated by the RFLP technique)
- Children (≤14 years and younger)
- Unavailability of social data
5.4 Patient Enrolment Procedures

Potentially eligible participants attending the TB clinic were identified by the research assistant and informed about the study. Eligible participants were also identified from the TB register each week and were invited to participate in the study.

5.4 Data Collection

Data for this study was collected as part of a parent study of TB epidemiology performed in Masiphumelele, Cape Town by the Desmond Tutu HIV Centre (DTHC). Following informed consent (Appendix 1), clinical and demographic data were extracted from the TB registers and clinical folders, and TB and socio-economic data were collected using social data questionnaires (Appendix 2) that were translated to the participant’s local language and administered by trained interviewers. The questionnaires captured data on TB history, TB contacts, socio-economic conditions such as occupation, income level, educational level, sexual history in addition to other social and demographic data. The study coordinator was responsible for reviewing and checking questionnaires for completeness and legibility. Study files were stored in a locked cupboard, and signed consent forms were kept separate from questionnaires and laboratory test results. Data entry was done through Datafax. Sputum specimens were collected and genotyping analysis performed on cultured isolates as described in the ensuing section.
5.5 Validity and variability of questionnaires

The questionnaire used in this study was not a validated tool but was adapted from tools that have been previously developed and used by the DTHC for similar studies. Questions were reviewed for appropriateness and language by the study team before being administered to the study participants.

5.6 Laboratory Procedures

HIV Testing and counselling

HIV testing and counselling (and treatment where required) for the participants was conducted according to the national HIV guidelines (24).

TB testing

Sputum specimens were obtained from TB suspects in accordance with the National TB testing, diagnostic and treatment guidelines (25). All sputum specimens were concentrated by centrifugation and then assessed for the presence of acid-fast bacilli (AFB) by fluorescent microscopy. Culture and isoniazid and rifampicin susceptibility testing was performed on all specimens, while second line drug susceptibility testing was done on all multi-drug resistant TB isolates (2).

Genotyping Analysis of TB strains

*M. tb* isolates from patients were analysed by *IS6110*-based RFLP as per parent study protocol (2). Genotyping and molecular analysis of strains and *IS6110*-RFLP typing was performed at the Public Health Research Institute (PHRI) Tuberculosis Centre laboratory, New Jersey. The *IS6110*-based RFLP method, often regarded as the gold-standard in *M. tb*
molecular epidemiology studies (8) is one method which has been instrumental in expanding our understanding of TB molecular epidemiology. This technique utilises DNA probes to visualise restriction fragments containing repetitive DNA sequence complementary to the probe (21). The most widely used probe for RFLP typing detects the insertion sequence, IS6110 which is unique to M. tb and is often found in multiple copies in most strains (21).

The IS6110-based RFLP method is highly discriminatory and because the IS6110 is relatively stable, it can be used to distinguish epidemiologically related from unrelated isolates, making it suitable for epidemiological studies (14,21). Several other techniques have been applied in studying the molecular epidemiology of TB, among them spoligotyping, the variable number tandem repeat (VNTR), mycobacterial interspersed repetitive unit-variable-number tandem repeat analysis (MIRU-VNTR), Single-Nucleotide Polymorphism (SNP) analysis, and more recently whole genome sequencing (WGS). The proposed study will use available secondary data generated from IS6110 RFLP.

Based on the genotyping data, strain families were classified using standard software and tools. Strains with <6 copies of IS6110 were excluded from analysis as they were considered to be poorly differentiated (8).
5.7 Sample Size and power calculations

Out of the total number of 2,237 patients recruited into the parent study, 825 patients were diagnosed with sputum/smear culture positive pulmonary TB. Of these, 772 completed the socio-economic questionnaire. Out of all the sputum positive patients, 631 had RFLP data available. For this study a total of 584 patients had both socio-economic data and RFLP data available. In general approximately 67% of the isolates in the parent study were clustered while at least 31% were unique strains and 2% were shown to be a reactivation of a previously treated TB infection (16). These figures are consistent with other research studies conducted in South Africa, which have reported clustering proportions in the range of 40-70%. Thus for this study, clustering is anticipated to be around 67%, in keeping with the parent study.

6. Data Analysis Plan

Data analysis will be performed using the statistical program STATA (Stata Corp. LP, College Station, TX, United States of America). Basic descriptive statistics will be employed to explore baseline demographic, clinical and socio-economic data and for outcome variables. Bivariate comparisons (such as Spearman or Pearson’ correlations, t-tests, Wilcoxon’s rank sum test, chi-squared and Fisher’s exact tests) will be employed as appropriate to explore the data and identify possible bivariate association.
6.1 Data Management

Completed questionnaires from the parent study were stored in a secure locked cupboard accessible to the study investigators. Data from these questionnaires was then kept in a secure and password protected electronic database, accessible only to researchers in the study team. For this study, the data to be received will be anonymized with no direct link to patients except for an allocated study and TB register number. Table 1 below shows a summary of the variables that will be explored, as well as the scales and coding used.
6.2 Measurements and variables

Table 1. Summary of Measurements and variables from questionnaire used in this study

<table>
<thead>
<tr>
<th>Operational variable</th>
<th>Scale</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Demographic Information</strong></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>Numerical- continuous, Binary, ordinal</td>
</tr>
<tr>
<td>Gender</td>
<td>Categorical- binary</td>
</tr>
<tr>
<td>HIV Status</td>
<td>Categorical- binary</td>
</tr>
<tr>
<td>Home language</td>
<td>Categorical- nominal</td>
</tr>
<tr>
<td>Highest level of education</td>
<td>Numerical- ordinal</td>
</tr>
<tr>
<td>Length of stay (time) at present address</td>
<td>Numerical- continuous</td>
</tr>
<tr>
<td>Length of stay in Masiphumelele</td>
<td>Numerical- continuous</td>
</tr>
<tr>
<td>Number of people living in the house</td>
<td>Numerical- continuous</td>
</tr>
<tr>
<td>Time spent outside Masiphumelele</td>
<td>Numerical- continuous</td>
</tr>
<tr>
<td>Number of people sleeping in the same room</td>
<td>Numerical- continuous</td>
</tr>
<tr>
<td>Employment status</td>
<td>Categorical- binary</td>
</tr>
<tr>
<td>Income level</td>
<td>Numerical-continuous, ordinal</td>
</tr>
<tr>
<td><strong>TB Contacts Information</strong></td>
<td></td>
</tr>
<tr>
<td>Anyone known to be currently on TB treatment</td>
<td>Categorical- binary</td>
</tr>
<tr>
<td>Number of people known who are currently on TB treatment</td>
<td>Numerical- continuous</td>
</tr>
<tr>
<td>Anyone known who was on TB treatment in the last 12 months</td>
<td>Categorical- binary</td>
</tr>
<tr>
<td>Number of people known who were on TB treatment in the past 12 months</td>
<td>Numerical- continuous</td>
</tr>
<tr>
<td>Anyone known to be on TB treatment ever</td>
<td>Categorical- binary</td>
</tr>
<tr>
<td>Number of people known who were ever on TB treatment</td>
<td>Numerical- continuous</td>
</tr>
<tr>
<td><strong>Housing Information</strong></td>
<td></td>
</tr>
<tr>
<td>Number of houses on a plot</td>
<td>Numerical- continuous</td>
</tr>
<tr>
<td><strong>Type of house</strong></td>
<td>Categorical-binary</td>
</tr>
<tr>
<td>-------------------</td>
<td>--------------------</td>
</tr>
<tr>
<td><strong>Number of bedrooms</strong></td>
<td>Numerical-continous</td>
</tr>
<tr>
<td><strong>Presence of water tap in house</strong></td>
<td>Categorical-binary</td>
</tr>
<tr>
<td><strong>Communal water tap</strong></td>
<td>Categorical-binary</td>
</tr>
<tr>
<td><strong>Availability of electricity in house</strong></td>
<td>Categorical-binary</td>
</tr>
<tr>
<td><strong>Availability of toilet in house</strong></td>
<td>Categorical-binary</td>
</tr>
<tr>
<td><strong>Number of adults in house</strong></td>
<td>Numerical-continous</td>
</tr>
<tr>
<td><strong>TB Risk factor information</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Alcohol use</strong></td>
<td>Categorical-binary</td>
</tr>
<tr>
<td><strong>Use shebeen/tavern in the last 12 months</strong></td>
<td>Categorical-binary</td>
</tr>
<tr>
<td><strong>Smoking status</strong></td>
<td>Categorical-binary</td>
</tr>
<tr>
<td><strong>History of recreational drug use</strong></td>
<td>Categorical-binary</td>
</tr>
<tr>
<td><strong>History of working in a mine</strong></td>
<td>Categorical-binary</td>
</tr>
<tr>
<td><strong>Duration of working in mine</strong></td>
<td>Numerical-continous</td>
</tr>
<tr>
<td><strong>History of being in prison</strong></td>
<td>Categorical-binary</td>
</tr>
<tr>
<td><strong>Recent death in the household in the last 12 months</strong></td>
<td>Categorical-binary</td>
</tr>
<tr>
<td><strong>Cause of death known</strong></td>
<td>Categorical-binary</td>
</tr>
<tr>
<td><strong>Number of sexual partners in the last 12 months</strong></td>
<td>Numerical-continous</td>
</tr>
<tr>
<td><strong>Occupation and other social activities</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Employment in the last 12 months</strong></td>
<td>Categorical-binary</td>
</tr>
<tr>
<td><strong>School attendance in the last 12 months</strong></td>
<td>Categorical-binary</td>
</tr>
<tr>
<td><strong>Regular Church attendance</strong></td>
<td>Categorical-binary</td>
</tr>
<tr>
<td><strong>Clinic visit in the last 12 months</strong></td>
<td>Categorical-nominal</td>
</tr>
<tr>
<td><strong>Travel to the Eastern Cape in the last 12 months</strong></td>
<td>Categorical-binary</td>
</tr>
<tr>
<td><strong>Regular use of taxi</strong></td>
<td>Categorical-binary</td>
</tr>
<tr>
<td><strong>Frequency of taxi use in a month</strong></td>
<td>Numerical-continous</td>
</tr>
<tr>
<td><strong>Social network groups in Masiphumelele</strong></td>
<td>Categorical-binary</td>
</tr>
</tbody>
</table>
7. Ethical Considerations

The proposed study will adhere to the principles laid down in the Helsinki Declaration of 2013 and the South African Medical Research Council guidelines on Ethics for Medical Research.

This protocol will be submitted to the University of Cape Town Human Research Ethics Committee (HREC) for ethical review.

During recruitment for the parent study, eligible participants were informed (in their home language) about the study objectives, and procedures as well as the associated the risks and benefits of participating in the study, by trained study personnel before enrolment. Participants were also informed about their right to refuse to participate or to withdraw consent at any point during the study. Thereafter, participants were asked to provide informed consent by signing an HREC-approved informed consent form. For minors participating in the study, they were asked to give assent while an adult guardian was asked to provide written informed consent before enrolling them into the study. The analysis proposed in this protocol falls within the informed consent that was obtained for the parent study.

**Risks and benefits of participating in the study**

Participants are TB patients who have been diagnosed and managed in a health care facility and were offered standard TB treatment as per the
national TB guidelines (26). Participants were thus not exposed to additional harmful activities or drugs apart from routine clinical procedures. However, participation in the study may be associated with a small risk of loss of confidentiality. All reasonable efforts were made and will be made to preserve confidentiality of participants’ information. Overall, this study could be deemed to be low risk research.

There are no direct benefits for participants in this study. However, this study may provide knowledge that could aid future TB control measures, in the community.

8. Communication

Stakeholders and Reporting

The principle stakeholders in this study include the City of Cape Town’s Department of Health, clinic staff as well as the Masiphumelele community.

Following study completion, findings will be disseminated back to the various stakeholders. Efforts will be made to ensure that research findings will be disseminated back to the community in the form of public information sessions and posters to be posted in public places. For other stakeholders such as the Department of Health and Provincial Health officials, findings will be presented in the form of written reports. Furthermore, findings will be submitted for publication in an appropriate
international peer-reviewed journal for dissemination to the broader research community and should ideally be available on Pubmed.

9. Logistics

Study Time Lines

10. References


10. Murray M, Nardell E. Molecular epidemiology of tuberculosis:


PART B: LITERATURE REVIEW
1. **Molecular Epidemiology of Tuberculosis:**

**Overview**

This review aims to discuss the general epidemiology and molecular epidemiology aspects of tuberculosis, first on a global scale and then the developing country and South African context. The main methods commonly used in molecular epidemiological studies of tuberculosis (TB) will be discussed briefly followed by an exploration of the key socio-economic risk factors for TB transmission. Molecular epidemiology of TB from the High Income Countries (HIC) is briefly discussed followed by discussion of molecular epidemiology of TB in the Low/medium income countries (LMIC), including the African and the South African context.

2. **Search Strategy and Selection Criteria**

Data for this review were identified through searches on PubMed and looking at references from key relevant articles. Search terms and MESH terms used included “Mycobacterium tuberculosis”, “tuberculosis”, “TB”, “molecular epidemiology”, “Mycobacterium tuberculosis” AND “molecular epidemiology”. Filters that were used included “humans”, “Africa”, “South Africa”, “developing country”, “risk factors”, “socio-economic factors”, and “transmission.” A total of 61 studies were identified that met the search criteria.
3. Global Epidemiology of Tuberculosis

Tuberculosis is a disease of poverty that thrives where social and economic determinants of ill health are most prevalent [1]. Following neglect of the disease during the 1980s, tuberculosis made resurgence in the 1990s, predominantly fueled by the HIV epidemic. This increase in the burden of TB can partly be attributed to the fact that HIV significantly increases the risk of reactivating latent *Mycobacterium tuberculosis* (*M.tb*) infection and recent *M.tb* infection to active disease (Bucher, 1999).

While the lifetime risk of developing TB ranges from 10-20% in HIV-uninfected persons infected with *M.tb*, in persons co-infected with *M.tb* and HIV, the annual risk can exceed 10% [2]. Efforts to control TB were thus revived in the 1990s with the advent of HIV/AIDS and the recognition of TB as a major global health public problem and the WHO declaring TB a global emergency [3].

The WHO reports that an estimated 9.0 million people developed TB and 1.5 million died from the disease in 2013, 360 000 of whom were HIV-positive [4]. While a large proportion of the TB cases were from South-East Asia and Western Pacific Regions (more than 56%), at least a quarter of these were from the African Region, which also has the highest rates of TB cases and deaths [4]. The achievement of global targets, widely known as Millennium Development Goals (MDGs) for reductions in the epidemiological burden of TB has been a major focus of national and international efforts in TB control in recent times [1]. As part of the MDGs, the Stop TB Partnership aimed to halve the TB prevalence and mortality
rates by 2015 as compared to the 1990 levels as well as eliminating TB by 2050 by reaching a global incidence of disease of less than one TB case per million[4].

Globally, the estimated rates of TB prevalence and mortality have been on the decline such that some of the regions have met or are on course to achieving the 2015 MDG targets [3,4]. Two out of the six WHO regions have achieved all three 2015 targets for reductions in TB disease. However in the other regions including the African, Eastern Mediterranean and European regions, the rate of decrease in prevalence and mortality rates is not sufficient for these targets to be met[4].

The variable impact of TB control measures, as seen by the achievement/non-achievement of global targets therefore necessitates not only better tools and interventions, but also better insights into the natural history and epidemiology of TB [5]. Such control strategies are also likely to differ in effectiveness between high and low incident settings. In order to reach long-term epidemiological targets for global TB control, preventive interventions targeting social risk factors are required, in addition to advances in biomedical interventions [6]. The re-invigorated efforts in TB research in recent years have led to several scientific advances and the discovery of new tools to fight and control TB. One of the notable successes in the past decade has been the development of a new discipline, molecular epidemiology [7].
4. Molecular epidemiology of TB

Molecular epidemiology, a blend of molecular biology and epidemiology, has added a further dimension to the conventional epidemiology of tuberculosis and has enhanced our understanding of how _M. tb_ continues to be successfully transmitted within populations [7,8]. To date, molecular epidemiologic studies of tuberculosis have been useful in addressing some of the short- and long-term epidemiologic questions adding much-needed accuracy and precision in describing and analyzing the transmission dynamics and the spread of _M. tb_ during outbreaks [9]. Molecular epidemiological studies have thus contributed useful insights by answering questions such as: the proportion of cases attributable to recent transmission; risk factors for recent transmission; the occurrence of multiple _M. tb_ infections in an individual, and the proportion of recurrent TB cases attributable to re-infection [6,10,11]. Additional successful application of molecular epidemiology to date include the identification of _M. tb_ lineages and their association with geographical origin [10,12].

Molecular epidemiology has also been applied to quantify the importance of institutional transmission and laboratory cross-contamination, hence helping to focus contact investigations as well as to investigate epidemics [10]. Furthermore, molecular epidemiologic approaches have enabled the impact of drug resistance on the transmission and pathogenesis of _M. tb_ to be assessed [13]. Increasing evidence has now been generated to show that specific strains of _M. tb_ belonging to distinct phylogenetic clusters/lineages may differ in virulence, pathogenesis, and epidemiologic
characteristics, findings which, in part, can be attributed to molecular epidemiological approaches [14].

When molecular methods are used in conjunction with traditional epidemiology methods such as active case finding and contact tracing, their utility for TB control can be fully realized [9]. The combination of molecular epidemiology and other disciplines is thus likely to continue to increase our understanding of tuberculosis transmission and pathogenesis. This enhanced understanding has the potential to result in new and more targeted strategies against TB which can potentially impact control of this important human pathogen [12,13]. Identifying tuberculosis transmission patterns using molecular biology affords the possibility of identification of transmission profile, detecting population groups for whom preferential measures can be taken, risk factors, which can inform appropriate TB prevention and appropriate control strategies in such settings [6,8,15-17].

4.1 Molecular Typing Techniques

Several new DNA typing methods have been applied in studying the molecular epidemiology of TB. Restriction fragment length polymorphism (RFLP), is one of the commonly used molecular typing techniques and is sometimes regarded as a gold standard for molecular epidemiology of the *M.tb* complex [9]. RFLP typing utilizes DNA probes to visualize the restriction fragments that contain the repetitive DNA sequences. The insertion sequence *IS6110*, which is unique to the *M.tb* complex and is found in multiple copies in most strains is the most widely used probe for
RFLP typing [18-20]. The *IS6110* RFLP typing method is highly discriminatory, with the *IS6110* being stable enough to distinguish epidemiologically related from unrelated isolates [20]. This makes it a very suitable and widely applied method in epidemiological studies of recent transmission of TB [21].

In a proportion of *M. tb* isolates which may contain fewer than 6 copies of *IS6110*, the RFLP method, however is becomes less discriminatory [20]. Hence, alternative DNA typing methods can be used to complement *IS6110* RFLP typing [20]. One such technique is PCR-based spoligotyping which has discriminatory ability and reproducibility that is almost as good as that of RFLP typing [21]. The method relies on amplification of DNA by PCR, and so can be performed with a smaller number of bacteria or genetic material [21]. Although it has a lower discriminatory power in comparison to *IS6110* RFLP typing, spoligotyping is frequently used for studying the phylogeography of the *M. tb* complex [22] and offers several advantages over RFLP, which include simplicity, cost effectiveness and speed [20,22] [23].

Yet another commonly used molecular approach is the variable number tandem repeats method (VNTR) and variations thereof. This method is based on the detection of polymorphism in the number of tandem repeats at seven loci in the *M. tb* genome [20]. Based on the size of PCR products, the exact number of tandem repeats is identified at each locus in each strain [24]. The MIRU-VNTR is slightly more discriminatory than
spoligotyping [25] and is currently being increasingly adopted worldwide [26]. In samples with low numbers of IS6110 copies, the MIRU VNTR techniques using 12-15 loci has been shown to have greater discriminatory power in comparison to RFLP [27,28].

Additional methods that have been employed in studying the molecular epidemiology of TB include Pulsed-Field Gel Electrophoresis (PFGE); Single nucleotide Polymorphism (SNP) typing; genomic deletion and mapping analysis; identification of strain-specific markers for rapid diagnosis (insertion site mapping and insertion site typing), and more recently, whole genome sequencing [9,22]. The use of whole genome sequencing, with its high degree of discrimination, circumvents some of the shortfalls of the older molecular techniques such as excluding mixed infections and knowing how much change to allow when comparing specimens from different episodes of disease such relapses and re-infections [29,30]. In the future, whole genome sequencing is likely to replace all other genotyping techniques, particularly with anticipated decreasing costs of this technology as availability increases [30,31].

The molecular epidemiologic approach chosen is often dependent on the research question (e.g. recent transmission vs. reactivation); the observed rate of polymorphism or stability of marker; and the genetic diversity of strains in the study population [9]. The rate of change of a biomarker should be rapid enough to distinguish non-epidemiologically related strains but slow enough to reliably link related cases. These
considerations, coupled with general background TB prevalence, need be taken into consideration when choosing molecular epidemiologic methods or in evaluating data [9]. In this study, the molecular data to be analyzed were generated by RFLP typing.

**4.2. Socio-economic risk factors for TB**

While the current global TB control programs mainly focus on reducing transmission through early case detection [6,32], socio-economic determinants should not be overlooked if the fight against TB is to be effective. Historically, TB has predominantly been a disease of poverty, associated with low socio-economic status and poor living conditions [33]. An enhanced focus on prevention and limiting the impact of the TB epidemic requires approaches focusing on mitigating common risk factors, including HIV/AIDS, smoking, malnutrition, alcoholism, diabetes, crowded living conditions, and indoor air pollution [34]. Additional factors identified as risk factors for TB transmission from studies done in the Low/medium income countries (LMIC) include homelessness, immigrant status or being foreign born, male sex and younger age [35-38]. However, there is no overall agreement with apparent inconsistencies often existing in the reported associations between socio-economic factors and TB, particularly in studies from LMIC. One study, which investigated the association between socio-economic status and TB and reported seemingly inconsistent results, was the study by Odone et al in Malawi [39]. This study showed an increased risk of TB in people working in the cash economy (potentially a higher socio-economic) compared to those in a subsistence (or lower socio-economic status) economy. The
authors attributed this unexpected association to potential mediating factors such as overcrowding and socialization as contributors to the observed increased TB transmission in those people from a higher socio-economic strata [39].

Even though the link between socio-economic factors may differ between countries (i.e. high vs. low incidence countries), populations (generalised epidemic vs. high risk groups or sporadic epidemics) and between studies, there is general agreement that marginalised populations are more at risk from on-going/recent TB transmission. Such associations have been shown in several studies from the HIC, but have not been clearly elucidated in the developing country setting as can be noted from the Malawian study by Odone et al. The identification of such socio-economic risk factors and their association with TB transmission and strain distribution is thus, likely to be critical for the control of TB in developing country settings including South Africa.

4.3. Public Health Applications of Molecular Epidemiology

In epidemiological investigations, a cluster is defined as two or more TB cases with the same genotype in a given geographical area over a time period [9]. In molecular epidemiology of tuberculosis, molecular genotyping is used to estimate the fraction of incident cases attributable to recent transmission of \( M.\text{tb} \) rather than reactivation disease and to identify risk factors for recent transmission. This approach is based on the premise that tuberculosis cases that share a DNA fingerprint are epidemiologically related while cases in which fingerprints are unique
result from remotely acquired or latent infection that has reactivated [1,40]. In other words, strain clustering is used as a proxy for recent transmission [18]. Previously, estimates of the contribution of re-infection relied on mathematical models, which identified the parameters most consistent with epidemiological data [20,41]. Molecular epidemiological methods have thus been instrumental in helping us to understand the substantial contribution of recent transmission to disease burden, particularly in LMIC.

4.4. Molecular epidemiology of TB in High Income countries

In many of the high-income countries (HIC), the tuberculosis epidemic is fuelled by immigration, as most of the TB cases are found in foreign migrants. In San Francisco, for example, 70% of the tuberculosis cases occur among foreign-born persons, mainly from China, the Philippines, and Mexico [42]. Even among the immigrants, the clinical, epidemiological, and microbiological characteristics may differ significantly among immigrants of different nationalities [1,35,42]. This may be a reflection of differences in the characteristics and risk factors in their population of origin and such information can be used to inform more targeted TB control policies and measures. Another study which used spoligotyping and MIRU-12 typing to investigate factors associated with genotype clustering of M.tbc isolates in an ethnically diverse region of southern California showed that TB patients with clustered isolates were more likely to be young, homeless and male [36,43]. Homelessness is well known to be associated with TB case clustering in regions of low TB incidence globally, but not in high incidence countries [44]. Yet another
American study conducted in New York City by Driver et al corroborated these research findings and reported that younger age, homelessness, substance abuse and presence of TB symptoms were independently associated with greater odds of clustering [37]. However, contrary to the other studies, which reported associations between clustering and foreign immigrant status, this study reported birth in the US to be a risk factor for TB clustering.

Similarly, TB molecular epidemiology studies in Western Europe are in close agreement with the American studies. In a study aimed to assess the degree of recent transmission of M.tb in London and sub-populations of the community with high rates of disease, M.tb isolates from persons in Greater London diagnosed with culture positive TB were genotyped [45]. The estimated rate of recent transmission was 14.4% with the bulk of disease largely caused by reactivation or importation of infection by recent immigrants. Young age, black Caribbean ethnic group, UK-birth, alcohol dependence and streptomycin resistance were independently associated with an increased risk of clustering [45]. Another British study showed that clustering was more common in those with pulmonary smear positive disease and in patients living in south London with a trend towards clustering being more common in those with an unknown HIV status [46].

In another European study, genotyping was used to analyze TB epidemiology and to identify risk factors related to clustering similar in
immigrant patients living in two major northern Italian urban areas [35]. In this study apart from country of origin, all other factors such as age, gender, HIV-seropositivity and drug resistance were not independent predictors of TB clustering. Similarly, another Italian study by Bandera and colleagues reported a fourfold-increased risk for reinfection in immigrant patients compared to Italian subjects [11]. Gender, age, HIV infection and drug resistance were not significantly linked to TB clustering. However, the risk factors associated with strain clustering in this study were country of origin and city of residence. Similar to studies done in San Francisco, variations in TB transmission were observed among immigrants from different countries and even within national groups, where living conditions have been found to exert a profound impact [35]. These results have been unequivocal in demonstrating the importance of improving social integration and targeted prevention strategies among immigrant populations [7]. Thus, most studies done in HIC such as the US and Europe have reported consistent associations between TB clustering and certain socio-economic risk factors [10,35,36,42,45-47].
Table 1: Summary of selected key primary studies from HIC and the social risk factors that were identified to be linked to TB transmission.

<table>
<thead>
<tr>
<th>Author, year of publication</th>
<th>Period of sampling/study</th>
<th>Country of Study</th>
<th>City</th>
<th>Study Design</th>
<th>Sample Size</th>
<th>Study Population</th>
<th>TB Transmission social risk factors</th>
<th>Risk factors not associated with TB transmission</th>
</tr>
</thead>
<tbody>
<tr>
<td>Franzetti et al 2010</td>
<td>1993-2000</td>
<td>Italy</td>
<td>Milan and Bresica</td>
<td>Cross-sectional</td>
<td>1,999</td>
<td>75% Italian born, 25% foreign born</td>
<td>Immigrant populations</td>
<td>Gender, age, HIV status, drug resistance</td>
</tr>
<tr>
<td>Rodwel et al 2012</td>
<td>2005-2008</td>
<td>USA</td>
<td>San Diego</td>
<td>Cross-sectional</td>
<td>1,164</td>
<td>Foreign and American born</td>
<td>Homelessness, male sex, younger age, Phillipine (foreign) born</td>
<td>Race (ethnicity), injection drug use, excessive alcohol consumption, homelessness, history of correctional facility, occupation, HIV status, previous</td>
</tr>
<tr>
<td>Study</td>
<td>Start of collections</td>
<td>Country</td>
<td>City</td>
<td>Study Design</td>
<td>Sample Size</td>
<td>Demographics</td>
<td>Associated Risk Factors</td>
<td></td>
</tr>
<tr>
<td>------------------</td>
<td>----------------------</td>
<td>---------</td>
<td>------</td>
<td>------------------</td>
<td>-------------</td>
<td>--------------</td>
<td>-------------------------</td>
<td></td>
</tr>
<tr>
<td>Fok et al 2008</td>
<td>Start of collections Up to 2006</td>
<td>17 different countries</td>
<td>Systematic review</td>
<td>36 studies</td>
<td>Male sex, injection drug use, alcohol abuse, homelessness</td>
<td>HIV status, drug resistance, previous imprisonment, smoking, employment status</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hayward et al 2002</td>
<td>1993</td>
<td>England</td>
<td>London</td>
<td>Surveillance</td>
<td>569</td>
<td>27% UK-born, 73% Other</td>
<td>UK-born, HIV-status, white race, gender, history of...</td>
<td></td>
</tr>
<tr>
<td>Study</td>
<td>Years</td>
<td>Country</td>
<td>Study Type</td>
<td>Sample Size</td>
<td>Gender</td>
<td>Age</td>
<td>HIV Status</td>
<td>Drug-resistant TB</td>
</tr>
<tr>
<td>------------------------</td>
<td>----------------</td>
<td>-------------</td>
<td>---------------</td>
<td>-------------</td>
<td>--------</td>
<td>-----</td>
<td>-------------</td>
<td>-------------------</td>
</tr>
<tr>
<td>Sloot et al 2013</td>
<td>2003-2008</td>
<td>Netherlands</td>
<td>Surveillance</td>
<td>3776</td>
<td>Mixed</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bandera et al 2001</td>
<td>4 years</td>
<td>Italy</td>
<td>Cross-sectional</td>
<td>2452</td>
<td>86% Italian born</td>
<td></td>
<td></td>
<td>Foreign Immigrants</td>
</tr>
<tr>
<td>Love</td>
<td>1998</td>
<td>England</td>
<td>Surveillance</td>
<td>3713</td>
<td>28% UK Born, 72% Other</td>
<td></td>
<td></td>
<td>Homelessness, alcoholism</td>
</tr>
<tr>
<td>Franzetti et al 2010</td>
<td>1993-2000</td>
<td>Italy</td>
<td>Cross-sectional</td>
<td>1999</td>
<td>75% Italian born, 25% foreign born</td>
<td></td>
<td></td>
<td>Immigrant populations</td>
</tr>
<tr>
<td>Study</td>
<td>Time Period</td>
<td>Location</td>
<td>Study Design</td>
<td>Sample Size</td>
<td>Key Findings</td>
<td>Relevant Factors</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------------------</td>
<td>-------------</td>
<td>----------</td>
<td>--------------</td>
<td>-------------</td>
<td>-------------------------------------------------------------------------------</td>
<td>-------------------------------------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rodwel et al. 2012</td>
<td>2005-2008</td>
<td>USA</td>
<td>Cross-sectional</td>
<td>1,164</td>
<td>Foreign and American born, Homelessness, male sex, younger age, Philippine (foreign) born</td>
<td>Race (ethnicity), injection drug use, excessive alcohol consumption, homelessness, history of correctional facility, occupation, HIV status, previous TB</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
4.5. Molecular Epidemiology of TB in LMIC

In many of the sub-Saharan African countries, despite high rates of ongoing TB transmission, our knowledge of the molecular epidemiology of TB and the socio-economic risk factors associated with clustering remains limited. In this section the molecular epidemiology of TB in LMIC, parts of Africa and in South Africa will be discussed based on the published literature available. A total of 22 studies from LMIC were found that met the criteria. Out of these, 12 studies, which met the criteria of being primary molecular epidemiology studies, were reviewed (Table 2).

Mechanisms and risk factors for tuberculosis transmission in a rural area with high TB prevalence in south India have been studied [8,48,49]. The findings from these studies showed that the majority of the tuberculosis cases in south India were due to reactivation.

A number of TB molecular studies have been performed in Africa. A study aimed at describing the molecular epidemiology of tuberculosis (TB) in Equatorial Guinea showed that at least 61.6% of the TB was from ongoing transmission [50]. Apart from the finding that isoniazid resistance was more frequent among patients with clustered strains, no other epidemiological links were established between clustering and any of the other demographic factors including age, gender, district of residence, level of overcrowding and alcohol consumption among other factors investigated [50]. Another Ugandan study conducted in a high TB incidence peri-urban setting of Kampala studied the relationship between
the risk of belonging to a TB cluster and the HIV sero-status of the individuals [51]. The study showed evidence of a high prevalence of recent transmission with a high average cluster size. However, infection with an isolate with a genotype found to be part of a cluster was not associated with any of the demographic factors such as age and sex or clinical characteristics, including HIV status [51,52]. Although small, yet another molecular epidemiology study in Kumasi, Ghana, identified several important factors that potentially promote disease transmission in a community. Among the reported risk factors were overcrowding, primary isoniazid resistance, all of which indicated poor TB control in this setting [53].

In an Ethiopian study combining socio-demographic, clinical and genotyping data, Garedew et al showed the presence of several clusters and new strains of *M.tb* circulating in the study area, suggestive of recent and ongoing TB transmission. However, no significant association were noted between transmission and any of the risk factors such as age, gender, rural or urban dwelling, or occupation among other demographic factors [54]. In another population based study aimed at assessing TB transmission risk factors in Botswana, Lockman et al showed that a history of imprisonment was the only statistically significant risk factor associated with TB clustering [55]. Other factors such as previous mine work, previous hospital work, previous TB infection, history of contact with a TB patient, age or gender did not show any significant association with clustering. While the lack of association may represent a valid finding,
the authors also noted that potential missing data and the mobility of the population might also have resulted in true associations between social risk factors and molecular data being missed [55].

4.6. Molecular Epidemiology of TB in South Africa

In South Africa, HIV-associated tuberculosis remains a major health problem among certain high-risk populations such as the gold-mining workforce. A recent study in gold miners by Mathema et al aimed at estimating levels of transmission, and examining risk factors for clustering reported a lack of individual-level risk factors for clustering such as age, sex, years in mining industry, or country of origin [56]. Despite the high \textit{M. tb} genetic diversity, the only factors to achieve borderline evidence of association with TB transmission were being non-South African, self-reported HIV infection and \textit{M. tb} genetic diversity [56]. The authors hypothesized that the high \textit{M. tb} genetic diversity and lack of risk factors for clustering may be indicative of a universal risk for disease among gold miners and likely mixing with non-mining populations.

TB molecular epidemiology studies in Western Cape have reported high rates of clustering which suggest ongoing transmission in both HIV-positive and HIV-negative individuals [57,58]. While an association between W-Beijing and HIV infection has been shown, no other associations with clustering have been identified. Another study from the Western Cape, although showing up to 73% clustering, could not show any clear links between transmission and risk factors such as age, gender or HIV-status [59]. Yet another study from Cape Town by Verver
et al reported smear positive re-treatment TB to be associated with clustering [60].

Thus, although we are well aware that TB transmission remains unchecked in many of the high burden countries such as South Africa, the links between transmission and socio-economic risk factors remain unclear and elusive. Table 2 below is a summary of selected key studies from LMIC and the main social risk factors that were identified to be linked to TB transmission.
Table 2: Summary of selected TB Molecular epidemiology studies from LMIC and the social risk factors that were identified to be linked to TB transmission.

<table>
<thead>
<tr>
<th>Author, year of publication</th>
<th>Period of sampling/study</th>
<th>Country of Study</th>
<th>City</th>
<th>Study Design</th>
<th>Sample Size</th>
<th>Study Population</th>
<th>TB Transmission social risk factors</th>
<th>Risk factors not associated with TB transmission</th>
</tr>
</thead>
<tbody>
<tr>
<td>Narayanan et al, 2002</td>
<td>1999-2000</td>
<td>India</td>
<td>Tiruvalur District</td>
<td>Surveillance</td>
<td>437</td>
<td>Indian</td>
<td>*Hospitalisation in a sanatorium, relapse TB infection</td>
<td>MDR, alcoholism, family size, close contact with a known TB patient</td>
</tr>
<tr>
<td>Muwonge et al 2013</td>
<td>2011</td>
<td>Uganda</td>
<td>Mubende District</td>
<td>Cross-sectional</td>
<td>344</td>
<td>Undefined</td>
<td>Previous TB, cigarette smoking, overcrowding</td>
<td>None reported</td>
</tr>
<tr>
<td>Study</td>
<td>Year</td>
<td>Country</td>
<td>Site</td>
<td>Study Type</td>
<td>N</td>
<td>Participant Characteristics</td>
<td>Outcome Characteristics</td>
<td></td>
</tr>
<tr>
<td>------------------------</td>
<td>------------</td>
<td>--------------</td>
<td>--------------------------</td>
<td>-------------</td>
<td>----</td>
<td>------------------------------------------------------------------------------------------------</td>
<td>----------------------------</td>
<td></td>
</tr>
<tr>
<td>Asiimwe et al 2009</td>
<td>2006</td>
<td>Uganda</td>
<td>Kampala</td>
<td>Survey</td>
<td>183</td>
<td>Male sex</td>
<td>HIV status, age, sex, drug resistance</td>
<td></td>
</tr>
<tr>
<td>Lawn et al 2001</td>
<td>2000</td>
<td>Ghana</td>
<td>Kumasi</td>
<td>Cross-sectional</td>
<td>30</td>
<td>Overcrowding</td>
<td>None noted</td>
<td></td>
</tr>
<tr>
<td>Garedew et al 2013</td>
<td>2010-2011</td>
<td>Ethiopia</td>
<td>Debre Birhan</td>
<td>Survey</td>
<td>99</td>
<td>Inconclusive</td>
<td>None noted</td>
<td></td>
</tr>
<tr>
<td>Mathema et al 2015</td>
<td>2006-2010</td>
<td>South Africa</td>
<td>Gauteng North West and Free State Provinces</td>
<td>Cross-sectional</td>
<td>1,602</td>
<td>* Non-South African born</td>
<td>Years in mine industry, previous TB, country of origin, type of residence, previous</td>
<td></td>
</tr>
<tr>
<td>Study</td>
<td>Year</td>
<td>Country</td>
<td>Region</td>
<td>Study Design</td>
<td>Cases</td>
<td>Sex, Age, HIV status, ART use</td>
<td>Hospitalisation Factors</td>
<td></td>
</tr>
<tr>
<td>-------------------</td>
<td>---------------</td>
<td>---------------</td>
<td>--------------</td>
<td>--------------</td>
<td>-------</td>
<td>-------------------------------</td>
<td>-------------------------</td>
<td></td>
</tr>
<tr>
<td>Middelkoop et al 2009</td>
<td>2001-2005</td>
<td>South Africa</td>
<td>Cape Town</td>
<td>Cross-sectional</td>
<td>467</td>
<td>Peri-urban population</td>
<td>No social risk factors identified</td>
<td></td>
</tr>
<tr>
<td>Middelkoop 2015</td>
<td>2001-2010</td>
<td>South Africa</td>
<td>Cape Town</td>
<td>Cross-sectional</td>
<td>710</td>
<td>Peri-urban population</td>
<td>No social risk factors identified</td>
<td></td>
</tr>
</tbody>
</table>

* Borderline significance or trend to significance
5. Summary

While studies from the HIC have reported data showing links between clustering and certain risk factors such as drug use, incarceration, homelessness among others, such clear associations have not been found in many of the studies from LMIC. Several studies from LMIC have reported social factors such as homelessness, alcohol and drug abuse, migrant status, and to some extent male sex and younger age to be factors that are associated with TB transmission. On the other hand, only a handful of the African studies have identified social risk factors that could be linked to TB transmission.

One reason of the lack or inconsistencies in associations may be study design flaws such as under-powering (most of the developing country studies were small and had sample sizes below 1000). This may mean that real associations could have been missed due to the small sample sizes. Another possible explanation maybe the differences in the study populations and nature of the TB epidemics between developed and LMIC. While many of the LMIC discussed have a generalized epidemic, this means that the populations studied are often homogeneous. In most cases, the developed country studies compare foreign born vs. indigenous population groups while developing country studies are more often focused only on indigenous populations which are often homogenous, and so equally exposed and equally at risk for the different factors.
The proposed study aims to explore the associations between different genotypically defined *M.tb* clusters and the social determinants and risk factors of tuberculosis transmission in a high TB and HIV burdened community. Based on previous studies from developing country settings, factors of interest include overcrowding and housing conditions and gender among others. The findings from this research may be of critical importance in devising specific and targeted interventions to control the extensive ongoing TB transmission in our study setting and may also inform interventions in similar high TB-burdened developing country settings.


7. Burgos MV, Pym AS. Molecular epidemiology of tuberculosis. *Eur*


Debre Birhan Hospital, Ethiopia. *Int J Tuberc Lung Dis.* 17(8), 1076–1081 (2013).


PART C: MANUSCRIPT
Title: Investigating the association between TB transmission and socio-economic risk factors in a high TB and HIV burdened community in Cape Town, South Africa.

Running Head: Socio-economic factors and TB transmission

Rebecca Tadokera, PhD¹, Linda-Gail Bekker, MBCHB, DTMH, DCH, FCP(SA), PhD¹,², Keren Middelkoop, MBCHB, MPH, PhD¹,²

¹ Desmond Tutu HIV Centre, Institute of Infectious Disease and Molecular Medicine, University of Cape Town.
² Department of Medicine, University of Cape Town.

Key Words
Tuberculosis, TB transmission, clustering, low socio-economic status, RFLP, molecular epidemiology
Abstract

Background
Several studies have assessed the associations between biological factors and TB transmission. However, our understanding of the associations between TB transmission and socio-economic factors remains incomplete. We explored associations between TB transmission and socio-economic risk factors in a high TB and HIV burdened community.

Methods
A cross-sectional molecular epidemiology study was conducted among adult TB patients. Demographic and clinical data were extracted from TB registers and clinical folders. Socio-economic data were collected using interviewer-administered questionnaires. Composite variables were generated for social and economic factors and analysed using multiple logistic regression analysis.

Results
352 (69%) of the participants were clustered and 157 (31%) non-clustered. Clustered cases were more likely to have lived longer in the study community, (OR= 1.06, C.I: 1.02; 1.10, p=0.006, adjusted analysis) and in the same house (OR=1.04, C.I: 0.99; 1.08, p=0.06). Clustered cases had increased household crowding conditions (OR=0.45, C.I: 0.22; 0.96, p=0.038) and tended to have a high social score lower economic score.

Conclusions
We report a novel association between TB transmission and prolonged stay within a high burdened community. Our findings further suggest that even within low socio-economic settings, individuals at the lower end of the economic scale tend to be at greater risk of acquiring TB disease.
Introduction

While significant progress has been made to halt and reverse tuberculosis (TB) cases and deaths, the global burden of TB remains enormous, with an estimated 9.0 million incident cases reported for 2014 [1]. The Millennium Development Goals (MDGs) set for 2015, which aimed to halve the prevalence and mortality rates compared with 1990 were achieved in some of the regions of the world, but huge challenges still remain, particularly in the LMIC [1,2]. The WHO estimates that more than 24% of the world’s TB cases and deaths occurred in the African region [1]. With arguably the highest TB prevalence rates in the world of over 860/100 000, South Africa, remains one of the world’s top 6 high burden countries. More than 60% of incident TB cases in South Africa are HIV-infected [1].

The increasing advances in genotyping tools such as IS6110 restriction fragment length polymorphism (RFLP), spoligotyping and Mycobacterial Interspersed Repetitive Units (MIRU), among others, have been instrumental in our ability to investigate the epidemiology and transmission of tuberculosis in recent times [3-5]. In many of the TB molecular epidemiology studies, the aim has often been to distinguish between recent transmission or reactivation, exogenous re-infection from relapse or treatment failure in recurrent cases for recent transmission [6].

When two or more cases occur within a given time in a given place with the same genotype (or “clustered” strains) these are often considered to
be part of a linked common transmission chain. Hence, clustering is often used as a proxy for recent transmission [7]. Studies from different parts of the world have reported varying findings on risk factors for clustering such as age, immigrant status, HIV infection, homelessness, alcoholism and intravenous drug use [8-13]. However, there are some discrepancies across studies, particularly between the developed and LMIC contexts [2,14].

Related South African studies in gold mine workers where TB transmission rates remain high, have reported few distinct clustering patterns [15]. Other work from high-burdened community settings has reported moderate associations between biological factors such as HIV-status and clustering [16]. There remains a need to explore further, possible associations between TB transmission and TB socio-economic risk factors. The identification of such social and economic risk factors would be valuable in informing targeted control measures and interventions to interrupt disease transmission chains [3].

In the light of this background, our study aimed to explore the possible associations between TB transmission and a number of socio-economic risk factors in a high TB and HIV burdened community setting.

**Methods**

We conducted a cross-sectional study among TB participants resident in a geographically well-defined peri-urban township in Cape Town, South
Africa. This community had a population of approximately 13 000 people in 2006 which grew to around 18 000 in 2010. At least 1 in every 4 adults in this community was HIV-infected as of 2006 [4,17]. In the same year TB case notifications were as high as 2 000/100 000, despite the presence of a functional primary care TB facility and increasing ART coverage [18].

Potentially eligible TB participants attending the clinic were identified and informed about the study. Inclusion criteria included: TB disease notified from 2006 to end 2010, residency in the study community and willingness to provide written consent.

Clinical and demographic data were extracted from the TB registers and clinical folders. TB and socio-economic data were collected using interviewer-administered questionnaires that were translated to the participant’s local language. The questionnaires captured data on TB history, TB contacts, socio-economic conditions such as occupation, income level, educational level, sexual history in addition to other social and demographic data.

HIV testing and counselling (and referral for treatment where required) was conducted according to the national HIV guidelines [19]. Sputum specimens were obtained from TB suspects in accordance with the National TB testing, diagnostic and treatment guidelines [20].
Mycobacteriological tests, including microscopy and culture, were performed on the sputum specimens as described elsewhere [5].

*M.tbc* isolates from participants were analysed using *IS6110*-based RFLP, [21] performed at the Public Health Research Institute (PHRI), Tuberculosis Centre Laboratory, New Jersey. Based on the genotyping data, strains were classified using standard software and tools [22]. Previous analysis of the *M.tbc* strains by RFLP showed that the strains families predominant in the study population were the W-Beijing (29% of participants), CC-related strains (24%) in addition to other less well-represented strains such as the BM strain [16].

**Assumptions and definitions**

A strain was defined as a genetic variant of an isolate. A unique strain was an isolate with an RFLP pattern that occurred in only one patient within the study dataset. Unique strains were defined as non-clustered strains. We have previously defined a strain family as a group of strains exhibiting similar but not identical *IS6110* profiles [16]. A cluster was defined as >1 specific strain detected in different individuals within the study population. Strains from dually infected participants were analysed as individual samples. Reactivation tuberculosis cases resulting from the same strain as the patient’s previous TB episode were presumed to be due to relapse and were excluded from analysis. Strains with <6 copies of *IS6110* (low bandwidth strains) are known to be poorly differentiated by the RFLP technique and so were excluded from further analysis [23].
For this sub-study, we restricted the analysis to adult participants (≥15 years of age) who had both RFLP and socio-economic questionnaire data. We excluded children, (12 children in dataset) as the social and economic behaviour of children (below 15 years of age) were presumed to be different from those of adults.

**Statistical Analysis**

Data was analysed using StataCorp version 12 software. A stratified analysis was performed to determine baseline differences in the socio-economic factors between the clustered and the non-clustered strains.

Composite scores were generated for social and economic risk factors. The following factors were incorporated in the social score: alcohol consumption, shebeen patronage, meeting regularly with a group, number of new sexual partners within the past 6 months and number of occupants living in the same house. It is also notable that while the majority of those participants who report visiting shebeens consume alcohol, there is also a proportion that visits shebeens for social or other reasons besides alcohol consumption. Furthermore not all alcohol consumption occurs on shebeen premises. An exploratory analysis showed that although there was a correlation between alcohol drinking and shebeen patronage (r=0.628), this correlation was not strong. Therefore we chose to keep both variables (shebeen patronage and alcohol consumption) in the social score. Each variable was assigned a
value of 0, 1 or 2 (depending on the number of categories in the variable), with a higher score corresponding to greater social interaction. For example shebeen patronage was scored 0 for not having attended a shebeen in the past 6 months and 1 if a participant had attended a shebeen in the time period. Education level, employment status, income level, electricity access, having a toilet in the house, number of people sleeping in the same room and number of rooms used for sleeping (size of house) were all classified as economic factors and were used to construct the composite economic score. Participants who had partial data in the questionnaire were not excluded but were assigned the lowest value for the missing score variable. A higher economic score correlated with higher economic status. Both the economic and social scores were divided into binary variables at the median (to generate a “low” and “high” economic and social score respectively). Additional risk factors not classifiable social or economic risk factors but considered important, and thus included in the analyses were: a history of TB contacts, recent death in family, tobacco smoking, period of residence in the same house/Masiphumelele, history of mine work, history of imprisonment and time spent outside Masiphumelele. Multiple logistic regression analyses were performed to determine associations between TB transmission, social and economic score and the other specified risk factors.

The University of Cape Town Human Research Ethics Committee (HREC) approved this study.
Results

Sample selection

Figure 1 is a consort diagram summarising the study sample selection. In summary, out of the 2,237 participants recruited into the parent study, 631 participants had complete RFLP data and of these, 584 also had socio-economic data available. Following additional exclusions as described earlier, the final sample comprised of 509 participants, and 511 TB strains (2 of the participants had dual infection). Dually infected participants were analysed as single participants for baseline demographic data but were then split into separate strains.

Of the 509 participants who had complete data available, 352 (69%) were classified as clustered while the remaining 157 participants (31%) were classified as non-clustered.

Demographic characteristics

In general, there was an even distribution in the demographic factors between the clustered and non-clustered cases. Gender and age were uniformly distributed across the two groups and the majority of the study participants were isiXhosa speaking (Table 1).

Economic risk factors

There were no significant differences in economic variables between clustered and non-clustered cases. While approximately 30% of the participants reported having acquired only primary level education in both groups, slightly more clustered cases reported “some” secondary
education (50% compared to 48% in non-clustered cases), and 16% of the clustered cases reported a matric or higher education compared to 18% of the non-clustered cases (p=0.65). Unemployment levels were very high in this study population at approximately 67% in both groups. Monthly household income was generally low with a higher proportion of the clustered cases (5% vs. 0.88%) earning > R5000 per month.

**Social factors and housing conditions**

In this cohort, the number of houses on a single plot ranged from 1 to ≥6 houses, with clustered cases reporting a higher percentage of ≥6 houses/plot compared to non-clustered cases (32% vs. 27% respectively; p=0.39). The majority of study participants lived in informal dwellings (89% across both strata). Overall, 10% of the participants reported having no toilet or running water in the house with 90% reporting the use of a communal water tap for household water supply (p=0.77 for comparison across groups). However, 96% did report having access to electricity in their house (Table 1).

The number of occupants living in a household ranged from 2 to ≥5. Up to one third of the households reported having 3-4 occupants, while 24% vs. 28% of the clustered participants reported having more than 5 occupants as compared to the non-clustered, respectively. The majority of households (59% vs. 57%) reported a single room for sleeping, however, non-clustered cases were more likely to have more rooms for sleeping (9% had ≥3 rooms compared to 5% of the clustered cases; p=0.04). No
significant differences were noted in the number of people sharing a room for sleeping, between strata (Table 1).

**Other TB risk factors**

Of the 483 study participants who had a known HIV status, proportions of HIV positive participants were comparable between clustered and non-clustered cases (63 % vs. 67 %, p=0.359). Across the cohort, 30% of participants reported knowing someone currently on TB treatment. A marginally greater proportion of clustered cases reported a recent death (of any cause) in their household (15% vs.10 %, p=0.09). Overall 23% of clustered cases reported a household member known to have died recently due to TB, compared to 21% of non-clustered cases (p=0.67).

Although not statistically significant, clustered cases were more likely to report alcohol consumption (39 %vs. 32 % p=0.10), and patronage of a shebeen in the past 12 months (28 % vs. 24 %, p=0.15). Only a small proportion of the participants (7% clustered vs. 8% non-clustered) reported a history of mine work. Similarly, only 2% of participants reported having been in prison in the past 6 months. The majority of the participants reported using a taxi regularly for transport (95% of clustered cases and 93% of non-clustered cases; p=0.42).

There was a strong negative association between clustering (as compared to non-clustering) and having 1 new sexual partner (18% vs. 25%, p=0.02). Clustered cases were more likely to have stayed for longer
period in the same house, (median=3 years vs. 2 years, p=0.06). There was also strong evidence of an association between longer stay in the study community and clustering (median of 6 years versus 4.5; p=0.01). The composite social score across the cohort ranged from 0-11 (median=5). When categorised into high and low scores, a moderately higher proportion of the clustered cases had a high social score: 61% compared to 53% of the non-clustered cases (p=0.08). The economic score ranged from 1-9, with a median of 4. When categorised, a lower proportion of the clustered cases reported a high economic score: 21% compared to 28% of the non-clustered cases (p=0.09) (Table 1).

**Multivariate analysis between TB transmission and socio-economic risk factors**

For the multivariate analysis, we explored the association between TB transmission and selected risk factors/variables. Variables that were included in the regression model included those variables which trended towards being associated with clustering in the bivariate analysis: including having been in contact with a known TB case within the past 12 months, knowing someone who had died of TB in the past 12 months and time spent outside the study community. Time lived in current house strongly correlated to time lived in the study community and so was not included in the regression model. Based on our prior knowledge about the risk factors for TB transmission, age, gender and HIV status are well known potentially confounding variables, and so were adjusted for in the regression model.
Table 2 shows a summary of the multivariate analysis, based on these variables. There was a strong positive association observed between a longer duration of stay in the study community and clustering (OR=1.06, C.I: 1.02 to 1.10, p=0.006). Also notable was weak evidence of an association between a lower economic score and clustering (OR=0.69, C.I: 0.45 to 1.06, p-value=0.09).

**Discussion**

The application of new tools in exploring the factors associated with TB transmission, among them, molecular epidemiological approaches, particularly in high burden settings remains pertinent. To date, limited studies have been conducted, particularly in sub-Saharan Africa to show associations between molecular genotypes and social risk factors for TB transmission. Our study is therefore one of the few reported molecular epidemiological studies investigating the role of socio-economic risk factors in TB transmission, in a high burden setting.

In summary, the demographic characteristics of the study participants were similar between the clustered and non-clustered cases across most of the risk factors investigated, with a few strong associations identified between socio-economic risk factors and clustering.

In an attempt to quantify social interaction, we created a composite social score. Although, we found no overall association between TB
transmission and the composite social score for this cohort, we did identify some individual measures that had evidence of association with transmission. Specifically, our results suggested that having fewer new sexual partners in the past 6 months might be protective against TB transmission. Others have posseted that sexual behaviour may be a social risk factor for TB transmission [24], but few studies have demonstrated this. We hypothesize that this study finding may be indicative of the exposure to fewer TB contacts. In addition, both a longer stay in the same house and longer time living in the community were associated with belonging to a TB transmission cluster. These associations may be a measure of prolonged and persistent exposure to *M. tb* in the community, with increasing chance of acquiring TB infection for participants living in the community for long periods of time. Although an intuitive finding, to our knowledge, this is the first study to show that prolonged stay within a high burden TB community with high rates of ongoing TB transmission results in an increased risk of being part of a TB transmission cluster.

We also explored economic variables, both individually and by creating a composite economic score. There was a significant association noted between TB transmission and the number of rooms used for sleeping, with participants who reported having more than 3 rooms for sleeping being less likely to be part of a transmission cluster. This association may point to reduced close indoor contact time, hence reduced risk of TB transmission for those who have more spacious houses. The remaining
economic factors, taken individually did not yield any strong associations with TB transmission. However, a weak association was observed with the composite economic score: a lower economic score was positively associated with TB transmission. To some extent, these findings are in agreement with other researchers who have reported that poor socio-economic conditions may predispose to TB transmission/clustering [25]. But further, given the setting of a low economic community, these finding suggest that even in poorer communities there is a “sliding-scale of poverty”, with individuals at the lower end of the economic scale being at greater risk for acquiring TB infection. The factors that linked to economic status, which in turn may explain this association are complex and may include poor nutritional and health status among other related and underlying factors.

While our results show potential epidemiological links between TB transmission and time lived in a high burdened community, we were surprised by the paucity of associations with many of the risk factors investigated or the composite social and economic scores. A recent study by Mathema et al in South African gold-miners also could not establish any risk factors for TB transmission and this was possibly due to a universally high risk for disease among that population [15]. Our findings in this study point to a similar scenario, with difficulty identifying specific transmission risk factors in a generally low socio-economic community with exceptionally high TB disease and TB transmission rates [26].
Potential limitations for our study may include information bias and recall bias, when answering the questionnaire. However, we sought to minimise information bias by using trained interviewers fluent in the local language. We also note that while all TB participants were invited to participate in the study, participation was voluntary and such “self-selection” into a study cannot be ruled out completely. Previous analyses of this dataset have reported few significant differences in patients with RFLP data and those without. One difference is that MDR-TB patients were more likely to have RFLP data and that patients who had died were less likely to have RFLP data [16]. In order to rule out any additional biases, we explored whether there were any differences between the participants who were enrolled in the parent study (n=2237) versus the 509 adult patients who had socio-economic questionnaire data. There were no differences in age (p=0.78), gender (p=0.15), ART status (p=0.14) or MDR-TB (p=0.56). Furthermore, there were no differences noted between the 509 patients included in this study and the remaining sputum positive patients who were not included in the analysis in terms of age (p=0.67), gender (p=0.92), HIV or ART status (p=0.36 and 0.66 respectively), MDR-TB (p=0.42) and treatment outcomes (p=0.28).

Another potential limitation in this study is that our sample size of 509 may not have been adequate to detect small statistical differences. This could potentially explain the marginal trends for some of the factors analysed in this study. Further work to study these trends in much larger populations could bring more definitive insights into the social and
economic factors linked with TB transmission in high burdened communities. We also note that our scoring algorithm has not been validated, and this may be another limitation to our study. Missingness of data for some of the variables used to generate the composite social/economic score variables may have biased our results. It is likely that we may have underestimated the social social/economic scores, and this, in turn, may have resulted in overestimation of any associations that we report in this study may. Another limitation, which is inherent with most molecular epidemiological studies, is the misclassification of participants due to missing epidemiological data or missing specimens. This is particularly notable given our definition of a cluster (>1 specific strain detected in different individuals). It is therefore possible that based on our definition; we may have under-classified strains as clustered due to missing specimens, which may have turned a unique strain into a cluster or two strains.

In summary, our study has shown that greater social interaction, poor economic conditions even in the context of a low socio-economic community, and prolonged residence in a high burdened community are potentially important factors linked to TB transmission in this setting. While the association between poverty and TB transmission is not new, the importance of degrees of poverty within low socio-economic setting is novel. The association between TB transmission and prolonged stay in a high transmission community is intuitive, but has not been shown before and was demonstrated in this study. Our study further re-enforces the
notion that TB is a social disease; hence social as well as individual level interventions may be recommendable in order to address the on-going TB transmission in low socio-economic status and high burdened communities where exposure is prolonged and TB transmission remains on-going. Further studies to analyse the associations we report here in larger sample sizes across different populations would be recommended in order to provide definite recommendations for National TB programmes programmes.
Figure 1: Consort diagram illustrating the sample selection process

2,237 TB cases diagnosed at Clinic

1,836 Pulmonary TB cases

1,353 Sputum Smear/culture confirmed TB cases

1,011 specimens collected

825 sputum positive specimens

772 completed social data questionnaire

631 with RFLP data

584 with completed questionnaire + RFLP data

Exclude age ≤14 (n=15)

Exclude low bandwidths and reactivations in same patient (n=62)

Final analysis sample = 509
Table 1: Summary of the baseline demographic socio-economic TB transmission risk factors stratified by clustering/non-clustering of *M.tuberculosis* strains. Also shown are in this table are the bivariate associations (n=509) participants.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Total: n (%) or median (IQR)</th>
<th>Non-clustered cases N=157 (31%) or median (IQR)</th>
<th>Clustered cases N=352 (69%) or median (IQR)</th>
<th>Bivariate Associations</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sex: Male</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>294 (58)</td>
<td>94 (59)</td>
<td>200 (57)</td>
<td>0.642</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.09 (0.75; 1.59)</td>
</tr>
<tr>
<td><strong>Known HIV Status:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positive</td>
<td>319 (66)</td>
<td>225 (67)</td>
<td>94 (63)</td>
<td>0.359</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.21 (0.81; 1.81)</td>
</tr>
<tr>
<td><strong>Age (continuous)</strong></td>
<td>32 (27; 40)</td>
<td></td>
<td></td>
<td>0.935</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.935 (0.98; 1.02)</td>
</tr>
<tr>
<td><strong>Age category:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15-24 years</td>
<td>81 (15)</td>
<td>31 (20)</td>
<td>50 (14)</td>
<td>1.0</td>
</tr>
<tr>
<td>25-34 years</td>
<td>214 (41)</td>
<td>61 (39)</td>
<td>153 (44)</td>
<td>0.107</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.55 (0.91; 2.66)</td>
</tr>
<tr>
<td>35-49 years</td>
<td>179 (33)</td>
<td>55 (35)</td>
<td>124 (35)</td>
<td>0.221</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.41 (0.81; 2.43)</td>
</tr>
<tr>
<td>&gt;= 50 years</td>
<td>35 (8)</td>
<td>12 (8)</td>
<td>23 (7)</td>
<td>0.548</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.29 (0.56; 3.02)</td>
</tr>
<tr>
<td><strong>Language</strong>:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IsiXhosa</td>
<td>472 (93)</td>
<td>145 (92)</td>
<td>327 (93)</td>
<td>1.0</td>
</tr>
<tr>
<td>Other</td>
<td>37 (7)</td>
<td>11 (8)</td>
<td>26 (7)</td>
<td>0.836</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.90 (0.34; 2.37)</td>
</tr>
<tr>
<td><strong>Education</strong>:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>None</td>
<td>31 (5)</td>
<td>10 (5)</td>
<td>21 (5)</td>
<td>1.00</td>
</tr>
<tr>
<td>Some primary</td>
<td>150 (30)</td>
<td>45 (29)</td>
<td>105 (30)</td>
<td>0.658</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.23 (0.49; 3.11)</td>
</tr>
<tr>
<td>Some high school</td>
<td>248 (49)</td>
<td>75 (48)</td>
<td>173 (50)</td>
<td>0.652</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.23 (0.50; 3.03)</td>
</tr>
<tr>
<td>Matric and higher</td>
<td>81 (16)</td>
<td>27 (18)</td>
<td>54 (16)</td>
<td>0.952</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.02 (0.39; 3.03)</td>
</tr>
</tbody>
</table>
Employment Status:  
<table>
<thead>
<tr>
<th></th>
<th>343 (67)</th>
<th>104 (66)</th>
<th>239 (67)</th>
<th>1.00</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unemployed</td>
<td>163 (32)</td>
<td>53 (33)</td>
<td>110 (32)</td>
<td>0.598</td>
</tr>
</tbody>
</table>

Income (n= 167)  
|       | 1500 (1000; 2200) | 0.176 | 1.0 (0.99; 1.00) |
|< R2000 | 113 (62) | 36 (58)  | 77 (64)  | 1.00 |
|R2001– R5000 | 65 (36) | 23 (37)  | 42 (35)  | 0.699| 0.88 (0.46; 1.72) |
|> R5000 | 4 (2)   | 3 (5)    | 1 (0.8)  | 0.115| 0.16 (0.16; 1.57) |

Number of houses on plot:  
|       | 140 (28) | 45 (29)  | 95 (27)  | 1.00 |
|1 house | 83 (16)  | 26 (17)  | 57 (16)  | 0.969| 1.01 (0.57; 1.80) |
|2-3 houses | 130 (26) | 43 (27)  | 87 (25)  | 0.974| 0.99 (0.60; 1.64) |
|4-5 houses | 153 (30) | 43 (27)  | 110 (32) | 0.391| 1.24 (0.76; 2.04) |
|6 or more houses | 57 (11) | 19 (12)  | 38 (11)  | 1.00 |
|Type of housing: formal | 57 (11) | 19 (12)  | 38 (11)  | 1.00 |
|Informal | 449 (89) | 138 (88) | 311 (89) | 0.760| 0.91 (0.51; 1.64) |

Source of Water*:  
|       | 458 (90) | 143 (91) | 315 (90) | 0.772| 0.91 (0.47; 1.74) |
|Communal tap | 487 (96) | 150 (95) | 337 (96) | 0.694| 1.21 (0.47; 3.09) |

Electricity in house: Yes  
|       | 487 (96) | 150 (95) | 337 (96) | 0.694| 1.21 (0.47; 3.09) |

Toilet in house: Yes  
|       | 48 (9)   | 13 (9)   | 35 (10)  | 0.533| 1.23 (0.63; 2.40) |

Number of occupants in participant’s house  
|       | 3 (2.5)  | 0.78     | 0.96 (0.89; 1.04) |
### Number of occupants in participant’s house*

<table>
<thead>
<tr>
<th>1-2</th>
<th>241 (42)</th>
<th>75 (41)</th>
<th>166 (42)</th>
<th>1.00</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>191 (33)</td>
<td>57 (31)</td>
<td>134 (34)</td>
<td>0.917</td>
</tr>
<tr>
<td></td>
<td>1.02 (0.66; 1.58)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3-4</td>
<td>148 (25)</td>
<td>51 (28)</td>
<td>97 (24)</td>
<td>0.535</td>
</tr>
<tr>
<td></td>
<td>0.86 (0.54; 1.38)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 or more</td>
<td>148 (25)</td>
<td>51 (28)</td>
<td>97 (24)</td>
<td>0.535</td>
</tr>
<tr>
<td></td>
<td>0.86 (0.54; 1.38)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Number of rooms used for sleeping*

<table>
<thead>
<tr>
<th>1 room</th>
<th>299 (59)</th>
<th>89 (57)</th>
<th>210 (59)</th>
<th>1.00</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 rooms</td>
<td>176 (35)</td>
<td>53 (34)</td>
<td>123 (37)</td>
<td>0.970</td>
</tr>
<tr>
<td></td>
<td>0.99 (0.66; 1.48)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3-6 rooms</td>
<td>31 (6)</td>
<td>15 (9)</td>
<td>16 (6)</td>
<td>0.038</td>
</tr>
<tr>
<td></td>
<td>0.45 (0.22; 0.66)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Number of people sleeping in same room

<table>
<thead>
<tr>
<th>1 to 2</th>
<th>402 (69)</th>
<th>133 (73)</th>
<th>269 (68)</th>
<th>1.00</th>
</tr>
</thead>
<tbody>
<tr>
<td>3-5</td>
<td>172 (30)</td>
<td>48 (26)</td>
<td>124 (31)</td>
<td>0.20</td>
</tr>
<tr>
<td></td>
<td>1.32 (0.87; 2.00)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6 or more</td>
<td>6 (1)</td>
<td>2 (1)</td>
<td>4 (1)</td>
<td>0.73</td>
</tr>
<tr>
<td></td>
<td>0.73 (0.12; 4.42)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Know anyone treated for TB currently

<table>
<thead>
<tr>
<th>174 (30)</th>
<th>55 (30)</th>
<th>119 (30)</th>
<th>0.83</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.05 (0.69; 1.58)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Know anyone treated for TB in last 12 months

<table>
<thead>
<tr>
<th>127 (22)</th>
<th>36 (20)</th>
<th>91 (23)</th>
<th>0.20</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.35 (0.85; 2.13)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Know anyone ever treated for TB

<table>
<thead>
<tr>
<th>111 (19)</th>
<th>37 (20)</th>
<th>74 (19)</th>
<th>0.91</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.03 (0.64; 1.67)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Recent death in House

<table>
<thead>
<tr>
<th>70 (14)</th>
<th>16 (10)</th>
<th>54 (15)</th>
<th>0.09</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.65 (0.91; 2.97)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Recent death due to TB

<table>
<thead>
<tr>
<th>18 (23)</th>
<th>4 (21)</th>
<th>14 (23)</th>
<th>0.59</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.51 (1.83; 12.58)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Alcohol consumption

<table>
<thead>
<tr>
<th>214 (37)</th>
<th>58 (32)</th>
<th>156 (39)</th>
<th>0.10</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.39 (0.94; 1.83)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Count (%)</td>
<td>Count (%)</td>
<td>Count (%)</td>
</tr>
<tr>
<td>------------------------------</td>
<td>-----------</td>
<td>-----------</td>
<td>-----------</td>
</tr>
<tr>
<td><strong>Been to a shebeen in the last 6 months</strong></td>
<td>81 (73)</td>
<td>21 (24)</td>
<td>60 (28)</td>
</tr>
<tr>
<td><strong>Smoked tobacco in last 6 months</strong></td>
<td>198 (34)</td>
<td>54 (30)</td>
<td>144 (37)</td>
</tr>
<tr>
<td><strong>Prison in last 6 months</strong></td>
<td>12 (2)</td>
<td>3 (2)</td>
<td>9 (2)</td>
</tr>
<tr>
<td><strong>Been to clinic or hospital in the last 6 months</strong></td>
<td>212 (42)</td>
<td>63 (40)</td>
<td>149 (43)</td>
</tr>
<tr>
<td><strong>Take taxi regularly</strong></td>
<td>478 (94)</td>
<td>146 (93)</td>
<td>332 (95)</td>
</tr>
<tr>
<td><strong>Meet group regularly</strong></td>
<td>49 (9)</td>
<td>17 (9)</td>
<td>32 (8)</td>
</tr>
<tr>
<td><strong>Number of new sex partners in last 6 months:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>405 (73)</td>
<td>118 (75)</td>
<td>287 (82)</td>
</tr>
<tr>
<td>1</td>
<td>102 (20)</td>
<td>39 (25)</td>
<td>63 (18)</td>
</tr>
<tr>
<td>2</td>
<td>5 (3)</td>
<td>2 (2)</td>
<td>3 (3)</td>
</tr>
<tr>
<td>3</td>
<td>3 (2)</td>
<td>1 (1)</td>
<td>2 (1)</td>
</tr>
<tr>
<td><strong>Time lived in current house (years)</strong></td>
<td>3.0 (0.7; 8.0)</td>
<td>2.0 (0.6; 7.0)</td>
<td>3.8 (0.8; 8.8)</td>
</tr>
<tr>
<td><strong>Time lived in study community (years)</strong></td>
<td>6.0 (2.0; 10.0)</td>
<td>4.5 (1.0; 10.0)</td>
<td>6.2 (2.1; 11.0)</td>
</tr>
<tr>
<td><strong>Time spent outside study community (months)</strong></td>
<td>3.0 (0.5; 5.0)</td>
<td>3.5 (1.5)</td>
<td>3.0 (0.5; 5.0)</td>
</tr>
</tbody>
</table>

Social Score
<table>
<thead>
<tr>
<th></th>
<th>Mean (Min; Max)</th>
<th>95% Confidence Interval</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>High Social Score (≥5)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Continuous)</td>
<td>5 (4; 6)</td>
<td>5 (4; 6)</td>
<td>0.49</td>
</tr>
<tr>
<td>E Economic Score</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Continuous)</td>
<td>4 (3; 5)</td>
<td>4 (3; 4)</td>
<td>0.22</td>
</tr>
<tr>
<td><strong>High Economic Score (≥4)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Continuous)</td>
<td>118 (23)</td>
<td>44 (28)</td>
<td>0.09</td>
</tr>
</tbody>
</table>
Table 2: Summary of the baseline socio-economic factors stratified by clustering/non-clustering of *M.tb* strains and the bivariate associations for n=509 participants

<table>
<thead>
<tr>
<th>Patient characteristic</th>
<th>Adjusted Analysis</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Odds Ratio</td>
<td>95% Confidence Interval</td>
<td>P-value</td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>0.99</td>
<td>0.42; 1.03</td>
<td>0.563</td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td>1.18</td>
<td>0.79; 1.77</td>
<td>0.411</td>
<td></td>
</tr>
<tr>
<td>HIV Status</td>
<td>1.00</td>
<td>0.92; 1.09</td>
<td>0.974</td>
<td></td>
</tr>
<tr>
<td>TB contact in the last 12 months</td>
<td>1.26</td>
<td>0.78; 2.04</td>
<td>0.340</td>
<td></td>
</tr>
<tr>
<td>Knowing someone who died in the last 12 months</td>
<td>1.48</td>
<td>0.81; 2.71</td>
<td>0.208</td>
<td></td>
</tr>
<tr>
<td>Time resident in study community</td>
<td>1.06</td>
<td>1.02; 1.10</td>
<td>0.006</td>
<td></td>
</tr>
<tr>
<td>Social Score</td>
<td>1.35</td>
<td>0.91; 2.00</td>
<td>0.134</td>
<td></td>
</tr>
<tr>
<td>Economic score</td>
<td>0.66</td>
<td>0.42; 1.03</td>
<td>0.067</td>
<td></td>
</tr>
</tbody>
</table>
Conflict of interest:

The authors declare that they have no conflicting interest.

Funding: This particular sub-study did not receive any external finding

Corresponding author: Dr Keren Middelkoop, Desmond Tutu HIV Centre, Institute of Infectious Disease & Molecular Medicine, University of Cape Town, PO Box 13801, Mowbray Observatory, Cape Town 7705, South Africa. Email: Keren.Middelkoop@hiv-research.org.za

Telephone: +27 21 650 6960

Fax: +27 21 650 6963
References


25. Fok A, Numata Y, Schulzer M, FitzGerald MJ. Risk factors for clustering of tuberculosis cases: a systematic review of population-

APPENDICES
INTRODUCTION:

You are being invited to take part in this research study because you are a resident of Masiphumelele. The doctor in charge of this study at Masiphumelele is Dr. LG Bekker. Before you decide if you want to be a part of this study, we want you to know about the study.

This is a consent form. It gives you information about this study. The research staff will talk with you about this information. You are free to ask questions about this study and discuss any worries you may have with the research staff. If you agree to take part in this study, you will be asked to sign this consent form. You will be given a copy of the form to keep.

WHY IS THIS STUDY BEING DONE?

We want to learn how and where tuberculosis (TB) is being spread from person to person in Masiphumelele. This information will help us to manage this health problem more effectively in the community.

WHO CAN TAKE PART IN THE STUDY?

We are asking everyone who lives in Masiphumelele who has TB and is receiving TB treatment at a clinic if they will take part in the study. We estimate that approximately 2900 people will participate in this study.

WHAT DO I HAVE TO DO IF I TAKE PART?

We will ask you for 40 minutes of your time, during which you will complete a questionnaire. The questionnaire asks about your age, gender, occupation, living conditions, previous TB and treatment, TB contacts, TB risk factors, and social activities.

You will also be asked for a sample of saliva (spit). For this, you will be asked to place a cotton pad in your mouth for at least 2 minutes to collect this sample. An HIV (Human Immunodeficiency Virus) test will be done on the spit, but your name will not be written on the specimen. In this way, we will be able to see how many TB patients have HIV, but we will not know your individual HIV status from this test. The research team strongly encourages everyone to know their HIV status. If you would like to know your
HIV status, you may attend the voluntary counselling and testing (VCT) service at the Masiphumelele clinic (Monday to Friday, 8:30 – 4:30 pm). This service is available on the same day as your study visit.

We will ask permission to perform extra tests on the slime sample you produced when diagnosed with TB, and on any other slime samples required during your TB treatment. We will be investigating the TB germ you are infected with to see how similar it is to other TB germs in the community. An example of this test is called Restriction Fragment Length Polymorphism. You do not need to give blood or any extra specimens for this test. The TB grown from your slime may be sent to the United States of America for confirmation of these tests.

We will ask permission to look at your clinic files and use your demographic, clinical and laboratory information recorded in them. This means that if you have attended VCT at the clinic, we will know your HIV status. This will always remain confidential and your name will not be attached to this information.

We will store the TB germ obtained from your sputum/slime specimen. This germ, as well as your clinical information, may be used in further, related studies and work, but no information identifying you will be used.

We may use this information to map the distribution or spread of TB disease in the community. Every precaution will be taken to ensure this information cannot be linked to you, and neither the community name, nor your name will be published with this information.

WHAT ARE THE RISKS OF TAKING PART IN THE STUDY?

Some of the questions asked in the questionnaire may make you feel uncomfortable. However there are no significant risks to participating in this study.

There is a potential risk of loss of confidentiality. Every effort will be made by the study staff to ensure that this does not happen and that your information is protected. The study team will keep your personal information confidential. You will be given a study number. The questionnaire and spit specimen for HIV testing will be labelled with this study number and NOT with your name. All information will be kept in a locked cupboard. We cannot guarantee absolute confidentiality.

WHAT ARE THE BENEFITS OF TAKING PART IN THE STUDY?

The benefit of this study is that we hope to learn better ways to reduce the spread of TB in your community. There are no direct benefits to you.

WHAT ARE MY ALTERNATIVES?
You may choose not to take part in this study. You may choose to know your HIV status and be tested at the Masiphumelele clinic.

WHAT ABOUT CONFIDENTIALITY?

Your records may be reviewed by the University of Cape Town Institutional Review Board, local Ministry of Health, study staff and study monitors.

Any publication of this study will not use your name or identify you personally.

WHAT ARE THE COSTS TO ME?

There are no costs to you if you are involved in this study.

WILL I RECEIVE ANY PAYMENT?

You will receive re-imbursment for your time for taking part in this study.

WHAT HAPPENS IF I AM INJURED?

If you are injured as a result of being in this study, you will be given immediate treatment for your injuries. The cost for this treatment will be provided by the Department of Health at the clinic or appropriate referral hospital. There is no program for compensation through this institution.

WHAT ARE MY RIGHTS AS A RESEARCH SUBJECT?

Taking part in this study is completely voluntary. You may choose not to take part in this study. You will not be giving up any of your legal rights by signing this consent form.

CAN I REFUSE TO TAKE PART IN THE STUDY?

Yes, you may refuse to take part in the study. If you decide not to take part in the study, this will not affect any medical care or treatment you may require now or in the future.

IN VOLUNTARY WITHDRAWAL OR EARLY WITHDRAWAL

If at any time during the study you refuse to complete one of the study procedures you will be withdrawn from the study. You may also decide at any point that you no longer wish to participate in the study and may then withdraw from the study by informing the study staff about your decision.

In either event, you will be free to withdraw from the study with no further obligations
The study may be discontinued at any time by the national health agency or by the Faculty of Health Sciences Human Research Ethics Committee (REC). (An REC is a committee that watches over the safety and rights of research subjects.)

WHAT DO I DO IF I HAVE QUESTIONS OR PROBLEMS?

If you have any questions about this study or experience a research-related injury, you may contact:

   Dr Linda Gail Bekker at (021) 650 6966

If you have any questions about your rights as a research participant, you may contact the following member of the ethics committee:

   Prof Marc Blockman (021) 406 6338
<table>
<thead>
<tr>
<th>Participant ID</th>
<th>Demographics</th>
<th>Date of Interview</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### 1. TB Register Number:

### 2. Interviewer:

### 3. If the patient is between the ages of 11 and 17 who is responding to the questionnaire?

- [ ] Not Applicable
- [ ] Participant only
- [ ] Both the participant and Parent/Legal guardian

#### 3a. If the parent/legal guardian is present, what is the relation of that person to the participant?

- [ ] Parent
- [ ] Aunt/Uncle
- [ ] Sibling
- [ ] Other, specify: __________

### 4. How old are you?

### 5. Gender?

- [ ] Male
- [ ] Female

### 6. What is your home language?

- [ ] Afrikaans
- [ ] Xhosa
- [ ] English
- [ ] Other, specify: __________

### 7. What is the highest level of education that you have completed?

- [ ] None
- [ ] Grade 1/Sub A
- [ ] Grade 2/Sub B
- [ ] Grade 3/Std 1
- [ ] Grade 4/Std 2
- [ ] Grade 5/Std 3
- [ ] Grade 6/Std 4
- [ ] Grade 7/Std 5
- [ ] Grade 8/Std 6
- [ ] Grade 9/Std 7
- [ ] Grade 10/Std 8
- [ ] Grade 11/Std 9
- [ ] Grade 12/Std 10
- [ ] Tertiary (university/Technikon/College with degree)
8. Approximately, how long have you lived in your present house?  
   Weeks or Months or Years

9. Approximately, how long have you lived in Masiphumelele?  
   Weeks or Months or Years

10. Have you spent at least one night outside of Masiphumelele in the last six months?  
    yes no  
    If NO skip to Q11

   10a. If yes, how many nights were you outside of Masiphumelele?  
       (Specify number of nights, weeks or months)

11. Including yourself, how many people (adults & children) live in your house?  

12. How many people sleep in the same room that you sleep in?  

13. Are you currently working?  
    yes no  
    If NO skip to Q14

   13a. If yes, what type of work do you do?  
       Employed full-time  
       Employed part-time  
       Informal job/hawker

14. What is your TOTAL monthly income for your household?  

15. Do you know anyone who is being treated for TB currently?  
    yes no  
    If NO skip to Q16

   15a. If yes, how many people currently being treated for TB do you know?  

Version: 1, 1 25 April 2006
15b. Now we are going to talk about the people you know who are currently being treated for TB. We only want to know where you meet with these people, starting with the person you see most frequently? (tick all that apply)

<table>
<thead>
<tr>
<th>People That You Know Who are Currently Being Treated for TB</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
</tr>
<tr>
<td>Neighbor</td>
</tr>
</tbody>
</table>

**Interviewer:** Use one column for each person that the participant knows who is currently being treated for TB. If the participant knows more than 10 people currently being treated for TB, write down the information for the 10 which he/she has the most contact with. A neighbor is a person living on the same or an adjacent plot.

16. Do you know anyone who has been treated for TB in the **last 12 months**? **Interviewer:** Note that this excludes individuals mentioned in question 15.  

16a. If yes, how many other people being treated for TB in the **last 12 months** do you know?
16b. Now we are going to talk about the people you know who have been treated for TB in the last 12 months. We only want to know where you meet with these people, starting with the person you see most frequently? (tick all that apply) This excludes individuals mentioned in Question 15.

<table>
<thead>
<tr>
<th>People That You Know Who Have Been Treated for TB in the Last 12 Months</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
</tr>
<tr>
<td>Neighbor</td>
</tr>
<tr>
<td>Work</td>
</tr>
<tr>
<td>School</td>
</tr>
<tr>
<td>Clinic</td>
</tr>
<tr>
<td>Church</td>
</tr>
<tr>
<td>Shebeen</td>
</tr>
<tr>
<td>Shared tap</td>
</tr>
<tr>
<td>Other</td>
</tr>
<tr>
<td>Specify</td>
</tr>
</tbody>
</table>

**Interviewer:** Use one column for each person that the participant knows who has been treated for TB in the last 12 months. If the participant knows more than 10 people who have been treated for TB in the last 12 months, write down the information for the 10 which he/she has the most contact with. A neighbor is a person living on the same or an adjacent plot. This excludes individuals mentioned in #15.

17. Do you know anyone else who has **EVER** been treated for TB in the past (who you have not mentioned before)? □ □ □ □ □ □ □ □ □ □ If NO skip to Q18

17a. If yes, how many other people do you know who have **EVER** been treated for TB in the past? □ □
17b. Now we are going to talk about the people you know who have EVER been treated for TB in the past, who you have not already mentioned before. We only want to know where you meet with these people, starting with the person you see most frequently? (tick all that apply) This excludes individuals mentioned in Question 15 and 16.

<table>
<thead>
<tr>
<th>People That You Know Who Have Ever Been Treated for TB in the Past</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
</tr>
<tr>
<td>Neighbor</td>
</tr>
<tr>
<td>Work</td>
</tr>
<tr>
<td>School</td>
</tr>
<tr>
<td>Clinic</td>
</tr>
<tr>
<td>Church</td>
</tr>
<tr>
<td>Shebeen</td>
</tr>
<tr>
<td>Shared tap</td>
</tr>
<tr>
<td>Other</td>
</tr>
<tr>
<td>Specify</td>
</tr>
</tbody>
</table>

**Interviewer:** Use one column for each person that the participant knows who has ever been treated for TB in the past. If the participant knows more than 10 people who have ever been treated for TB in the past, write down the information for the 10 which he/she has the most contact with. A neighbor is a person living on the same or an adjacent plot.
Participant ID

Housing

18. How many houses are on your plot/erf? .............................................

19. What type of house do you live in? .........................................................
   □ Informal (zinc/plastic/wood)
   □ Formal (Brick)

19a. How many rooms in the house are used for sleeping? ..........................

20. Does your house have a water tap? .......................................................  
   □ Yes  □ No
   If YES skip to Q21

20a. If no, do you use a communal tap? .....................................................
   □ Yes  □ No
   If NO then skip to Q21
   If no, what other source(s) of water do you use?

20b. Where is the communal tap located? ...................................................

20c. When during the day do you go to the tap? (tick all that apply) ..........  
   □ Early morning  □ Early afternoon
   □ Mid morning   □ Late afternoon
   □ Noon         □ Night
   □ Don’t collect water

21. Does your house have electricity? .........................................................  
   □ Yes  □ No

22. Does your house have a toilet? ..............................................................  
   □ Yes  □ No

23. How many adults live in your house? (Excluding the participant) An adult is anyone 18 years or older. ....................

24. How many children live in your house? (Excluding the participant) A child is anyone less than 18 years of age. .........

Version: 1.1 25 April 2006
25. Have you drunk any kind of alcohol in the last 12 months?  

25a. If yes, how many drinks have you had in the last week?  

25b. In the last twelve months, have you been to a shebeen (tavern)?  

25c. If yes, how many times have you been to a shebeen in the last week?  

25d. Which taverns do you visit? (List of taverns to be updated every 6 months)  

INTERVIEWER: READ EACH OPTION TO PARTICIPANT

<table>
<thead>
<tr>
<th>In last 12m</th>
<th>Ever</th>
<th>In last 12m</th>
<th>Ever</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beja</td>
<td></td>
<td>Q&amp;J</td>
<td></td>
</tr>
<tr>
<td>D.D</td>
<td></td>
<td>Rasta</td>
<td></td>
</tr>
<tr>
<td>Hagwini</td>
<td></td>
<td>Sinethemba</td>
<td></td>
</tr>
<tr>
<td>Khaya</td>
<td></td>
<td>Solly</td>
<td></td>
</tr>
<tr>
<td>Kuphumleni</td>
<td></td>
<td>Take it Easy</td>
<td></td>
</tr>
<tr>
<td>Mancanye</td>
<td></td>
<td>Terra</td>
<td></td>
</tr>
<tr>
<td>Mavuyie Spaza</td>
<td></td>
<td>Tosh</td>
<td></td>
</tr>
<tr>
<td>Mazizini</td>
<td></td>
<td>Tshepang</td>
<td></td>
</tr>
<tr>
<td>Mniqini</td>
<td></td>
<td>Wendy</td>
<td></td>
</tr>
<tr>
<td>Nana’s</td>
<td></td>
<td>Ziyaqgobha</td>
<td></td>
</tr>
<tr>
<td>Ndulini</td>
<td></td>
<td>Other</td>
<td></td>
</tr>
</tbody>
</table>

Specify: ________________________________
26. Have you smoked a cigarette or tobacco in the last six months?  
   [ ] yes  [ ] no  
   If NO skip to Q27

26a. If yes, how many bought cigarettes have you smoked in the last week?  
   [ ]

26b. If yes, how many rolled tobacco cigarettes have you smoked in the last week?  
   [ ]

27. Have you smoked any recreational drugs in the last six months?  
   (Recreational drugs include Mandrax, Dagga, Tik)  
   [ ] yes  [ ] no  
   If NO skip to Q28

27a. If yes, how many times in the last month have you smoked a recreational drug?  
   [ ]

28. Have you ever worked in a mine?  
   [ ] yes  [ ] no  
   If NO skip to Q29

28a. If yes, what kind of mine was it?  
   [ ] Gold  [ ] Stone/quarry  
   [ ] Asbestos  [ ] Other  
   [ ] Coal

28b. If yes, for how many years in total did you work in the mine?  
   [ ]

29. Are you a health care worker who may come into contact with TB patients or TB suspects?  
   [ ] yes  [ ] no

30. Have you been in prison in the last 6 months?  
   [ ] yes  [ ] no

31. Has anyone who was living in your house died in the last 2 years?  
   [ ] yes  [ ] no  
   If NO then skip to Q32

31a. If yes, do you know the cause of death?  
   [ ] yes  [ ] no  
   If NO then skip to Q32

31b. If yes, what was the cause of death?
Interviewer prompt: Now I am going to ask you some short questions about your sexual activity. Remember, everything that you say is completely confidential. It is important that you answer accurately.

32. How many sexual partners have you had in the last six months? ........................................

33. How many new sexual partners have you had in the last six months? (Please note that this total must be equal or less than the total specified for Question 32) ........................................

34. In the last 6 months, have you used a condom? . . . .

35. Did you use a condom at your last sexual encounter? ........................................

36. Have you been treated for a sexually transmitted infection in the last six months? ........................................

37. Have you worked at any job in the last 12 months? . . . .

37a. If yes, what is your job? (Participant can name more than one) ........................................

37b. If yes, where is your job? (specify location(s)) ........................................

37c. How do you get to work? (Interviewer: read each option; tick all that apply)

- Walk
- Taxi
- Private Car
- Train
- Bus
- Other, specify: ........................................

Version: 1. 1 25 April 2006
38. Have you attended school in the last 12 months? ............... [ ] yes [ ] no If NO skip to Q39

38a. If yes, which school do you attend? ............... [ ] Ukhanyo School
[ ] Masiphumelele High School
[ ] False Bay High School
[ ] Ocean View High School
[ ] Other, specify: __________________________

38b. What grade are you in? .......................

39. Have you attended church in the last 12 months? ............... [ ] yes [ ] no If NO skip to Q40

39a. If yes, which churches have you attended? ............... [ ] Masiphumelele Methodist Church
[ ] Masiphumelele Baptist Church
[ ] Universal Church
[ ] Catholic Church
[ ] Jehovah Witnesses of God
[ ] King of Kings Baptist Church
[ ] Other, specify: __________________________

40. Where do you regularly (every day, week or month) buy your groceries and do other shopping?

[ ] Sunvalley Pick ’n Pay
[ ] Fishhoek Shoprite
[ ] Ocean view: __________________________ (which shop)
[ ] Spaza shop in Masi: __________________________
[ ] Spaza shop outside Masi: __________________________
[ ] Other, specify: __________________________
41. How do you get to the shops?  
   - Walk  
   - Bus  
   - Train  
   - Private car  
   - Taxi  
   - Other, specify:  

42. Have you been to a clinic or hospital in the last 12 months for any reason other than your TB treatment?  
   - Yes  
   - No  
   If No skip to Q43  

42a. If yes, which clinic/hospital have you been to in the last 12 months?  
   - Masi Clinic  
   - Other Clinic specify:  
   - False Bay Hospital  
   - Other Hospital specify:  

42b. If yes, what day of the week do you attend regularly?  
   - Monday  
   - Tuesday  
   - Wednesday  
   - Thursday  
   - Friday  
   - No set day of the week  

43. Have you travelled to the Eastern Cape in the previous 12 months?  
   - Yes  
   - No  
   If No skip to Q44  

43a. If yes, when did you go? (Specify the month)  

43b. If yes, where specifically did you go?  

44. Do you take a taxi regularly (daily, weekly or monthly) around Cape Town?  
   - Yes  
   - No  
   If No skip to Q45  

44a. If yes, how many times in the last month have you been in a taxi?  

44b. If yes, where do you catch the taxi from?  

44c. If yes, where do you go?  

45. Do you meet regularly with any other group in Masiphumelele?  
   - Yes  
   - No  
   Interviewer: Examples of other groups are Yizani, Sosabenza, Vuka mama or any other organisation.  
   - Specify  

Version: 1 1 25 April 2006
**Adult and Adolescent Questionnaire**

**Please Initial and date the appropriate section below:**

<table>
<thead>
<tr>
<th>1st Review:</th>
<th><em><strong>/</strong></em>/20___</th>
<th>Faxed by:</th>
<th><em><strong>/</strong></em>/20___</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st Review:</td>
<td><em><strong>/</strong></em>/20___</td>
<td>Faxed by:</td>
<td><em><strong>/</strong></em>/20___</td>
</tr>
<tr>
<td>1st Review:</td>
<td><em><strong>/</strong></em>/20___</td>
<td>Faxed by:</td>
<td><em><strong>/</strong></em>/20___</td>
</tr>
<tr>
<td>1st Review:</td>
<td><em><strong>/</strong></em>/20___</td>
<td>Faxed by:</td>
<td><em><strong>/</strong></em>/20___</td>
</tr>
</tbody>
</table>
FHS016: Annual Progress Report / Renewal

This serves as notification of annual approval, including any documentation described below.

☐ Approved
☐ Not approved

Annual progress report
See attached comments

Approved until/next renewal date 30.8.2015

Signature Chairperson of the HREC

Date Signed 11/8/2014

Comments to PI from the HREC

Principal Investigator to complete the following:

1. Protocol information

<table>
<thead>
<tr>
<th>Date form submitted</th>
<th>06 August 2014</th>
</tr>
</thead>
<tbody>
<tr>
<td>HREC REF Number</td>
<td>321/2005</td>
</tr>
<tr>
<td></td>
<td>Current Ethics Approval was granted until 28 Aug 2014</td>
</tr>
<tr>
<td>Protocol title</td>
<td>A Study of the Effects of Antiretroviral Therapy on Rates and Transmission of Tuberculosis</td>
</tr>
<tr>
<td>Protocol number (if applicable)</td>
<td>3.0 Dated 03 August 2011</td>
</tr>
</tbody>
</table>

Are there any sub-studies linked to this study? ☐ Yes √ No

If yes, could you please provide the HREC Ref’s for all sub-studies? Note: A separate FHS016 must be submitted for each sub-study.

Principal Investigator: LG Bekker
Department / Office: Desmond Tutu HIV Centre
Internal Mail Address: Werner and Beit North Rm1.21.3

1.1 Does this protocol receive US Federal funding? ☐ Yes √ No

1.2 If the study receives US Federal Funding, does the annual report require full committee approval? ☐ Yes √ No

(Nota: Please complete the Closure form (FHS010) if the study is completed within the approval period)
Form FHS006: Protocol Amendment

HREC office use only (FWA00001637; IRB00001938)

☑ Approved  ☐ Type of review: Expedited  ☐ Full committee

This serves as notification that all changes and documentation described below are approved.

Signature Chairperson of the HREC  
Date 11/8/2014

Note: All major amendments should include a PI Synopsis justifying the changes for the amendment (please see notice dated 23 April 2012)

Principal Investigator to complete the following:

1. Protocol information

<table>
<thead>
<tr>
<th>Date (when submitting this form)</th>
<th>08 August 2014</th>
</tr>
</thead>
<tbody>
<tr>
<td>HREC REF Number</td>
<td>321/2005</td>
</tr>
<tr>
<td>Protocol title</td>
<td>A Study of the Effects of Antiretroviral Therapy on Rates and Transmission of Tuberculosis</td>
</tr>
<tr>
<td>Protocol number (if applicable)</td>
<td>3.0 Dated 03 August 2011</td>
</tr>
<tr>
<td>Principal Investigator</td>
<td>Prof LG Bekker</td>
</tr>
<tr>
<td>Department / Office</td>
<td>Desmond Tutu HIV Centre</td>
</tr>
<tr>
<td>Internal Mail Address</td>
<td>Werner and Beit North Rm1.21.3</td>
</tr>
</tbody>
</table>

1.1 Is this a major or a minor amendment? (see FHS006hlp)

<table>
<thead>
<tr>
<th>Major (tick box)</th>
<th>Minor (tick box)</th>
</tr>
</thead>
<tbody>
<tr>
<td>☐ Major</td>
<td>✓ Minor</td>
</tr>
</tbody>
</table>

1.2 Does this protocol receive US Federal funding?

<table>
<thead>
<tr>
<th>Yes (tick box)</th>
<th>No (tick box)</th>
</tr>
</thead>
<tbody>
<tr>
<td>☐ Yes</td>
<td>✓ No</td>
</tr>
</tbody>
</table>

1.3 If the amendment is a major amendment and receives US Federal Funding, does the amendment require full committee approval?

<table>
<thead>
<tr>
<th>Yes (tick box)</th>
<th>No (tick box)</th>
</tr>
</thead>
<tbody>
<tr>
<td>☐ Yes</td>
<td>✓ No</td>
</tr>
</tbody>
</table>

2. List of Proposed Amendments with Revised Version Numbers and Dates

Please itemise on the page below, all amendments with revised version numbers and dates, which need approval.

This page will be detached, signed and returned to the PI as notification of approval. Please add extra pages if necessary.

Informed Consent forms:

- Informed Consent for Adults (Version 3.2 Dated 08 Aug 2014)
- Informed Consent for Adult with Dependent Child (Version 3.2 Dated 08 Aug 2014)
- Child Assent Form (Version 3.2 Dated 08 Aug 2014)
14 July 2015

HREC REF: 470/2015

Dr K Middelkoop
Desmond Tutu HIV Foundation
IIDMM

Dear Dr Middelkoop

PROJECT TITLE: INVESTIGATING THE SOCIO-DEMOGRAPHIC AND EPIDEMIOLOGICAL RISK FACTORS ASSOCIATED WITH TB TRANSMISSION IN A HIGH TB AND HIV BURDENED COMMUNITY IN CAPE TOWN, SOUTH AFRICA (Masters Candidate – Ms R Tadokera) Sub-study linked to 321/2005

Thank you for submitting your study to the Faculty of Health Sciences Human Research Ethics Committee (HREC) for review.

It is a pleasure to inform you that the HREC has formally approved the above-mentioned study.

Approval is granted for one year until the 30th July 2016.

Please submit a progress form, using the standardised Annual Report Form if the study continues beyond the approval period. Please submit a Standard Closure form if the study is completed within the approval period.

(Forms can be found on our website: www.health.uct.ac.za/fhs/research/humanethics/forms)

We acknowledge that the following student:- Rabecca Tadokera is also involved in this project

Please note that the on-going ethical conduct of the study remains the responsibility of the principal investigator.

Please quote the HREC REF in all your correspondence.

Yours sincerely

[Signature]

PROFESSOR M BLOCKMAN
CHAIRPERSON, FHS HUMAN RESEARCH ETHICS COMMITTEE
Federal Wide Assurance Number: FWA00001637.
Institutional Review Board (IRB) number: IRB00001938
This serves to confirm that the University of Cape Town Research Ethics Committee complies to the Ethics Standards for Clinical Research with a new drug in patients, based on the Medical Research Council (MRC-SA), Food and Drug Administration (FDA-USA), International Convention on Harmonisation Good Clinical Practice (ICH GCP) and Declaration of Helsinki guidelines.

The Research Ethics Committee granting this approval is in compliance with the ICH Harmonised Tripartite Guidelines E6: Note for Guidance on Good Clinical Practice (CPMP/ICH/135/95) and FDA Code Federal Regulation Part 50, 56 and 312.
GENERAL INFORMATION

The *Journal of Infectious Diseases* (JID) is sponsored by the Infectious Diseases Society of America. Reports of research related to any aspect of the fields of microbiology, infection, and host response, whether laboratory, clinical, or epidemiologic, are considered for publication in JID. Major Articles and Brief Reports are peer reviewed; Correspondence is reviewed by the Editors.

*JID* reserves the right to edit for journal style, clarity, precision of expression, and grammar. Authors review these changes at the proof stage but must limit their alterations in proof to correction of errors and clarification of misleading statements.

**Manuscript Categories**

All manuscripts—Major Articles, Brief Reports, Correspondence, Perspectives, Editorials, Reviews, and Supplement articles—must have conflict of interest and funding statements.

**Major Articles** describe original investigations that are an important advance in the field and that have been brought to an acceptable degree of completion. Major Articles must be no longer than 3500 words of text (from the beginning of the introduction to the end of the discussion; do not count the abstract or the references), and illustrations must be limited to the minimum necessary for clear and concise presentation. For Major Articles describing results of clinical trials (see “Clinical trials registration,” below), the abstract must be structured, using the headings Background, Methods, Results, and Conclusions, and must be no more than 200 words; for other Major Article manuscripts, the abstract may be structured (200-word limit) or unstructured (150-word limit). Major Articles are limited to a maximum of 7 inserts (tables and figures combined) and 50 references.

**Brief Reports** present complete studies that are narrower in scope than those described in Major Articles or that represent new developments. Manuscripts that are descriptive or primarily methodologic in nature, that report results of phase I and II vaccine trials (see “Clinical trials registration,” below), or that describe in vitro chemotherapeutic studies should, in general, be submitted as Brief Reports. Brief Reports include an abstract (no more than 100 words) and are limited to no more than 2000 words of text, a total of 2 inserts (tables or figures), and 15 references.

**Correspondence (letters)** must be submitted in reference to a previous publication in JID (within the preceding 12 months); otherwise they will not be considered. Please prepare the letter in manuscript format, including a title page. The letter cannot exceed 750 words of text, 1 insert (table or figure), and 10 references.

**Perspectives and Editorials** are generally invited by the Editor and are overviews of articles in JID or of other research in infectious diseases. Unsolicited perspectives are also considered.

**Reviews** that are research oriented will be considered by JID. Authors should contact the Editor in advance of submission, to determine whether a specific topic is deemed appropriate and timely. Review articles will be peer reviewed.

**Supplements**

Supplements are published by JID. Requirements for supplement manuscripts follow those for JID manuscripts (e.g., cover letter, conflict of interest and funding statements, list of potential reviewers). Inquiries related to suitability of topic, program organization, and production should be made in writing to the Editor.

**Notice of Page Charges**

Page charges for regular issues are assessed as follows: $55 per page for the first 6 pages and $85 for each additional page. Authors are also charged for the print reproduction of color figures: the cost is $500 for the first page and $400 for each additional page. If 2 or more figures can be placed on a single printed page, the cost for that page is $500. Authors should state in the cover letter whether they will bear the cost of reproducing their color figures or whether they prefer to have them published in black and white at no additional cost. The publisher will bill the author concurrently for page and color charges and for reprints. Invited Perspectives, Editorials, and Correspondence related to articles recently published in *JID* will not be subject to page charges. In unusual circumstances, the Editor may waive page charges.

**Online-Only Supplemental Content**

Designation of supplemental content as online only is reserved for content that cannot be printed because of space or technical limitations, such as large data sets that cannot easily be displayed in printed form. The editors reserve the right to limit the size and content of file submissions.

Content that has been published previously (print or online) is not appropriate for posting only online but rather should be referenced in the manuscript text. Other reasons that are insufficient for designating content as online only are to avoid paying charges for color figures ($500 for the first color page, $400 for each subsequent page) or to circumvent journal limits for inserts.
(a maximum of 7 inserts for Major Articles and 2 for Brief Reports) by providing additional figures or tables and designating them as online only.

Any material that is intended for publication only online must be submitted for review, along with the manuscript, via the JID Editorial Manager system. Software required to view or use the material should be embedded in the file or easily accessible.

The editor will make the final decision whether to publish (i.e., post online) the online-only material with the article. Acceptance of the article for publication does not necessitate acceptance of the online-only material. Accepted online-only material will remain associated with the full article and will not be modified after publication.

**DOCUMENT REQUIREMENTS**

**Checklist**

Your manuscript will be returned if you do not

1. Specify the type of article and adhere to the following limits:
   - Major Article: 3500 words (*including* tables and figure legends), 50 references, 7 figures/tables
   - Brief Report: 2000 words (*including* tables and figure legends), 15 references, 2 figures/tables
2. Include a cover letter with the following information:
   - A statement that the manuscript has not been submitted or accepted elsewhere
   - A statement that all authors have contributed to, seen, and approved the final, submitted version of the manuscript
   - A list of 5 potential reviewers, with their e-mail addresses
3. All file names—for manuscript, cover letter, figures, tables—should contain no spaces between numbers or letters. File names may be run together (e.g., authornamearticletitleversion1) or contain punctuation such as underscores, hyphens, or both (e.g., author_name_article_title_version_1 or author_name_article_title_version_1). **Manuscript Preparation**

   *JID* complies with the Uniform Requirements for Manuscripts Submitted to Biomedical Journals (*Ann Intern Med* 2000; 133: 229–31 [editorial]; for the complete text, see http://www.icmje.org), except that reference citations should appear in the text in square brackets (not parentheses). Text, tables, references, and legends must be double spaced. Italics should be used for genus and species names and for genes, but not for in vivo, in vitro, in situ, et al., or other Latin-derived expressions listed in Webster’s Collegiate Dictionary; see a recent issue of *JID* for appropriate style.

   **Title page.** On the title page, please supply a running head of not more than 40 characters and spaces, a title of not more than 160 characters and spaces, the names and affiliations of all of the authors, and word counts of the abstract and the text. Each author’s full name must be used. If there is potential confusion with respect to whether the first name presented is actually the last name of the author, please identify the last name.

   **Footnote page.** Footnotes must include (1) a statement that the authors either have or do not have a commercial or other association that might pose a conflict of interest (e.g., pharmaceutical stock ownership, consultancy, advisory board membership, relevant patents, or research funding); (2) a statement naming sources of financial support (including grant numbers); (3) the name, date (month and year), and location (city, state, and country if not USA) of a meeting at which all or part of the information has been presented (include an abstract number if possible); (4) the name, address, telephone and fax numbers, and e-mail address of the person to whom correspondence and requests for reprints should be addressed; and (5) current affiliations and addresses for authors whose affiliations have changed since completion of the study.

**INSTRUCTIONS FOR AUTHORS**
Conflict of interest. There is a potential conflict of interest when anyone involved in the publication process has a financial or other beneficial interest in the products or concepts mentioned in a submitted manuscript or in competing products that might bias his or her judgment. A footnote in the manuscript must include a statement either that the authors have or do not have an association that might pose a conflict of interest (e.g., pharmaceutical stock ownership, consultancy, advisory board membership, relevant patents, or research funding). The name of each contributing author and any potential conflict of interest for each author for the previous 2 years should be listed.

Examples:

- Potential conflicts of interest. S.A. and K.H. are shareholders in Loke Diagnostics (Aarhus, Denmark).
- Potential conflicts of interest. The author has served as consultant to and has received research grants from all of the manufacturers of the lipid formulations of amphotericin.
- Potential conflicts of interest. E.H.P. has been a consultant to Fujisawa Healthcare, Inc., Gilead Sciences, Novartis, and GlaxoSmithKline and is a member of the speakers’ bureaus for Pharmacia and Novartis. J.A.S. has received research funding from Bayer and Pharmacia, has been a consultant for Bayer and Pfizer, and has been on the speakers’ bureau for Pfizer and Ortho McNeil.

If no potential conflict exists, the phrase “No conflict” should appear after the author’s name.

The online submission system will require the corresponding author to certify that all potential conflicts of interest have been disclosed. If the article is accepted for publication, the disclosures will be published in the footnotes section.

Abstract. The abstract for a Major Article describing results of a clinical trial must be no more than 200 words and must be structured with the headings Background, Methods, Results, and Conclusions. The trial must be registered (see “Clinical trials registration”), and the abstract must include the registry’s URL and the trial’s registration number. Abstracts of other Major Articles may be structured (200-word limit) or unstructured (150-word limit). Abstracts of Brief Reports should be no more than 100 words. Whether structured or unstructured, the abstract must state the purpose of the research, the methods used, the results, and the conclusions. Do not cite references in the abstract. Include 3 to 10 key words, separate from the abstract. Authors are reminded that the abstract is of particular value to producers and users of online literature retrieval systems such as MEDLINE.

Text. The text of Major Articles must be no longer than 3500 words, and that of Brief Reports no longer than 2000 words. The Methods section must include a statement that informed consent was obtained from patients or their parents or guardians and that human experimentation guidelines of the US Department of Health and Human Services and/or those of the authors’ institution(s) were followed in the conduct of clinical research or that animal experimentation guidelines were followed in animal studies.

References. Major Articles are limited to 50 references, and Brief Reports to 15 references. Only works that have been published or accepted for publication can be included in the reference list. Unpublished observations by the authors (authors’ unpublished data), personal communications (J. L. Searle, personal communication), and manuscripts submitted for publication (H. Chapin and M. Perkins, submitted) should be mentioned parenthetically in the text. Please number references in order of appearance; those cited only or first in tables or figures are numbered according to the order in which the table or figure is cited in the text. Example: If table 2 is cited in the text after reference 25, a new reference cited in table 2 will be reference 26.

References must follow the National Library of Medicine format as used in MEDLINE and Uniform Requirements (see above). Provide all authors’ (or editors’) names when there are fewer than 7; for 7 or more, list the first 3 and add “et al.” Titles of journals not listed in Index Medicus should be spelled out in full. Reference to a doctoral dissertation should include the author, title, institution, location, year, and publication information, if published. For online resources, include a URL and date accessed. Accuracy of references is the responsibility of the authors.

Examples of the proper format are as follows:

Journal article

Book chapter

Conference program

Internet site
Acknowledgment(s). The page preceding the references can include a statement thanking those who assisted substantially with work relevant to the study.

Statistical analysis. The statistical analyses used should be identified both in the text and in all tables and figures where the results of statistical comparison are shown.

Units of measure. All data should be expressed in metric units; use of SI units is encouraged. Use °C for temperature.

Tables and figures. Major Articles are limited to a maximum of 7 inserts (tables and figures combined), and Brief Reports to a maximum of 2 inserts. A single insert should not contain both tables and figures; possible exceptions are survival plots or time-to-event outcomes and pooled data in meta-analyses or other analyses that combine data from individual studies.

Data should not be repeated in both a table and a figure. Abbreviations and acronyms used in tables and figures must be explained in the table footnotes and figure legends, even if already defined in the text.

Tables should be numbered in the order of mention in the text. Tables should be typed double spaced throughout, with no vertical or internal rules. Footnotes and accompanying explanatory material should be kept to a minimum. Footnotes should be placed below the table and designated by superscript lowercase letters (listed in order of location when the table is read horizontally). Each column must have an appropriate heading describing the data in the column below, and units of measure must be clearly indicated. For further instructions on the preparation of tables in Word or WordPerfect, consult the Guidelines for Tables (available online at http://www.journals.uchicago.edu/page/jid/msprep-tables.html).

Figures should be also numbered in the order of mention in the text. Please prepare your figures in accordance with the Guidelines for Submission of Artwork (available online at http://www.journals.uchicago.edu/page/jid/info/artguide/artguide.html).

Falsified data or images: Should it be determined that data or images in a manuscript have been deliberately falsified, the manuscript will be rejected and the corresponding author’s superior(s) will be notified.


For commercially obtained products mentioned in the text, list full names of manufacturers. Generic names of drugs and other chemical compounds should be used.

CONSORT guidelines. Authors of manuscripts describing results of randomized, controlled trials, whether in Major Articles or Brief Reports, should consult guidelines recommended in the revised CONSORT statement (available online at http://www.consort-statement.org).

Clinical trials registration. For their results to be considered for publication in JID, all clinical trials must be included in a registry that is electronically accessible to the public free of charge. In common with the International Committee of Medical Journal Editors (ICMJE), JID defines a clinical trial as “any research project that prospectively assigns human subjects to intervention and concurrently assigned comparison groups to study the cause-and-effect relationship between a medical intervention and a health outcome.” Studies designed for other purposes—for example, phase I trials designed to evaluate pharmacokinetics or major toxicities—are exempt.

Registration should occur before patient enrollment. The corresponding author must provide the registry’s URL and the trial’s registration number at the end of the manuscript’s abstract. Information on acceptable registries may be obtained at the ICMJE Web site (http://www.icmje.org/). The National Library of Medicine’s registry (http://www.clinicaltrials.gov/) is free, is open to all investigators, and meets JID requirements.


Nucleotide sequences. Nucleotide sequences must be submitted to the International Nucleotide Sequence Databases prior to acceptance, and the accession numbers must appear in the final revision of the manuscript. For more information, please see the GenBank (http://www.ncbi.nlm.nih.gov/Genbank), EMBL Nucleotide Sequence Database (http://www.ebi.ac.uk/embl), and DNA Data Bank of Japan (http://www.ddbj.nig.ac.jp) Web sites.

Human genetic nomenclature and notation. For human genes, use genetic notation and symbols approved by the HUGO Gene Nomenclature Committee (see Wain HM, Bruford EA, Lovering RC, Lush MJ, Wright MW, Povey S. Guidelines for human gene nomenclature. Genomics 2002; 79:464–470). Human gene names and loci should be written in italicized capital letters and Arabic numerals. Human protein product names are not italicized. For genetic mutations in humans, use the notation approved by the Nomenclature Working Group (see Antonarakis SE. Recommendations for a nomenclature sys-
Human single-nucleotide polymorphisms. For human genes, newly described single-nucleotide polymorphisms (SNPs) should be submitted to an appropriate database, such as dbSNP (available online at http://www.ncbi.nlm.nih.gov/SNP), prior to submission of revised manuscripts. The identification numbers of previously recognized SNPs (rs numbers) or recently submitted SNPs (ss numbers) should be provided in the manuscript if the number of SNPs is small, or submitted as supplemental online material if the number is large.

Controls for genetically engineered mice. For all experiments involving genetically engineered mice, inbred-strain background effects have become an important concern. The Editorial Board generally considers only the following as appropriate controls: the parental wild-type inbred strain (if the engineered animals studied have been backcrossed onto this strain for >4 generations) or littermates. If an alternative control was used, an explanation must be included in the cover letter to the Editor.

Supporting Documents

Cover letter. All manuscripts submitted to JID, online or otherwise, must be accompanied by a letter declaring that the manuscript has not been submitted or accepted for publication elsewhere. This letter must warrant that all authors have seen and approved the content and have contributed significantly to the work. Authors should suggest potential unbiased reviewers who are qualified to review their manuscript. A cover letter must also accompany a revised submission and must address, point by point, issues raised in the review process.

Related manuscripts. If there appears to be significant overlap between a manuscript submitted to JID and another manuscript submitted (to JID or another journal) by the same authors or if there is an overlap between a manuscript submitted to JID and one published by JID or another journal, the editors will ask the corresponding author to respond to the question of overlap. On the basis of the response, the editors may choose to consider the manuscript further, ask for modification of the manuscript, or reject the manuscript; in the case of a rejected manuscript, the editors will provide the author(s) with a detailed explanation. The editors may also choose to take further action, which could include contacting the appropriate superior at the home institution and/or suspending publishing privileges in JID for a designated period of time. If it is found that the authors of an article published in JID had published an overlapping or identical manuscript in another journal, the editors will publish an announcement to that effect in JID.

Permissions. Written permission must be obtained from all investigators cited in a personal communication who are not coauthors of the present manuscript and from the copyright owner if a previously published table or figure is to be reproduced. Additionally, authors should include a copy of “in press” references and articles by the same authors on the same subject that were submitted elsewhere. All letters of permission and references/articles should be submitted as a separate file, in TIFF (.tif) or EPS (.eps) (not GIF [.gif] or JPEG [.jpg]) format.

If you submit a LaTeX file, all figures and tables should be external files that are referenced within the main “ms.tex” file by using appropriate commands. LaTeX is especially suited for articles containing extensive math; authors of math-intensive manuscripts can use the AAS TeX manuscript package available at http://www.journals.uchicago.edu/AAS/AAS TeX.

Please note that authors of accepted manuscripts may be required to submit high-resolution hard copies of all figures during production, as not all digital art files are usable.

In addition to the main manuscript file, submit your cover letter as a separate file in the same format as your main file. If you used any revision or editorial tracking tools in your word-

INSTRUCTIONS FOR AUTHORS
processing program, be sure the final version of your manuscript does not contain tracked changes.

File Compression and Archives
If you have more than two files to upload into the system (e.g., manuscript, figures, and cover letter), we recommend you combine these files into an archive, so you upload only a single file when submitting the manuscript. Applications for Mac OS (such as StuffIt) and Windows (such as WinZip) support the formats listed below.

The following archive formats may be used:

- Zip (e.g., “archive.zip”)
- Unix tar (e.g., “archive.tar” or compressed with a gzip as “archive.tar.gz” or “archive.tgz”)

If you cannot create archives in these formats, you may still compress individual files to shorten upload time:

- Unix compress (e.g., “ms.tex.Z”)
- gzip (e.g., “ms.tex.gz”)

Revised and Final Versions of Manuscripts
If you are submitting a revised manuscript, please include your responses to the reviewers’ comments as part of the cover letter file. When submitting a revised manuscript with figures, include all figures, even if they have not changed since the previous version. The final version of your manuscript must be submitted in Word (.doc.), Rich Text (.rtf), or LaTeX (.tex) format, because your keystrokes will be used in publication; a PDF does not contain usable character data and is thus not adequate. For both revised and final versions of manuscripts, please observe the same formatting instructions outlined above.

Online Submission Instructions
Please have the following items readily available before beginning the online submission process:

- Manuscript in an acceptable format as described above
- Cover letter as a separate file
- Information from title page (to be typed into the peer review database): title, short title, list of authors and affiliations, and contact information for the corresponding author
- Abstract of the manuscript (to be copied and pasted into a field in Editorial Manager)

Go to the JID Editorial Manager system at http://jid.edmgr.com to submit your manuscript. If you do not have access to the Web, please send one hard copy of your manuscript and a CD-R containing all relevant electronic files to the editorial office. It is no longer necessary to submit a hard copy in addition to an electronic submission.