Heterogeneity in Tuberculosis infectiousness: understanding its impact on epidemiology and control

Yayehirad Alemu Melsew

BSc (Medical Laboratory Science)

MPH (Epidemiology and Biostatistics)

A Thesis Submitted for the Degree of Doctor of Philosophy

at

Monash University

in 2020

Department of Epidemiology & Preventive Medicine

School of Public Health & Preventive Medicine

Copyright Notice

Table of Contents

Contents

Abstract

Background

Tuberculosis (TB) is one of the top ten causes of death worldwide, responsible for 10.0 million cases and 1.2 million deaths in 2018. Heterogeneity in infectiousness is particularly essential to *Mycobacterium tuberculosis* transmission since TB has profoundly heterogeneous clinical manifestations and variable duration of disease. This heterogeneity of transmission has the potential to disrupt global TB elimination strategies, which are often applied consistently across areas and populations, making it essential to understand and quantify.

Aims

The overall aim of this thesis was to establish an evidence base concerning TB patients' infectiousness heterogeneity and to integrate this evidence with a mathematical TB transmission model to predict the effectiveness of targeted interventions.

Methods

Systematic reviews and meta-analyses of published articles were undertaken to identify risk factors associated with the infectiousness of patients with TB and explore past TB transmission models. Offspring distributions for the number of secondary infections per index patients were produced using TB contact investigation data from Victoria, a state in Australia with a low burden of TB and high-quality data. This empiric distribution was then fitted with a negative binomial distribution to quantify heterogeneity and super-spreading. Survival analyses were employed to describe profiles and patterns of TB disease progression among contacts of TB patients. Finally, a transmission dynamic model TB was constructed using ordinary differential equations to evaluate the comparative effectiveness of targeted active case finding intervention towards the most infectious type of patients. Model parameters were estimated from the observed distribution of secondary infections with a Hamiltonian Monte

Carlo algorithm, with a particular focus on parameters pertaining to heterogeneous infectiousness.

Results

The systematic review showed that patient characteristics significantly associated with higher infectiousness included clinical characteristics such as treatment delay, positive sputum smear, cavitary disease and HIV seronegative status; and behavioural characteristics such as smoking and heavy alcohol consumption. After assessing the distribution of the number of secondary infections per index from contact tracing data, super-spreaders were estimated to be around 10% of all index patients, which were responsible for 75.2% of secondary infections. The large majority of past TB models did not consider heterogeneous infectiousness at all, and among those that did incorporate heterogeneous infectiousness, the commonest approach to stratification of the active TB compartment was to incorporate two levels of infectiousness. However, various clinical characteristics were used to stratify the active TB compartments, and models differed as to whether they permitted transition between these states. The proposed model showed that heterogeneity parameters such as the proportion of super-spreaders and their relative infectiousness compared to other low-spreaders were as important as other previously recognised epidemiological parameters, such as the latency progression parameters, in determining the burden of TB. This model also suggested that targeted active case finding interventions directed towards people likely to be super-spreaders can have a substantial impact on TB burden, particularly in settings with high transmission but low case detection rates.

Conclusions

Heterogeneity of transmission and super-spreading are critical issues to consider in the design of interventions and models of TB transmission dynamics. Behavioural and clinical

characteristics of TB patients can be used to identify highly infectious patients for targeted contact tracing interventions. Targeting active case-finding interventions to people likely to be super-spreaders is particularly beneficial in settings where case detection is poor.

Declaration

This thesis is an original work of my research and contains no material which has been accepted for the award of any other degree or diploma at any university or equivalent institution and that, to the best of my knowledge and belief, this thesis contains no material previously published or written by another person, except where due reference is made in the text of the thesis.

Signature:

Print Name: Yayehirad Alemu Melsew

Date: 12th March 2020

Publications during enrolment

Publications comprising part of the thesis

1. **Melsew YA**, Doan TN, Gambhir M, Cheng AC, McBryde E, Trauer JM. Risk factors for infectiousness of patients with tuberculosis: a systematic review and meta-analysis. *Epidemiology and Infection*. 2018 Feb;146(3):345-53.
2. **Melsew YA**, Gambhir M, Cheng AC, McBryde ES, Denholm JT, Tay EL, Trauer JM. The role of super-spreading events in *Mycobacterium tuberculosis* transmission: evidence from contact tracing. *BMC Infectious Diseases*. 2019 Dec;19(1):244.
3. **Melsew YA**, Cheng AC, McBryde ES, Denholm JT, Tay E, Ragonnet R, Trauer JM. Profiles of tuberculosis disease activation among contacts of patients with tuberculosis. *European Respiratory Journal*. 2019 Jan 1:1900353.
4. **Melsew YA**, Adekunle AI, Cheng AC, McBryde ES, Ragonnet R, Trauer JM. Heterogeneous infectiousness in mathematical models of tuberculosis: a systematic review. *Epidemics*. 2019 Oct 17:100374.

Publications relevant to the thesis but not comprising part of it

1. Trauer JM, Dodd PJ, Gomes MG, Gomez GB, Houben RM, McBryde ES, **Melsew YA**, Menzies NA, Arinaminpathy N, Shrestha S, Dowdy DW. The importance of heterogeneity to the epidemiology of tuberculosis. *Clinical Infectious Diseases*. 2018 Nov 1;69(1):159-66.

Additional publications during PhD candidature

1. Vogt F, Mengesha B, Asmamaw H, Mekonnen T, Fikre H, Takele Y, Adem E, Mohammed R, Ritmeijer K, Adriaensen W, **Melsew Y**, Griensven JV, Diro E. Antigen Detection in Urine for Noninvasive Diagnosis and Treatment Monitoring of

Visceral Leishmaniasis in Human Immunodeficiency Virus Coinfected Patients: An Exploratory Analysis from Ethiopia. T*he American journal of tropical medicine and hygiene.* 2018 Oct 3;99(4):957-66.

2. Geberselassie SB, Abebe SM, **Melsew YA**, Mutuku SM, Wassie MM. Prevalence of stunting and its associated factors among children 6-59 months of age in Libo-Kemekem district, Northwest Ethiopia; A community based cross sectional study. *PloS one*. 2018 May 3;13(5):e0195361.
3. Mekonnen TH, Tefera MA, **Melsew YA**. Sick at work: prevalence and determinants among healthcare workers, western Ethiopia: an institution based cross-sectional study. *Annals of occupational and environmental medicine*. 2018 Dec;30(1):2.
4. Chekol DA, Biks GA, Gelaw YA, **Melsew YA**. Exclusive breastfeeding and mothers' employment status in Gondar town, Northwest Ethiopia: a comparative cross-sectional study. *International breastfeeding journal*. 2017 Dec;12(1):27.
5. Jember G, **Melsew YA**, Fisseha B, Sany K, Gelaw AY, Janakiraman B. Peripheral Sensory Neuropathy and associated factors among adult diabetes mellitus patients in Bahr Dar, Ethiopia. *Journal of Diabetes & Metabolic Disorders*. 2017 Apr 4;16(1):16.
6. Zeleke AJ, **Melsew YA**. Seroprevalence of Toxoplasma gondii and associated risk factors among HIV-infected women within reproductive age group at Mizan Aman General Hospital, Southwest Ethiopia: a cross sectional study. *BMC research notes*. 2017 Dec;10(1):70.
7. Kifle D, Azale T, Gelaw YA, **Melsew YA**. Maternal health care service seeking behaviors and associated factors among women in rural Haramaya District, Eastern Ethiopia: a triangulated community-based cross-sectional study. *Reproductive Health*. 2017 Jan 13;14(1):6.

International/national conference abstracts/presentations

1. Melsew YA, Gambhir M, Cheng AC, McBryde ES, Justin T Denholm JT, Tay E, Trauer JM. The role of super-spreading events in Mycobacterium tuberculosis transmission: Evidence from contact tracing. *PRISM^2 International Conference*, August 20-24, Palm Cove, Australia.
2. Melsew YA, Ragonnet R, Cheng AC, McBryde ES, Trauer JM. Capturing heterogeneous infectiousness in transmission dynamic models of tuberculosis: a compartmental modelling approach. *PRISM^2 annual conference*, Melbourne 2019
3. Melsew YA, Ragonnet R, Cheng AC, McBryde ES, Trauer JM. Capturing heterogeneous infectiousness in transmission dynamic models of tuberculosis: a compartmental modelling approach. *Epidemics 7th International Conference on Infectious Disease Dynamics*, 3-6 December 2019, Charleston, SC, USA.

Scholarships

2016-2019	Monash University Graduate Scholarship (MGS) Monash University International Postgraduate Research Scholarship (MIPRS)
2018	National Institute of General Medical Sciences. Tuition scholarship for the "Social Network Analysis" course at Columbia University, (US$1,000)
2018	National Institute of General Medical Sciences. Tuition scholarship for the "Agent Based Models for Population Health" course Columbia University online (US$1,000)

Awards

2018	PRISM^2 CRE Postdoc and PhD Student Funding Scheme travel grant to present at International Conference Palm Cove, QLD, Australia, August 2018.
2019	PRISM^2 CRE Postdoc and PhD Student Funding Scheme travel grant research visit at John Hopkins University Bloomberg School of Public Health, August 2019.
2019	TB MAC conference funding travel grant to present at Epidemics Conference on Infectious Disease Dynamics in the USA 2019

Professional roles during candidature

2018 – Present	Teaching Associate/Tutor, School of Public Health and Preventive Medicine, Monash University, Australia.
2016 – Present:	Peer reviewer: *BMC Archives of Public Health*, *Hindawi Journal of Nutrition and Metabolism, GERMS, Ethiopian Journal of Health Sciences*

Thesis including published works declaration

I hereby declare that this thesis contains no material which has been accepted for the award of any other degree or diploma at any university or equivalent institution and that, to the best of my knowledge and belief, this thesis contains no material previously published or written by another person, except where due reference is made in the text of the thesis.

This thesis includes *four* original papers published in peer-reviewed journals and *one* submitted publication. The core theme of the thesis is to understand heterogeneous infectiousness among patients with TB and its impact on *Mtb* transmission. The ideas, development and writing up of all the papers in the thesis were the principal responsibility of myself, the student, working within the School of Public Health and Preventive Medicine, Department of Epidemiology and Preventive Medicine under the supervision of Dr James M Trauer, Professor Allen C Cheng, Professor Emma McBryde and Dr Romain Ragonnet.

The inclusion of co-authors reflects the fact that the work came from active collaboration between researchers and acknowledges input into team-based research.
In the case of Chapter 2-6 included in this thesis, my contribution to the work involved the following:

Thesis Chapter	**Publication Title**	**Status (***published, in press, accepted or returned for revision, submitted***)**	**Nature and % of student contribution**	**Co-author name(s) Nature and % of Co-author's contribution***	**Co-author(s), Monash student Y/N***
2	Risk factors for infectiousness of patients with tuberculosis: a systematic review and meta-analysis	Published *Epidemiology and Infection 2018*	70%, literature search, data extraction, analysis and interpretation, conceptualisation and writing of manuscript, critical revision,	1. James Trauer, *providing guidance, ideas, reviewing* and *input into manuscript 10%,* 2. Tan Doan, *data extraction and*	No for all

			approval of final draft for publication	*input into manuscript 7%*, 3. Manoj Gambhir, *reviewing and input into manuscript 4%*, 4. Allen Cheng, r*eviewing and input into manuscript 5%*, 5. Emma McBryde, *reviewing and input into manuscript 4%*,	
3	The role of super-spreading events in Mycobacterium tuberculosis transmission: evidence from contact tracing	Published *BMC Infectious Diseases 2019*	70%, acquisition of data, data management, analysis and interpretation of data, conceptualisation and writing of manuscript, critical revision, approval of final draft for publication	1. James Trauer, *providing guidance, ideas, reviewing* and *input into manuscript 10%*, 2. Manoj Gambhir, *reviewing and input into manuscript 3%*, 3. Allen Cheng, *reviewing and input into manuscript 4%*, 4. Emma McBryde, *reviwing and input into manuscript 4%*, 5. Justin Denholm, *ideas, reviewing and input into manuscript 5%*, 6. EE Laine Tay, *input into manuscript 4%*,	No for all
4	Profiles of tuberculosis disease activation among contacts of patients with tuberculosis	Published *European Respiratory Journal 2019*	70% acquisition of data, data management, analysis and interpretation of data,	1. James Trauer, *providing guidance, ideas, reviewing* and *input into manuscript 10%*,	No for all

			conceptualisation and writing of manuscript, critical revision, approval of final draft for publication	2. Romain Ragonnet, *reviewing and input into manuscript 2%*, 3. Allen Cheng, reviewing and *input into manuscript 4%*, 4. Emma McBryde, *reviewing and input into manuscript 5%*, 5. Justin Denholm, *reviewing and input into manuscript 5%*, 6. EE Laine Tay, *data exptraction and input into manuscript 4%*,	
5	Heterogeneous infectiousness in mathematical models of tuberculosis: a systematic review	Under review	70%, literature search, data extraction, analysis and interpretation, conceptualisation and writing of manuscript, critical revision, approval of final draft for publication	1. James Trauer, *providing guidance, ideas, reviewing* and *input into manuscript 10%*, 2. Adeshina I Adekunle, *data extraction and input into manuscript 8%*, 3. Romain Ragonnet, *reviewing and input into manuscript 5%*, 4. Allen Cheng, *reviewing and input into manuscript 3%*, 5. Emma McBryde, *reviewing and input into manuscript 4%*,	No for all

6	Capturing heterogeneous patient infectiousness in transmission dynamic models of tuberculosis: a compartmental modelling approach	Submitted	70%, concept, design structure, writing equations and codes, writing of manuscript, critical revision, approval of final draft for publication	1. James Trauer, *providing guidance, ideas, reviewing* and *input into manuscript 10%,* 2. Romain Ragonnet, *providing mathematical and coding guidance* and *input into manuscript 10%,* 3. Allen Cheng, *reviewing and input into manuscript 5%,* 4. Emma McBryde, *reviewing and input into manuscript 5%*	No for all

I have not renumbered sections of submitted or published papers in order to generate a consistent presentation within the thesis.

Student name: Yayehirad Alemu Melsew

Student signature: **Date:** 12th March 2020

I hereby certify that the above declaration correctly reflects the nature and extent of the student's and co-authors' contributions to this work. In instances where I am not the responsible author, I have consulted with the responsible author to agree on the respective contributions of the authors.

Main Supervisor name: Dr James Trauer

Main Supervisor signature: **Date:** 12th March 2020

Acknowledgements

Several individuals and institutions have supported me to produce this thesis. First, I would like to forward my heartfelt gratitude to my enthusiastic PhD supervisors, Dr James Trauer, Associate Professor Manoj Gambhir, Professor Emma McBryde, Professor Allen Cheng, and Dr Romain Ragonnet. Each of the members of my PhD supervisory team has provided me extensive personal and professional guidance and taught me a great deal about both scientific research and life in general. My special appreciation goes to Dr James Trauer, my primary supervisor; James had been an extraordinary supervisor and mentor for the whole course of my study. More than I could ever thank him; he has shown me what a virtuous researcher and person should be by his example, not to mention the emotional supports.

Second, my PhD would not have been possible without the financial support of the following organisations. Foremost, I thank Monash University for giving me scholarships entitled Monash Graduate Scholarship and Monash University International Postgraduate Research Scholarship that covered the cost of all my PhD study. It is also a pleasure to mention my gratitude to the University of Gondar in Ethiopia for allowing me to complete my PhD with a partially paid study leave. I also want to thank the Centre for Research Excellence in Policy Relevant Infectious diseases Simulation and Mathematical Modelling ($PRISM^2$) for affording me many opportunities to network and attend conferences. My thanks also extend to the Victorian TB Program for providing TB contact investigation data.

Third, I am also grateful to the collaborators of my papers: Associate Professor Justin Denholm, Ee-Laine Tay, Tan Doan and Adeshina Adekule. I would also like to thank Louisa Wright for providing me feedback on the English writing of this thesis. I would like to acknowledge Kathryn Daly for her positive attitude and support in relation to the School's administrative issues. I also thank my friends Dr Debebe Shaweno, Dr Solomon Kelemu, Dr

Sam Brilleman, Dr Richard Ofori-Asenso, Asmare Gelaw and Neal Smith for their unreserved friendship and support.

Finally, my wife, Mrs Tigist Agegnehu please accept my heartfelt love and gratitude for carrying the huge burden of raising our young children Leul and Hanos in the course of my study. This would not happen without your whole-hearted support and love.

Abbreviations

AHR	Adjusted hazard ratio
AOR	Adjusted odds ratio
BCG	Bacillus Calmette–Guérin
CDR	Case detection rate
CI	Confidence interval
CXR	Chest x-ray
DHHS	(Victorian) Department of Health and Human Services
DS-TB	Drug-susceptible tuberculosis
EIR	Exposed infectious removed
EPTB	Extrapulmonary tuberculosis
HIV/AIDS	Human immunodeficiency virus/acquired immunodeficiency syndrome
HMC	Hamiltonian Monte Carlo
IGRA	Interferon-gamma release assay
LHS	Latin hypercube sampling
LTBI	Latent tuberculosis infection
MDR-TB	Multidrug-resistant tuberculosis
MLE	Maximum likelihood estimation
MOOSE	Meta-analyses and Systematic Review of Observational studies in Epidemiology
Mtb	*Mycobacterium tuberculosis*
NAT	Nucleic acid test
NBD	Negative binomial distribution
ODE	Ordinary differential equation
PCR	Polymerase chain reaction
PRISMA	Preferred Reporting Items for Systematic reviews and Meta-Analyses
PTB	Pulmonary tuberculosis
RR	Rate ratio
SARS	Severe acute respiratory syndrome
SEI	Susceptible exposed infectious
SEIE	Susceptible exposed infectious exposed
SEIR	Susceptible exposed infectious removed

SEIS	Susceptible exposed infectious susceptible
TB	Tuberculosis
TST	Tuberculin skin test
USA	United States of America
VTP	Victorian Tuberculosis Program
WHO	World Health Organization
XDR-TB	Extensively drug-resistant TB

Structure of the thesis

Chapter 1: Introduction: This first Chapter provides the context to the PhD project by describing the natural history of TB, its current global epidemiology, heterogeneity in infectiousness and methods of quantifying it for infectious diseases in general. This Chapter also presents the data sources and methods of data management and analyses.

Chapter 2: Risk factors for infectiousness of patients with tuberculosis: This Chapter summarises evidence relating to the heterogeneity of patients with TB with regards to their capacity to transmit the infection to their susceptible contacts by systematically reviewing TB contact investigation studies. This chapter highlights how patient-related factors are associated with the risk of *Mtb* transmission to contacts.

Chapter 3: The role of super-spreading events in *Mycobacterium tuberculosis* transmission: This Chapter quantifies the magnitude of heterogeneity in infectiousness of patients with TB and presents evidence of super-spreading in *Mtb* transmission by analysing prospectively collected TB contact investigation data from the Victorian TB Program, Australia.

Chapter 4: Profiles of tuberculosis disease activation among contacts of patients with tuberculosis: Using contact investigation data from the Victorian TB Program, this Chapter explores and quantifies heterogeneity in the rates of progression from *Mtb* infection to active TB disease among contacts of TB patients.

Chapter 5: Heterogeneous infectiousness in mathematical models of tuberculosis: By systematically reviewing previous TB modelling studies, this Chapter highlights how previous TB modelling studies approached heterogeneous infectiousness and identifies the factors which have been used to stratify TB patients into various infectiousness levels.

Chapter 6: Capturing heterogeneous infectiousness in transmission dynamic models of tuberculosis: This Chapter proposed a data-driven deterministic model of *Mtb* transmission, capturing heterogeneous infectiousness by considering three levels of TB patients' infectivity (namely non-spreaders, low-spreaders and super-spreaders). With this approach, this Chapter compares the effectiveness of a targeted active case finding intervention to that of an untargeted intervention.

Chapter 7: Conclusions and Future Directions: This Chapter presents the main findings from the thesis and their implications and proposes future research directions.

Chapter 1

1. Introduction

1.1. Chapter overview

This Chapter outlines five major themes that comprise the background to the issue of heterogeneous infectiousness of individuals with active TB disease and its implications. First, it presents the current understanding of the natural history and control methods of TB. Second, it describes the global epidemiology of TB and international efforts to control the epidemic. Third, it describes the current understanding of heterogeneity in infectiousness in communicable disease transmission and methods of quantifying this heterogeneity. Fourth, it introduces mathematical models of infectious diseases generally and TB models in particular. Finally, it summarises the background to this doctoral research, research questions, and the overall aim of the doctoral research. The Chapter concludes with a brief description of the methods used, particularly the data sources and approaches to data management.

1.2. Epidemiology of tuberculosis

1.2.1. Introduction to tuberculosis

Causation and history

Tuberculosis (TB) is an infectious disease caused by several species of gram-positive bacteria belonging to the *Mycobacterium tuberculosis complex*, which includes several human pathogens. However, human TB disease is mainly caused by the *Mycobacterium tuberculosis* (*Mtb*) species [1]. TB is an ancient disease of humankind, believed to have caused epidemics starting from human prehistory, such that it is estimated that *Mtb* might have killed more people than any other microbial species [2]. Although the clinical features of TB were identified by Hippocrates (460–370 BCE), it was Robert Koch's discovery of the tubercle bacillus in 1882 that fundamentally changed the understanding of TB and the quest for its treatment [3-6].

Infection

The transmission of *Mtb* from diseased individuals to uninfected persons is almost universally through the airways. Infection occurs when droplet nuclei (typically containing one to three tubercle bacilli) are exhaled from the diseased person and suspended in the air. The nucleus is then inhaled and ingested by alveolar macrophage in the previously uninfected person's lungs [7-9]. After the production of specific immunity, the bacilli enter a state of latency, called latent TB infection (LTBI). A proportion of infected people then proceed to symptomatic active TB disease within a short period from infection (e.g. less than one year), which may be termed "early progression". Some may remain latently infected but healthy for the rest of their lives, while others may later reactivate to symptomatic disease, sometimes many years form exposure. Globally, an estimated 1.7 billion people have LTBI, and these people form a critical reservoir for continued TB disease activation [10-12].

Although infection starts in the lungs, TB can affect any site in the human body, as the infection can spread via the blood to all body parts. For TB epidemiology and control, TB disease is commonly classified based on the involvement of the lungs, with TB involving the lungs termed pulmonary TB (PTB) and representing the infectious form of the disease, while extrapulmonary TB (EPTB) refers to TB at other body sites. The approximate duration of clinical disease from onset to cure or death in the absence of any intervention has been estimated to be three years [13]. Clinical signs and symptoms may differ depending on the age and immune status of the patient, in addition to the body organs involved. However, a person with TB disease often manifests a combination of several of the following typical symptoms: persistent cough, fatigue, weight loss, fever, loss of appetite and night sweats [14, 15].

Diagnosis

Early diagnosis of active TB disease ensures optimal outcomes for the individual patient and the prevention of further transmission. TB diagnostic delays are a central problem to the programmatic control of TB and are often due to the patient having no access to care or not seeking care; clinicians not considering or suspecting TB as a diagnosis; or the absence of a sensitive, readily available diagnostic test. Although the focus of diagnostic and treatment interventions in TB control is predominantly on active TB cases, it is also desirable to diagnose LTBI among those who are at increased risk of TB, to inform a decision on the provision of preventive therapy. Identifying healthy individuals with LTBI can be achieved through a reactive tuberculin skin testing (TST) and or a positive interferon-gamma release assay (IGRA), although these tests are unable to distinguish those at higher risk of progressing to active disease [16-18].

The World Health Organization (WHO) has approved four microbiologic tests and chest radiography as diagnostic tests for TB [17]. Conventional sputum smear microscopy is the

primary type of test used to diagnose pulmonary TB in resource-limited settings. However, the identification of acid-fast bacilli with light microscopy has imperfect sensitivity, especially in paucibacillary disease or low sputum bacillary concentration, such as paediatric TB and in TB-HIV co-infection. Optimisation techniques for smear microscopy, such as using fluorescence microscopy instead of conventional light microscopy, sputum processing, and testing serial sputum specimens, may also be used to improve sensitivity [19, 20].

Unlike sputum smear, the relatively recent technology of nucleic acid amplification tests (NAATs) can be used to diagnose both pulmonary and extrapulmonary TB. However, due to lack of access and inconsistent evidence, NAATs are not recommended to replace sputum smear testing but can be preferably used in conjunction with conventional tests and clinical data to confirm the diagnosis of TB [21-23]. Automated liquid culture is more sensitive and returns results faster than solid culture; however, this technique requires higher lab standards to avoid contamination. MPT64-based rapid immunochromatographic tests (ICT) are highly sensitive and specific tests used on mycobacteria growing on culture media that can confirm that the organism is *Mtb*. This test uses colonies that grow in either liquid or most typical culture, but the turnaround time is shorter for liquid cultures [24-26]. Chest radiography is also useful for detecting pulmonary abnormalities due to PTB and can be used as a triage tool, diagnostic aid or for screening purposes [27]. In December 2010, the WHO endorsed a molecular assay called Xpert MTB/RIF that enables simultaneous detection of *Mtb* and rifampicin (RIF) resistance [22, 28-30]. A further development in this TB diagnostic technology, called GeneXpert Omni, was released in 2015 as a more sensitive and accurate point-of-care molecular assay, although not yet endorsed by the WHO [31, 32].

Treatment

Once patients are diagnosed with TB, treatment is typically undertaken in two phases. The standard initial phase of treatment for drug-susceptible TB consists of two months of

intensive isoniazid, rifampicin, pyrazinamide and ethambutol, while the continuation phase consists of isoniazid and rifampicin given for four months. Patients with tuberculosis caused by drug-resistant (especially MDR/XDR) organisms should be treated with specialised regimens containing quality-assured second-line anti-tuberculosis drugs. Treating MDR-TB is more complicated than drug-susceptible ones due to the need to identify the type of resistance and longer duration of treatment; however, there is a major move towards shorter MDR-TB regimen instead of the traditional longer regimens [17, 33].

1.2.2. Global burden of tuberculosis

Although the development of effective drugs in the 1960s dramatically reduced TB-related mortality and morbidity, TB still caused an estimated 10.0 million cases and 1.2 million deaths worldwide in 2018. TB ranks as the ninth leading cause of death from all sources and the number one killer disease from a single infectious pathogen [5, 34, 35].

Dramatic declines in TB-related deaths have been seen in history before the development of effective drugs. A classic example is a decline in TB annual mortality in England and Wales from more than 300 deaths per 100,000 population in 1815 to fewer than 100 deaths in 1935 [36]. The story is similar in other developed countries, which used to suffer from considerable TB epidemics before economic development and improvements in living standards were achieved. The trend of TB morbidity from 1990 to the present day is generally towards a continued decline in high-income countries. However, from the beginning of the 1990s, the global TB burden began to increase due to high rates of HIV infection in Africa and the economic deterioration of former Soviet countries [37].

The global TB burden is heterogeneously distributed, with low- and middle-income countries bearing the highest burden of the disease [38]. For the period 2016-2020, the WHO classified countries as a low-burden country or high-burden country based on the absolute numbers of

cases and in terms of case rates per capita. The WHO lists for high-burden countries consists of three intersecting groups: TB, TB/HIV and MDR-TB; with each list containing 30 countries. Forty-eight countries are in at least one of these groups, and these high-burden countries together account for 85-89% of the global burden [39].

Locations that currently enjoy a low-burden of TB, such as the European countries, endured severe TB epidemics in the 19th century but transitioned to low-burden status over time. Generally, developed or high-income countries, where the quality of life improved dramatically with economic and industrial development during the 20th century, are now classified as low-burden countries [36]. In 2017 in most high-income countries, there were fewer than ten new cases per 100,000 population, while there were between 150 and 400 new cases per 100,000 population in most high-burden countries. In countries such as South Africa, Mozambique and the Philippines, there is an extreme rate of TB (more than 500 new cases per 100,000 population); interestingly, some of these countries are high HIV burden settings while others are not, highlighting the diverse drivers of the epidemic by context and region [35]. Tuberculosis remains a disease of poverty that is inseparably associated with overcrowding and poor nutrition. HIV infection, diabetes, alcohol use, poor nutrition and smoking are among the most prominent risk factors accounting for the high TB burden in low- and middle-income countries although these factors cannot fully explain the difference in TB burden [40].

1.2.3. Global efforts to control tuberculosis

The re-emergence of TB in developed countries due to international migration and the resurgence of the disease in lower-income settings has required the global community to re-think coordination and funding of TB-related research and control strategies [41]. In 1993, the WHO declared TB a global public health emergency and the disease to be a matter of global public health concern. Soon after, it launched the “directly observed treatment, short

course” (DOTS) strategy to ensure that the correct drugs are taken at the right time for the full duration of treatment, to prevent further transmission [42]. Since then, the inclusion of tuberculosis-related indicators in the Millennium Development Goals and the development and implementation of the Stop TB Strategy underpinning the Global Plan to Stop TB 2006–2015 have helped to co-ordinate and accelerate global TB control interventions, resulting in a significant reduction in TB incidence rates and mortality over this period [37, 43].

Following these achievements, the WHO developed the new End TB Strategy for post-2015 TB elimination activities. The End TB Strategy’s objective is a 95% reduction in TB deaths and 90% reduction in the TB incidence rate by 2035, compared to 2015 rates [44]. This ambitious plan faces challenges from increased migration, poverty and unpredictable factors, such as human-made and natural disasters that could lead to the destruction of health care facilities and derail the expected progress. Moreover, other epidemiological factors, such as TB’s poorly understood natural history, the extent to which past disease episodes confer immunity, and the uncertain impact of future interventions make it difficult to be confident about the possibility of its elimination.

1.3. Heterogeneity in disease transmission

The emergence and re-emergence of infectious disease may occur due to combinations of environmental, social, political and economic factors. The infectious agent can be transmitted from its reservoir, typically a human host, to a new host via direct or indirect transmission mechanisms. In the case of direct transmission, the infection spreads by immediate transfer of the infectious agent from the infectious host to a new host via direct physical contact, or body fluid or droplet spread. In the case of an indirect mechanism, there are three primary pathways of agent transmission: vehicle-borne, vector-borne, and airborne [45].

The chance of an infectious individual coming into contact with a susceptible host and transmitting the infectious agent depends on the environment enabling transmission, the infectivity of the infectious individual and the susceptibility of the uninfected host [46]. Thus, heterogeneity in pathogen transmission can arise due to these myriad factors related to the pathogen, the infected host, susceptible host and the environment.

1.3.2. Heterogeneous infectiousness in disease transmission

Heterogeneity in the ability of individuals to transmit infections is an essential concept in disease transmission dynamics and has implications for evaluating disease control interventions. In infectious disease transmission, it has been recognised that there are heterogeneities in the capacity of infective individuals to transmit infection [46]. Heterogeneity between infectious individuals in their capacity to transmit the infectious agents can be explained thus: a very few highly infectious individuals produce a large number of secondary infections, while others may infect very few contacts or maybe entirely non-infectious [47, 48].

This individual infectiousness heterogeneity has been explained concerning the transmission of various infectious disease. The established statistical pattern of the 20/80 rule has been

applied to vector-borne parasitic infections and sexually transmitted infections; this rule of thumb states that a "core" 20% of patients are highly infectious, causing 80% of secondary infections. The rule suggests that control programs targeting this highly infectious segment of the infectious population are likely to be considerably more effective [49]. Similarly, in the spread of the severe acute respiratory syndrome (SARS), infectious persons that disproportionally infected more secondary contacts than other (super-spreaders) were a crucial feature of the epidemic [50, 51]. Super-spreading events whereby very few individuals produce a large number of secondary infections are not an exceptional phenomenon; instead, they are a core characteristic of the transmission of most infectious diseases.

Methods for quantifying heterogeneous infectiousness

Approximations of the number of secondary infections per index using appropriate statistical distributions provide opportunities for gaining new insights into the spread and control of infectious disease in heterogeneous populations [52]. With the availability of contact investigation data, a particular method of analysing heterogeneity in infectiousness has become available. The distribution of the number of secondary infections per index produced from contact tracing can be fitted with a negative binomial distribution to describe heterogeneous infectiousness and quantify the presence of super-spreading [48].

An offspring distribution can be approximated by using a probability distribution that describes the number of secondary infections produced across the range of patients with the disease. This process would be expected to follow a negative binomial distribution, which is similar to a Poisson distribution in that it is discrete, but which also allows for "over-dispersion", where the variance may be greater than the mean [53]. The negative binomial distribution can accommodate such over-dispersed count data such as the number of secondary infections or offspring per index case [53-55]. As has been established by *Lloyd-Smith et al.* [48], offspring

distributions can be well described by the negative binomial distribution, which permits sufficient flexibility with only two parameters (the shape parameter and the mean). Denoting the individual reproductive number by v and the distribution of individual reproductive numbers (offspring distribution) by Z, v follows a negative binomial offspring distribution with dispersion parameter k and mean m, such that $Z \sim NegB(m, k)$. The dispersion parameter k quantifies the extent of over-dispersion in the count data. There will be some heterogeneity in the realisation of a simple Poisson distribution, however, if there is extra-heterogeneity between index cases in the number of secondary infections produced, dispersion increases and the parameter k approaches zero $(k \to 0)$. In the absence of over-dispersion, $k \to \infty$ and the mean and the variance approach parity, with the negative binomial distribution reducing to the Poisson distribution. If $k = 1$, the negative binomial distribution reduces to the geometric distribution, such that the negative binomial model can accommodate Poisson, geometric and over-dispersed distributions [54].

Assuming that the distribution of secondary infections per index follows a negative binomial distribution, the probability of observing index cases with $v \geq 0$ number of infected contacts is given by:

$$P(Z = v) = \frac{(k+v-1)!}{v!(k-1)!} . \left(\frac{m}{m+k}\right)^{v} \left(1 + \frac{m}{k}\right)^{-k}, m > 0, k > 0$$

Where, m= arithmetic mean offspring distribution, k= dispersion parameter. As $k \to \infty$, the variance $\mathrm{m}(1 + (\mathrm{m}/\mathrm{k}))$ approaches the mean, m or the overdispersion decreases, i.e. the distribution will become homogeneous.

The parameters k and m can be estimated by maximum likelihood estimation (MLE), which provides unbiased estimates, especially for large sample sizes [54]. The MLE of the mean of the offspring distribution, m, is the sample mean of Z and can be interpreted as the mean

number of secondary infections. The dispersion parameter, k, can be estimated after substituting the MLE of the mean into the likelihood expression, with a value of k less than one interpreted as evidence of super-spreading [56].

Although there is no standard definition for classifying super-spreading, it was described by a given cut-off point in the number of secondary infections per index patient. After producing the distribution for the number of secondary infections per index and computing the mean of the distribution (including any over-dispersion), a Poisson distribution can be produced (without over-dispersion) with distribution mean equal to the mean number of infections per index, m. The cut-off point for defining a super-spreader is then an index case that produces a number of secondary infections that is greater than the 99th centile of the Poisson distribution [48, 57].

Heterogeneous infectiousness in *Mtb* transmission

Heterogeneous infectiousness is likely to be more substantial in *Mtb* transmission than for other infectious diseases since TB has protean clinical manifestations and highly variable durations of disease episodes that can directly affect the risk of transmission [58]. The involvement of the lungs defines patients with pulmonary TB as infectious, while extrapulmonary TB patients are almost universally considered non-infectious. However, the number of secondary infections that an infectious case produces is likely to be the result of the interplay of a multitude of factors related to the clinical and behavioural characteristics of TB, the patient, the agent and the enabling environment [59, 60]. The systematic review in Chapter 2, will present factors associated with patients' infectiousness in detail by bringing together information from TB contact tracing studies.

Epidemiological quantification of TB patients' heterogeneity in the number of secondary infections produced using genotypic data has shown super-spreading events to be a prominent

feature of *Mtb* transmission [56]. This study found that the negative binomial dispersion parameter for TB genotypic clustering was 0.1, which represents a similar level of dispersion to that reported for SARS super-spreading. However, the genotypic data used in this previous study did not consider epidemiologically linked contacts, i.e. genotypic data was not collected by contact tracing.

I contributed to a published review on the importance of heterogeneity on TB epidemiology [61], and this article is attached to this thesis as an Appendix (*Appendix C: Additional publication relevant to the thesis*). In this review, the drivers of heterogeneity in TB epidemiology at the level of the infectious host, organism, susceptible host, environment and distal determinants were summarised and discussed. At the level of the infectious host, the extent to which each infectious person transmits *Mtb* infection depends on several clinical, demographic and behavioural characteristics. Transmission from adults with pulmonary TB (particularly people with smear-positive and cavitary TB) is extensive, while transmission from people with extrapulmonary is very rare. *Mtb* displays multiple lineages, which differ in their genomic make-up and several aspects of their clinical and epidemiological behaviour, including disease progression, disease severity, transmissibility and geographic distribution. In particular, the differences between drug-susceptible and drug-resistant strains in their ability to survive and reproduce influence the spread of infection. The susceptibility level of apparently healthy individuals affects the risk of TB disease progression after the initial infection.

Behavioural or social factors such as social mixing also influence the spread of infection by modifying the number of contacts exposed to infectious host. The setting or the physical environment in which *Mtb* transmitted is also an important source of heterogeneity in two ways: both due to increased population density and through environmental features that facilitate airborne transmissions such as poor ventilation or high levels of indoor air pollution.

More distal factors such as migration, urbanisation, demographic transition and other broad global trends combined with weak and inequitable policy and planning lead to pockets of poverty, unhealthy behaviours and weak health systems, which are also sources of spatial heterogeneity in *Mtb* transmission.Infectious disease modelling

1.3.3. Overview

Infectious disease models are mathematical descriptions of the spread of infection and are simplified representations of the system, rather than entirely accurate representations of the natural processes. Although models cannot be entirely accurate and will never predict the precise course of the disease epidemic, they can provide broad guidance to policy-makers, and uncertainty estimates around their predictions. Infectious disease models typically have one of two primary purposes: either to understand the behaviour of the system or to inform interventions through predictions [62-64]. A model that is appropriate to the question at hand and whose parameters are informed by the best available empiric data is generally preferred over models that are not. A model that is designed to help us understand the behaviour of infectious disease transmission should concentrate on the characteristics that are of interest, while usually simplifying others because of a desire for parsimony. Although a predictive model may require the inclusion of many epidemiological and clinical features, the implementation of these elaborations must be informed by available data.

Construction of mechanistic models for an infectious disease should start from a fundamental natural history of the infection at the individual level. Frequently, this approach classifies the stages of the infection process into susceptible (S), exposed (E), infectious (I) and removed (R), which may be abbreviated to SEIR. Models may also represent the infection process as SIR dynamics, ignoring the exposed class depending on the infection under consideration. In cases where people are expected to become fully susceptible again immediately after the first episode of infection, which results in SEIS or SIS dynamics. These simplified classes may

not be appropriate for all infectious diseases; thus, modellers use different class structures based on the infection under study and other epidemiological features [62].

Mathematical models may be classified as either deterministic or stochastic on the bases of the type of transitions between disease states. Deterministic models are models in which the values for the outputs of the system are entirely determined by the parameters of the model with no randomness involved, while stochastic models allow the number of people transitioning between states of infection and disease to vary due to chance. Perhaps an even more important distinction is between compartmental models and individual-based models, based on the method of representing the population. Compartmental models group individuals in the population into states or compartments, based on characteristics relevant to the infection process, while individual-based models (also known as agent-based models) track the infection process for every individual in the population. Compartmental models can be deterministic or stochastic, while individual-based models are almost universally stochastic. Other specific types of individual-based models include network models, in which the network of individual contacts is explicitly modelled, and infection is dependent on this contact network [65]. Most mechanistic infectious disease models are dynamic transmission models that incorporate contact between infectious and susceptible persons, which results in transmission. Thus, the force of infection in dynamic models is a function of the number of infectious individuals, which distinguishes these models from static or catalytic models that assume a constant force of infection [66]. Dynamic models are therefore powerful tools for capturing the non-linear features of an epidemic.

The principle of parsimony decrees that models should be as simple as possible and discourages adding complexity to a model without reasonable justification. A model that contain as few variables as possible to describe the essential attributes of a system is a parsimonious model. If the purpose of the model is to be used as an operational tool for

making predictions, sticking to the parsimony principle is very important. However, when the purpose of the model is to provide explanations of the actual natural process, the application of parsimony principle to evaluate this model may not be plausible [67-69].

1.3.4. Transmission heterogeneity in infectious disease models

Infectious disease models may attempt to capture transmission heterogeneity due to socio-demographic factors and host infectiousness. Models that have incorporated heterogeneity due contact structures have reported that epidemiologically relevant social contact patterns are not random and that contacts made by children and adolescents are more assortative (preferentially made with similar population groups, or "with-like") than contacts made by other age groups [70]. Network models are more intuitive and increase realism than compartmental models that assume homogeneous-mixing for predicting disease spread in heterogeneously mixing populations [71, 72]. For instance, a stochastic small-world network model was employed to model transmission of the SARS virus with both geographically localised outbreaks and super-spreaders [50].

On the other hand, models may attempt to capture transmission heterogeneity that arises due to variation in host infectiousness. For example, HIV transmission models have incorporated heterogeneous infectiousness by allowing infectiousness to vary throughout the infectious period [73, 74], while another HIV model stratified the population by age and level of risk as well as variable infectivity [75]. These features have important implications for the effectiveness of interventions; a compartmental model employed to investigate multiple interventions suggested that in situations where few individuals are responsible for a large proportion of transmission, interventions targeted towards high-risk subgroups of the population may be vastly more effective in reducing the prevalence and achieving synergy between interventions [76]. Similarly, in an SIR model of salmonella transmission within cattle groups, heterogeneous infectiousness was captured by stratifying the infectious

population, with each stratum taking on different infectious periods and ‘transmission coefficients’ [77].

1.3.5. Models of tuberculosis

Since *Waaler, H. T.* published the first TB model in 1962 [78], the use of dynamic transmission models in TB has increased. The most frequent types of models are compartmental models, while relatively few individual-based and network models have been developed [79]. However, the assumptions used by different TB models may differ due to the complex natural history of TB and the stated aim of each model. Most TB models follow the susceptible, exposed, infectious, and recovered (SEIR) compartmental structures, with one of the most familiar TB model structures following five health states, namely; susceptible (uninfected), latently infected (early), latently infected (late), active TB and recovered. A review study of TB models presented this commonly used basic compartmental structure of TB transmission model [80]. Additional complexities to the basic structure may then be added, depending on the aim of the model.

TB models have been used to inform policy by predicting the future impact of different control interventions and understanding the disease system or epidemiology in a specific setting. *Blower, S. M., et al.* (1995) used a model to estimate the time required for a TB epidemic to rise, fall and reach a stable endemic level. This classic model helped to explain the decline of TB epidemics in developed countries before the discovery of TB treatment drugs [81]. Models have also been used to quantify TB-HIV co-infection and the effect of different treatment strategies for both HIV and TB [82, 83]. Possible impacts of risk behaviours such as smoking [84-86] and chronic disease comorbidities such as diabetes mellitus on TB epidemiology have also been assessed with models [87, 88]. Models have been used to evaluate the expected benefits of novel or hypothetical interventions in addition to informing an evaluation of currently available control strategies. Such modelling studies

may also incorporate cost-effectiveness evaluations. *Abu-Raddad, L. J., et al.* used a compartmental model to project the effectiveness of a novel neonatal vaccination which was still under development. This study also evaluated the effects of new drug regimens with a shorter duration of treatment and new diagnostic techniques by computing the possible reduction in TB incidence [89]. *Channing, L. and Sinanovic E.* also modelled the cost-effectiveness of a new vaccine which was under development [90].

Heterogeneous infectiousness in TB models

As described in previous sections, heterogeneous infectiousness is important in *Mtb* transmission; however, TB models usually assume homogenous infectivity. Ignoring heterogeneity has a profound effect on modelling epidemics and leads to substantial biases in estimating the proportion of individuals who will become infected, which in turn influences the associated policy decisions [52]. Some past compartmental models of TB transmission dynamics have attempted to capture heterogeneity in patients' infectiousness by stratifying the active TB compartment into different levels of infectivity. Among the most typical approaches is stratification as either infectious (usually representing pulmonary TB) or non-infectious (extrapulmonary TB) [91-95]. Other TB models have considered sputum smear status as a factor in stratifying patients' levels of infectiousness, considering both smear-negative and smear-positive pulmonary TB to be infectious, with the relative infectiousness of smear-negative patients compared to smear-positive typically set between 15 and 25% [96-111]. In Chapter 5, a systematic review of TB models that assumed heterogeneous infectiousness will present details about past TB modelling approaches to stratifying the active TB compartment by the level of infectiousness.

1.4. The rationale for this doctoral research

TB is one of the top ten causes of death worldwide, with an estimated 10.0 million cases and 1.2 million deaths in 2018 [112]. However, this global TB epidemic is not homogeneously

distributed in a given population but is instead a collection of heterogeneous local micro-epidemics [38]. Heterogeneous infectiousness is one of the many ways in which individuals with active TB differ in relation to TB transmission dynamics. This heterogeneity of transmission has the potential to disrupt elimination strategies, which are often applied consistently across areas and populations, making it essential to understand and quantify.

Several previous infectious disease studies have reported heterogeneity in the capacity of individual source patients to transmit the infection to their contacts [48, 50, 51, 113]. The variation between infectious individuals in their capacity to transmit infectious agents is well described, with some infecting a large number of contacts, while others may infect very few or none [47, 48, 51, 56, 114]. This heterogeneous infectiousness and its extreme event, super-spreading, are recognised characteristics of the transmission of several globally important infections [113].

Heterogeneous infectiousness is important for *Mtb* transmission in particular, as TB has profoundly variable clinical manifestations and disease durations, which in turn affect the risk of *Mtb* transmission [115]. The clinical manifestations involving the lungs allow the bacilli to access the airway and transmission to occur; however, this is also dependent on the patients' ability to expectorate and their bacillary load, as well as proximity and duration of exposure of the susceptible contacts [59, 116]. Thus, quantification of heterogeneity in *Mtb* transmission will help to understand better its transmission characteristics. Nevertheless, how TB patients vary with respect to their capacity to produce secondary infection including the extent to which super-spreaders exist and are responsible for driving transmission remains poorly understood. As the global TB response moves towards ending TB, understanding transmission heterogeneity and characterising those with greater capacity to spread the infection will assist with targeting interventions. Quantification of heterogeneity in *Mtb*

transmission will also help in constructing TB transmission dynamic models for informing policy, such as evaluating the effectiveness of targeted interventions.

TB transmission dynamic models should ideally be governed by the reality of heterogeneous infectiousness, rather than universally assuming homogeneous populations. By systematically reviewing previous TB transmission modelling studies, we will analyse the existing methods used to capture heterogeneity in infectiousness. The review will help to construct a new data-informed model, which can explain current TB transmission dynamics capturing heterogeneous infectiousness assumptions.

1.5. Research aims

The overall aim of this thesis is to establish evidence concerning the variation in TB patients' infectiousness, to characterise highly infectious patients, and to integrate this evidence with a mathematical TB transmission model to predict the effectiveness of targeted interventions.

Thus, the specific aims of this thesis are to answer the following research questions:

1. What are the risk factors for the infectiousness of patients with TB?
2. What is the role of super-spreading in *Mtb* transmission?
3. How heterogeneous is the risk of progression towards active TB disease among contacts?
4. How did past TB mathematical models capture heterogeneous infectiousness?
5. What is the likely impact of active case finding interventions targeting super-spreaders?

1.6. Thesis scope

The thesis includes two broad methodological approaches: statistical analysis and mathematical modelling.

Theme 1: using statistical analyses to establish evidence of TB patients' infectiousness heterogeneity and risk factors

- Describe the risk factors for the infectiousness of patients with TB, using systematic reviews and meta-analyses of contact investigation studies
- Conduct offspring distribution analyses of TB contact investigation data to quantify infectiousness heterogeneity and super-spreading
- Conduct survival analyses using TB contact investigation data to describe patterns of active TB progression

Theme 2: using mathematical modelling to develop a model that can capture infectiousness heterogeneity and evaluate targeted active case-finding interventions.

- Describe methods of capturing heterogeneous patient infectivity by systematically reviewing mathematical models of TB
- Develop a compartmental mathematical model of TB transmission capturing heterogeneous infectiousness and evaluate the effectiveness of active case finding targeted to super-spreaders

1.7. Overview of methods and materials

Study setting and data sources

Victoria is a state of Australia with approximately 5.6 million people and a single centralised tuberculosis program (The Victorian Tuberculosis Program; VTP). In Victoria, notification of all confirmed or suspected cases of TB disease is mandatory for both laboratories and clinicians. Culture and molecular confirmation of *Mtb* are routine in this setting. While hospitalisation of cases is not mandatory, those with pulmonary disease are typically maintained in isolation until they are considered non-infectious (received more than two weeks of TB treatment or smear negativity) [117, 118]. On receipt of a notification, a public health

nurse from the VTP is allocated to the patient to provide support, assist with treatment compliance, and assess the requirements and extent of contact tracing necessary. Household contacts and others with greater than an estimated eight hours of contact are considered eligible for screening, with an individualised assessment of high-risk contacts performed (e.g. immunosuppressed or high-intensity exposure).

Contact investigation initially consists of clinical assessment and serial testing for *Mtb* infection. Testing of contacts is conducted by either tuberculin skin testing (TST) using the Mantoux procedure or an interferon-gamma release assay (IGRA), although during the study period the vast majority of tests were undertaken using TST. Those negative on initial testing are tested 8-12 weeks again following exposure. Contacts with either symptom suggestive of active disease or a positive test for *Mtb* infection undergo chest x-ray (CXR) and further clinical assessment, with isoniazid preventive treatment offered for those for whom the active disease has been excluded [117].

Data collection and management

Contact investigation data from VTP, which are stored by the Victorian Department of Health and Human Services, were extracted. Index patients were confirmed cases of TB notified from 1st January 2005 to 31st December 2015 among residents of Victoria. The dataset includes linked contact tracing information, along with the results of their testing for infection with *Mtb*, which were investigated up to March 2017. Contacts from this dataset were linked to subsequent notifications with active TB during the same period [119]. Data were extracted after ethical approval was obtained from Monash University, Human Research Ethics Committee Project Number: 7776 (***Appendix A***) and the VTP gave permission.

Data analysis

In this thesis, three broad analyses were conducted: systematic reviews and meta-analyses, statistical modelling analyses, and mathematical transmission dynamic modelling. The systematic review and meta-analysis of contact investigation studies pooled evidence for risk factors of patients' infectiousness. Several statistical analyses of contact tracing data were performed, including fitting the negative binomial distribution to the offspring distribution, survival analyses, univariate and multivariate regressions with negative binomial, logistic and Cox-proportional hazard models.

Previous modelling approaches were also described by systematically reviewing TB transmission models. Finally, a data-driven transmission dynamic model was constructed to capture TB patients' infectiousness heterogeneity and to evaluate the impact of targeting case-finding interventions towards the most infectious patients. In both the statistical analyses and the dynamic transmission modelling, R version 3.4.0 (R Project for Statistical Computing) was used [120].

In the case of the systematic reviews, abstract screening, full text review and data extractions were performed by collaborating with other researchers.

1.8. References

1. Comas, I. and S. Gagneux, *The past and future of tuberculosis research.* PLoS pathog, 2009. **5**(10): p. e1000600.
2. Daniel, T.M., *The history of tuberculosis.* Respiratory medicine, 2006. **100**(11): p. 1862-1870.
3. Sakula, A., *Robert Koch: centenary of the discovery of the tubercle bacillus, 1882.* Thorax, 1982. **37**(4): p. 246-251.
4. Cambau, E. and M. Drancourt, *Steps towards the discovery of Mycobacterium tuberculosis by Robert Koch, 1882.* Clinical Microbiology and Infection, 2014. **20**(3): p. 196-201.
5. Sensi, P., *History of the development of rifampin.* Reviews of infectious diseases, 1983. **5**(Supplement_3): p. S402-S406.
6. Herzog, H. and Basel, *History of tuberculosis. Respiration. 1998;65(1):5-15.* Respiration, 1998. **65**(1): p. 5-15.

7. Orme, I.M., *A new unifying theory of the pathogenesis of tuberculosis.* Tuberculosis, 2014. **94**(1): p. 8-14.
8. Cardona, P.J., *Revisiting the natural history of tuberculosis. The inclusion of constant reinfection, host tolerance, and damage-response frameworks leads to a better understanding of latent infection and its evolution towards active disease.* Arch Immunol Ther Exp (Warsz), 2010. **58**(1): p. 7-14.
9. Dannenberg, A.M., Jr., and P.J. Converse, *Pathophysiology and Immunology*, in *Tuberculosis and Nontuberculous Mycobacterial Infections*, D. Schlossberg, Editor. 2011, American Society for Microbiology Press.
10. Houben, R.M. and P.J. Dodd, *The global burden of latent tuberculosis infection: a re-estimation using mathematical modelling.* PLoS medicine, 2016. **13**(10): p. e1002152.
11. Behr, M.A., P.H. Edelstein, and L. Ramakrishnan, *Revisiting the timetable of tuberculosis.* BMJ, 2018. **362**: p. k2738.
12. Reichler, M.R., et al., *Risk and Timing of Tuberculosis Among Close Contacts of Persons with Infectious Tuberculosis.* The Journal of infectious diseases, 2018. **40**: p. 1-9.
13. Tiemersma, E.W., et al., *Natural history of tuberculosis: duration and fatality of untreated pulmonary tuberculosis in HIV negative patients: a systematic review.* PloS one, 2011. **6**(4): p. e17601.
14. Leung, A.N., *Pulmonary tuberculosis: the essentials.* Radiology, 1999. **210**(2): p. 307-322.
15. Association, A.L. *Tuberculosis Symptoms, Causes & Risk Factors*. Lung Health & Diseases 2018 October 2018 [cited 2018 December 13]; Available from: https://www.lung.org/lung-health-and-diseases/lung-disease-lookup/tuberculosis/tuberculosis-symptoms.html.
16. Salgame, P., et al., *Latent tuberculosis infection–revisiting and revising concepts.* Tuberculosis, 2015. **95**(4): p. 373-384.
17. TBCAREI, *International standards for tuberculosis*. 3 ed. 2014, The Hague: WHO.
18. WHO, *Recommendations for investigating contacts of persons with infectious tuberculosis in low- and middle-income countries*. 2012, World Health Organization: Geneva.
19. Steingart, K.R., A. Ramsay, and M. Pai, *Optimizing sputum smear microscopy for the diagnosis of pulmonary tuberculosis.* Expert review of anti-infective therapy, 2007. **5**(3): p. 327-331.
20. Steingart, K.R., et al., *Fluorescence versus conventional sputum smear microscopy for tuberculosis: a systematic review.* The Lancet infectious diseases, 2006. **6**(9): p. 570-581.
21. Ling, D.I., et al., *Commercial nucleic-acid amplification tests for diagnosis of pulmonary tuberculosis in respiratory specimens: meta-analysis and meta-regression.* PloS one, 2008. **3**(2): p. e1536.
22. Nicol, M.P., et al., *Accuracy of the Xpert MTB/RIF test for the diagnosis of pulmonary tuberculosis in children admitted to hospital in Cape Town, South Africa: a descriptive study.* The Lancet Infectious Diseases, 2011. **11**(11): p. 819-824.
23. Pai, M. and D.I. Ling, *Rapid diagnosis of extrapulmonary tuberculosis using nucleic acid amplification tests: what is the evidence?* 2008.
24. Yin, X., et al., *Commercial MPT64-based tests for rapid identification of Mycobacterium tuberculosis complex: a meta-analysis.* Journal of Infection, 2013. **67**(5): p. 369-377.
25. Drobniewski, F.A., et al., *Modern laboratory diagnosis of tuberculosis.* The Lancet Infectious Diseases, 2003. **3**(3): p. 141-147.

26. Chihota, V., et al., *Liquid vs. solid culture for tuberculosis: performance and cost in a resource-constrained setting.* The international journal of tuberculosis and lung disease, 2010. **14**(8): p. 1024-1031.
27. WHO, *Chest radiography in tuberculosis detection* 2016: Geneva.
28. Vadwai, V., et al., *Xpert MTB/RIF: a New Pillar in Diagnosis of Extrapulmonary Tuberculosis?* Journal of Clinical Microbiology, 2011. **49**(7): p. 2540-2545.
29. Tortoli, E., et al., *Clinical validation of Xpert MTB/RIF for the diagnosis of extrapulmonary tuberculosis.* European Respiratory Journal, 2012: p. erj01763-2011.
30. Lawn, S.D., et al., *Advances in tuberculosis diagnostics: the Xpert MTB/RIF assay and future prospects for a point-of-care test.* The Lancet Infectious Diseases, 2013. **13**(4): p. 349-361.
31. FIND, C.a. *World's most portable molecular diagnostics system unveiled at AACC.* 2015 [cited 2019 May 14]; Available from: https://www.prnewswire.com/news-releases/worlds-most-portable-molecular-diagnostics-system-unveiled-at-aacc-300119213.html.
32. Drain, P.K. and N.J. Garrett, *The arrival of a true point-of-care molecular assay—ready for global implementation?* The Lancet Global Health, 2015. **3**(11): p. e663-e664.
33. WHO, *WHO consolidated guidelines on drug-resistant tuberculosis treatment*. 2019: Geneva.
34. Chakraborty, S. and K.Y. Rhee, *Tuberculosis drug development: history and evolution of the mechanism-based paradigm.* Cold Spring Harbor perspectives in medicine, 2015: p. a021147.
35. WHO, *Global Tuberculosis Report 2018*. 2018: Geneva.
36. Wilson, L.G., *The historical decline of tuberculosis in Europe and America: its causes and significance.* Journal of the history of medicine and allied sciences, 1990. **45**(3): p. 366-396.
37. Dye, C. and P. Glaziou, *The Global Tuberculosis Epidemic: Scale, Dynamics, and Prospects for Control.* Tuberculosis: The Essentials, 2016. **237**: p. 1.
38. Pai, M., et al., *Tuberculosis.* Nature Reviews Disease Primers, 2016. **2**: p. 16076.
39. WHO, *Use of high burden country lists for TB by WHO in the post-2015 era.* Geneva: World Health Organization, 2015.
40. Lawn, S.D. and A.I. Zumla, *Tuberculosis.* Lancet, 2011. **378**: p. 57-52.
41. Borgdorff, M. and D.v. Soolingen, *The re-emergence of tuberculosis: what have we learnt from molecular epidemiology?* Clinical Microbiology and Infection, 2013. **19**(10): p. 889-901.
42. WHO, *What is DOTS?: A guide to understanding the WHO-recommended TB Control Strategy Known as DOTS*. 1999, Geneva: World Health Organization.
43. Raviglione, M.C. and M.W. Uplekar, *WHO's new Stop TB Strategy.* The Lancet, 2006. **367**(9514): p. 952-955.
44. WHO, *The End TB Strategy: Global strategy and targets for tuberculosis prevention, care and control after 2015.* 2014: Geneva, Switzerland
45. Osterholm, M.T. and C.W. Hedberg, *13 - Epidemiologic Principles*, in *Mandell, Douglas, and Bennett's Principles and Practice of Infectious Diseases (Eighth Edition)*, J.E. Bennett, R. Dolin, and M.J. Blaser, Editors. 2015, Content Repository Only!: Philadelphia. p. 146-157.e2.
46. VanderWaal, K.L. and V.O. Ezenwa, *Heterogeneity in pathogen transmission: mechanisms and methodology.* Functional Ecology, 2016. **30**(10): p. 1606-1622.
47. Galvani, A.P. and R.M. May, *Epidemiology: dimensions of superspreading.* Nature, 2005. **438**(7066): p. 293-295.

48. Lloyd-Smith, J.O., et al., *Superspreading and the effect of individual variation on disease emergence.* Nature, 2005. **438**(7066): p. 355-359.
49. Woolhouse, M.E., et al., *Heterogeneities in the transmission of infectious agents: implications for the design of control programs.* Proceedings of the National Academy of Sciences, 1997. **94**(1): p. 338-342.
50. Small, M., C. Tse, and D.M. Walker, *Super-spreaders and the rate of transmission of the SARS virus.* Physica D: Nonlinear Phenomena, 2006. **215**(2): p. 146-158.
51. Shen, Z., et al., *Superspreading SARS events, Beijing, 2003.* Emerging infectious diseases, 2004. **10**(2): p. 256-260.
52. Getz, W.M., et al., *Modeling the invasion and spread of contagious diseases in heterogeneous populations.* Disease Evolution: Models, Concepts, and Data Analyses, 2006. **71**: p. 113-144.
53. Bliss, C.I. and R.A. Fisher, *Fitting the negative binomial distribution to biological data.* Biometrics, 1953. **9**(2): p. 176-200.
54. Lloyd-Smith, J.O., *Maximum likelihood estimation of the negative binomial dispersion parameter for highly overdispersed data, with applications to infectious diseases.* PloS one, 2007. **2**(2): p. e180.
55. Fisher, R.A., *The negative binomial distribution.* Annals of Eugenics, 1941. **11**(1): p. 182-187.
56. Ypma, R.J., et al., *A sign of superspreading in tuberculosis: highly skewed distribution of genotypic cluster sizes.* Epidemiology, 2013. **24**(3): p. 395-400.
57. Lloyd-Smith, J.O., S.J. Schreiber, and W.M. Getz. *Moving beyond averages: Individual-level variation in*. in *Mathematical Studies on Human Disease Dynamics: Emerging Paradigms and Challenges: AMS-IMS-SIAM Joint Summer Research Conference on Modeling the Dynamics of Human Diseases: Emerging Paradigms and Challenges, July 17-21, 2005, Snowbird, Utah*. 2006. American Mathematical Soc.
58. Heemskerk, D., et al., *Tuberculosis in adults and children*. Vol. 2. 2015: Springer London.
59. Turner, R.D., et al., *Tuberculosis infectiousness and host susceptibility.* The Journal of infectious diseases, 2017. **216**(suppl_6): p. S636-S643.
60. Saunders, M.J., et al., *A score to predict and stratify risk of tuberculosis in adult contacts of tuberculosis index cases: a prospective derivation and external validation cohort study.* The Lancet Infectious Diseases, 2017. **17**(11): p. 1190-1199.
61. Trauer, J.M., et al., *The Importance of Heterogeneity to the Epidemiology of Tuberculosis.* Clinical Infectious Diseases, 2018. **69**(1): p. 159-66.
62. Keeling, M.J. and P. Rohani, *Modeling infectious diseases in humans and animals*. 2008: Princeton University Press.
63. Cohen, T. and P. White, *Transmission-dynamic models of infectious diseases.* Infectious Disease Epidemiology, 2016: p. 223.
64. McLean, A.R., *Infectious disease modeling*, in *Infectious Diseases*. 2013, Springer. p. 99-115.
65. Vynnycky, E. and R. White, *An introduction to infectious disease modelling*. 2010: Oxford University Press.
66. Kretzschmar, M. and J. Wallinga, *Mathematical models in infectious disease epidemiology*, in *Modern Infectious Disease Epidemiology*. 2009, Springer. p. 209-221.
67. Stoica, P. and T. Söderström, *On the parsimony principle.* International Journal of Control, 1982. **36**(3): p. 409-418.
68. Otto, S.P. and T. Day, *A biologist's guide to mathematical modeling in ecology and evolution*. 2011: Princeton University Press.

69. Coelho, M.T.P., J.A. Diniz-Filho, and T.F. Rangel, *A parsimonious view of the parsimony principle in ecology and evolution.* Ecography, 2019. **42**(5): p. 968-976.
70. Mossong, J., et al., *Social contacts and mixing patterns relevant to the spread of infectious diseases.* PLoS Med, 2008. **5**(3): p. e74.
71. Bansal, S., B.T. Grenfell, and L.A. Meyers, *When individual behaviour matters: homogeneous and network models in epidemiology.* Journal of the Royal Society Interface, 2007. **4**(16): p. 879-891.
72. Miller, J.C., A.C. Slim, and E.M. Volz, *Edge-based compartmental modelling for infectious disease spread.* Journal of the Royal Society Interface, 2012. **9**(70): p. 890-906.
73. Jacquez, J.A., et al., *Modeling and analyzing HIV transmission: the effect of contact patterns.* Mathematical Biosciences, 1988. **92**(2): p. 119-199.
74. Blythe, S. and R. Anderson, *Variable infectiousness in HFV transmission models.* Mathematical Medicine and Biology: A Journal of the IMA, 1988. **5**(3): p. 181-200.
75. Busenberg, S. and C. Castillo-Chavez, *A general solution of the problem of mixing of subpopulations and its application to risk-and age-structured epidemic models for the spread of AIDS.* Mathematical Medicine and Biology: A Journal of the IMA, 1991. **8**(1): p. 1-29.
76. Dodd, P.J., P.J. White, and G.P. Garnett, *Notions of synergy for combinations of interventions against infectious diseases in heterogeneously mixing populations.* Mathematical biosciences, 2010. **227**(2): p. 94-104.
77. Lanzas, C., et al., *The effect of heterogeneous infectious period and contagiousness on the dynamics of Salmonella transmission in dairy cattle.* Epidemiology & Infection, 2008. **136**(11): p. 1496-1510.
78. Waaler, H., A. Geser, and S. Andersen, *The use of mathematical models in the study of the epidemiology of tuberculosis.* American Journal of Public Health and the Nations Health, 1962. **52**(6): p. 1002-1013.
79. Colijn, C., T. Cohen, and M. MURRAY, *Mathematical models of tuberculosis: accomplishments and future challenges*, in *BIOMAT 2006*. 2007, World Scientific. p. 123-148.
80. Dowdy, D., C. Dye, and T. Cohen, *Data needs for evidence-based decisions: A tuberculosis modeler's 'wish list'.* International Journal of Tuberculosis and Lung Disease, 2013. **17**(7): p. 866-877.
81. Blower, S.M., et al., *The intrinsic transmission dynamics of tuberculosis epidemics.* Nature Medicine, 1995. **1**(8): p. 815-21.
82. Houben, R.M., et al., *How can mathematical models advance tuberculosis control in high HIV prevalence settings?* International Journal of Tuberculosis & Lung Disease, 2014. **18**(5): p. 509-14.
83. Bacaer, N., et al., *Modeling the joint epidemics of TB and HIV in a South African township.* Journal of Mathematical Biology, 2008. **57**(4): p. 557-93.
84. Lin, H.H., et al., *Effects of smoking and solid-fuel use on COPD, lung cancer, and tuberculosis in China: a time-based, multiple risk factor, modelling study.* Lancet, 2008. **372**(9648): p. 1473-83.
85. Basu, S., et al., *Projected effects of tobacco smoking on worldwide tuberculosis control: mathematical modelling analysis.* Bmj, 2011. **343**: p. d5506.
86. Bhunu, C., S. Mushayabasa, and J. Tchuenche, *A theoretical assessment of the effects of smoking on the transmission dynamics of tuberculosis.* Bulletin of mathematical biology, 2011. **73**(6): p. 1333-1357.

87. Moualeu, D., et al., *Analysis of the impact of diabetes on the dynamical transmission of tuberculosis.* Mathematical Modelling of Natural Phenomena, 2012. **7**(3): p. 117-146.
88. Pan, S.C., et al., *Effect of diabetes on tuberculosis control in 13 countries with high tuberculosis: a modelling study.* The Lancet Diabetes & Endocrinology, 2015. **3**(5): p. 323-30.
89. Abu-Raddad, L.J., et al., *Epidemiological benefits of more-effective tuberculosis vaccines, drugs, and diagnostics.* Proceedings of the National Academy of Sciences, 2009. **106**(33): p. 13980-13985.
90. Channing, L. and E. Sinanovic, *Modelling the cost-effectiveness of a new infant vaccine to prevent tuberculosis disease in children in South Africa.* Cost Effectiveness and Resource Allocation, 2014. **12**(1).
91. Aparicio, J.P. and C. Castillo-Chavez, *Mathematical modelling of tuberculosis epidemics.* Mathematical Biosciences & Engineering: MBE, 2009. **6**(2): p. 209-37.
92. Hickson, R., G. Mercer, and K. Lokuge. *Sensitivity analysis of a model for tuberculosis*. in *19th international congress on modelling and simulation*. 2011.
93. Hickson, R.I., G.N. Mercer, and K.M. Lokuge, *A metapopulation model of tuberculosis transmission with a case study from high to low burden areas.* PloS one, 2012. **7**(4): p. e34411.
94. Korthals Altes, H., et al., *Latent tuberculosis infection in foreign-born communities: Import vs. transmission in The Netherlands derived through mathematical modelling.[Erratum appears in PLoS One. 2018 May 24;13(5):e0198376; PMID: 29795693]*. PLoS ONE [Electronic Resource], 2018. **13**(2): p. e0192282.
95. Thomas, E.G., et al., *Modelling the spread of tuberculosis, including drug resistance and HIV: a case study in Papua New Guinea's Western Province.* The ANZIAM Journal, 2010. **52**(1): p. 26-45.
96. Dowdy, D.W., et al., *Impact of enhanced tuberculosis diagnosis in South Africa: a mathematical model of expanded culture and drug susceptibility testing.* Proceedings of the National Academy of Sciences, 2008.
97. Dowdy, D.W., et al., *The potential impact of enhanced diagnostic techniques for tuberculosis driven by HIV: a mathematical model.* AIDS, 2006. **20**(5): p. 751-62.
98. Dowdy, D.W., et al., *Population-level impact of same-day microscopy and Xpert MTB/RIF for tuberculosis diagnosis in Africa.* PLoS One, 2013. **8**(8): p. e70485.
99. Lin, H.-H., et al., *The impact of new tuberculosis diagnostics on transmission: why context matters.* Bulletin of the World Health Organization, 2012. **90**: p. 739-747.
100. Oxlade, O., et al., *Developing a tuberculosis transmission model that accounts for changes in population health.* Medical Decision Making, 2011. **31**(1): p. 53-68.
101. Pandey, S., et al., *Estimating tuberculosis incidence from primary survey data: a mathematical modeling approach.* The International Journal of Tuberculosis and Lung Disease, 2017. **21**(4): p. 366-374.
102. Sun, A.Y., et al., *Modeling the impact of alternative strategies for rapid molecular diagnosis of tuberculosis in Southeast Asia.* American journal of epidemiology, 2013. **178**(12): p. s1740-1749.
103. Houben, R., et al., *TIME Impact–a new user-friendly tuberculosis (TB) model to inform TB policy decisions.* BMC medicine, 2016. **14**(1): p. 56.
104. Menzies, N.A., et al., *Prospects for Tuberculosis Elimination in the United States: Results of a Transmission Dynamic Model.* American journal of epidemiology, 2018.
105. Menzies, N.A., et al., *Population health impact and cost-effectiveness of tuberculosis diagnosis with Xpert MTB/RIF: a dynamic simulation and economic evaluation.* PLoS medicine, 2012. **9**(11): p. e1001347.

106. Garcia, A., J. Maccario, and S. Richardson, *Modelling the annual risk of tuberculosis infection.* International Journal of Epidemiology, 1997. **26**(1): p. 190-203.
107. Dowdy, D.W., et al., *A user-friendly, open-source tool to project impact and cost of diagnostic tests for tuberculosis.* Elife, 2014. **3**: p. e02565.
108. Arregui, S., et al., *Data-driven model for the assessment of Mycobacterium tuberculosis transmission in evolving demographic structures.* Proceedings of the National Academy of Sciences of the United States of America, 2018. **115**(14): p. E3238-E3245.
109. Dowdy, D.W., et al., *Population-level impact of active tuberculosis case finding in an Asian megacity.* PLoS One, 2013. **8**(10): p. e77517.
110. Salomon, J.A., et al., *Prospects for advancing tuberculosis control efforts through novel therapies.* PLoS Medicine, 2006. **3**(8): p. e273.
111. Vynnycky, E., et al., *Tuberculosis control in South African gold mines: mathematical modeling of a trial of community-wide isoniazid preventive therapy.* American journal of epidemiology, 2015. **181**(8): p. 619-632.
112. WHO, *Global Tuberculosis Report 2019*. 2019: Geneva.
113. Stein, R.A., *Super-spreaders in infectious diseases.* International Journal of Infectious Diseases, 2011. **15**(8): p. e510-e513.
114. Curtis, A.B., et al., *Extensive transmission of Mycobacterium tuberculosis from a child.* New England Journal of Medicine, 1999. **341**(20): p. 1491-1495.
115. Heemskerk, D., et al., *Clinical Manifestations*, in *Tuberculosis in Adults and Children*. 2015, Springer. p. 17-26.
116. Escombe, A.R., et al., *The infectiousness of tuberculosis patients coinfected with HIV.* PLoS medicine, 2008. **5**(9): p. e188.
117. DepartmentofHealth&HumanServices, *Management, control and prevention of tuberculosis: Guidelines for health care providers*. 2015, State of Victorias: Victoria, Australia.
118. Services, D.o.H.H. *Mycobacterial infections (tuberculosis)*. 2017 [cited 2017 April 07]; Available from: https://www2.health.vic.gov.au/public-health/infectious-diseases/disease-information-advice/tuberculosis.
119. Moyo, N., E. Tay, and J. Denholm, *Evaluation of tuberculin skin testing in tuberculosis contacts in Victoria, Australia, 2005–2013.* Public health action, 2015. **5**(3): p. 188-193.
120. RCoreTeam, *R: A Language and Environment for Statistical Computing*. 2014, R Foundation for Statistical Computing: Vienna, Austria.

Chapter 2

2. Risk factors for infectiousness of patients with tuberculosis: a systematic review and meta-analysis

2.1. Chapter overview

Having described the rationale and objectives of the overall thesis, this Chapter presents a study synthesising evidence on the variation among patients with TB as to their capacity to transmit the infection. To identify risk factors for the infectiousness of patients with TB, this Chapter presents a systematic review and meta-analyses of published TB contact investigation studies. The Chapter addresses thesis question 1 by identifying patient-related factors associated with the infectiousness and provides associated risk ratios. Importantly, this Chapter provides information on the characteristics of highly infectious individuals with TB that could be used in targeting interventions such as contact tracing and provision of preventive treatment. It also identifies the gaps in understanding heterogeneous infectiousness, which is the focus of subsequent chapters of the thesis.

Citation: **Melsew YA**, Doan TN, Gambhir M, Cheng AC, McBryde E, Trauer JM. Risk factors for infectiousness of patients with tuberculosis: a systematic review and meta-analysis. *Epidemiology & Infection*. 2018 Feb;146(3):345-53.

Epidemiology and Infection

cambridge.org/hyg

Risk factors for infectiousness of patients with tuberculosis: a systematic review and meta-analysis

Y.A. Melsew[1,2], T.N. Doan[3,4], M. Gambhir[1], A.C. Cheng[1,5], E. McBryde[3,4] and J.M. Trauer[1]

[1]School of Public Health and Preventive Medicine, Medicine, Nursing and Health Sciences, Monash University, Melbourne, Australia; [2]Department of Epidemiology and Biostatistics, Institute of Public Health, University of Gondar, Gondar, Ethiopia; [3]Australian Institute of Tropical Health and Medicine, James Cook University, Townsville, Australia; [4]Department of Medicine at the Royal Melbourne Hospital, University of Melbourne, Melbourne, Australia and [5]Infection Prevention and Healthcare Epidemiology Unit, Alfred Health, Melbourne, Australia

Original Papers

Cite this article: Melsew YA, Doan TN, Gambhir M, Cheng AC, McBryde E, Trauer JM (2018). Risk factors for infectiousness of patients with tuberculosis: a systematic review and meta-analysis. *Epidemiology and Infection* **146**, 345–353. https://doi.org/10.1017/S0950268817003041

Received: 28 May 2017
Revised: 28 November 2017
Accepted: 5 December 2017
First published online: 17 January 2018

Key words:
Patient infectiousness; risk factors; systematic review; tuberculosis

Author for correspondence:
Y.A. Melsew, E-mail: 078yayu@gmail.com or yayehirad.melsew@monash.edu

Abstract

We performed a systematic review and meta-analyses of studies assessing tuberculosis (TB) patient-related risk factors for transmission of *Mycobacterium tuberculosis* infection. Meta-analyses were conducted for sputum smear-positivity, lung cavitation and HIV seropositivity of index patients with both crude and adjusted odds ratios (AORs) pooled using random effect models. Thirty-seven studies were included in the review. We found that demographic characteristics such as age and sex were not significant risk factors, while behaviours such as smoking and alcohol intake were associated with infectiousness although inconsistently. Treatment delay of >28 days was a significant predictor of greater infectiousness. Contacts of sputum smear-positive index patients were found to be more likely to be infected than contacts of sputum smear-negative patients, with a pooled AOR of 2.15 (95% confidence interval (CI) 1.47–3.17, $I^2 = 38\%$). Similarly, contacts of patients with the cavitary disease were around twice as likely to be infected as contacts of patients without cavitation (pooled AOR 1.9, 95% CI 1.26–2.84, $I^2 = 63\%$). In contrast, HIV seropositive patients were associated with few contact infections than HIV seronegative patients (AOR 0.45, 95% CI 0.26–0.80, $I^2 = 52\%$). In conclusion, behavioural and clinical characteristics of TB patients can be used to identify highly infectious patients for targeted interventions.

Introduction

Despite recent improvements following the successful implementation of the Millennium Development Goals, tuberculosis (TB) is still the leading cause of mortality due to a single infectious pathogen [1].

The transmission of infectious diseases is affected by characteristics of the environment, infectious agent, and host, including both the infecting and infected hosts. For directly transmissible diseases, understanding the characteristics of the index patient that influence the capacity to transmit infection is essential to design the effective control strategies [2, 3]. The spread of TB is strongly affected by the infectious individual's characteristics, with some individuals causing considerably greater numbers of new infections, while others produce very few or none [4].

A central goal of TB control strategies is the prevention of transmission. Hence, identifying highly contagious TB patients, by analysing risk factors from contact investigation studies, is critical for designing effective and efficient TB control interventions. We performed a systematic review of patient factors associated with the risk of transmission of TB to contacts.

Methods

The review protocol was prospectively registered with PROSPERO (International Prospective Register of Systematic Reviews, registration number: CRD42016053913) and we adhered to the guidelines of Meta-analyses and Systematic Review of Observational studies in Epidemiology (MOOSE) throughout [5].

Search strategy

We searched the following electronic bibliographic databases from inception to 3 January 2017: MEDLINE, EMBASE, Web of Science, Global Health and Global Health Archive. All

search terms were 'exploded' to include all resources and consisted of 'Tuberculosis', '*Mycobacterium tuberculosis*', 'TB', 'contact', 'contact investigation', 'transmission', 'index case', 'infectiousness', 'risk factors', 'secondary case' and 'latent infection'. The detailed search strategy is provided in Table S1.

Study selection

Search results were exported to EndNoteX8 (Clarivate Analytics, NY, USA) and duplicates were removed. Titles and abstracts were screened to identify potentially relevant articles. The outcome of interest was latent TB infection (LTBI) or TB disease among contacts of index TB patients. The full texts of the studies that were deemed relevant in this initial screening phase were then obtained and reviewed by two authors (Y. M. and T. D.). Studies were included if they diagnosed index patients according to the clinical, radiological or microbiological criteria, assessed LTBI or TB disease among contacts, reported the number of index patients and contacts, and quantified index patient-related risk factors for infection or disease in their contacts. Editorials, commentaries, case reports, outbreak reports, systematic reviews, mathematical modelling studies and studies in languages other than English were excluded, as were studies that did not assess index patient-related risk factors.

Data extraction

The following data were extracted: country, year, study design, index patient definition, contact definition, outcome status of contacts (LTBI or TB disease), number of index patients, number of contacts, study population, contact screening methods, results of screening and index risk factors investigated. Estimates of the effect of exposures on the outcomes of the proportion of contacts with active TB or with LTBI were also extracted, including adjusted risk ratio or odds ratio and their 95% confidence intervals (CIs). Y. M. and T. D. extracted the data independently using the same data extraction form with all discordant data resolved by consensus.

Assessment of study quality

Each article was evaluated using the Newcastle–Ottawa Scale [6] for its validity with respect to the selection of study participants, comparability between exposure groups and outcome measurements, by two authors (Y. M. and T. D.) (Table S2).

Meta-analyses

Exposure variables were selected based on the availability of sufficient data and were not specified *a priori*, three variables were chosen for meta-analysis based on importance and consistent availability: sputum smear-positivity, lung cavitation and HIV seropositivity. Meta-analyses were performed with MetaXL (Version 5.3, EpiGear International Pty Ltd, QLD, Australia) to estimate the pooled effects of each of these three exposures on the binary outcomes of contact LTBI or TB disease. Random effects models were used to estimate pooled effect measures. Heterogeneity among estimates was quantitatively evaluated with the I^2-statistic.

Results

Search results

The search strategy resulted in 4780 references, of which 1725 were duplicates, 321 were deemed eligible for full-text review and 37 studies were included in the systematic review. The detailed flow diagram of study selection is presented in Fig. S1. Of the included studies, 17 were cross-sectional, seven cohort studies, 12 follow-up studies without control groups and one case–control study. Twenty-four studies originated from one of the 30 high burden TB countries.

Contact and index definitions

Studies varied by the types of index patients and contacts included. Eight defined index patients as newly diagnosed smear-positive pulmonary TB patients, 12 studies considered confirmed pulmonary TB and six included all types of TB. With regard to contact definition, 20 studies considered only household contacts, six studies traced only children and adolescent contacts and three considered only close contacts who shared enclosed spaces (Table S3).

Outcome measures

The outcome of interest was LTBI or TB disease amongst contacts of different index patients, which was used as a marker of patients' infectiousness; that is the capacity to transmit *Mycobacterium tuberculosis* (*M. tb*) to contacts. The included studies compared contacts who developed the outcome (LTBI or TB disease) to those who did not, based on the characteristics of index patients as the exposure variable. Thus, the measures of association (risk ratio or odds ratio) quantified the effect of exposure to an index patient with a given characteristic on the probability or odds of infection or disease in the exposed contact. The majority of included studies (29/37) considered LTBI as the main outcome, with 25 using tuberculin skin test (TST), one using interferon-gamma release assay (IGRA) and three using both TST and IGRA to determine infection. All studies using IGRAs to define infection used the manufacturer's recommended cut-off assessed at a single time point to define infection. Studies using TST mostly used a cut-off $\geqslant$10 mm for HIV negative and $\geqslant$5 mm for HIV positive contacts, although six studies used a cut-off of $\geqslant$5 mm for all contacts and two other studies used 15 mm as the cut-off point for children having BCG vaccination. The remaining eight studies used the development of active TB in contacts as the measure of transmission (Table S4).

Index patient risk factors for M. tb transmission

The included studies considered a range of characteristics of index patients for their association with transmissibility. We categorised these index risk factors as demographic, behavioural, co-morbidities, TB-related clinical and household characteristics. We preferentially used the adjusted association reported from multivariate analyses. We also classified studies based on the type of association they had reported as: positive, negative or non-significant based on a *P*-value of <0.05 or a 95% CI not including unity.

https://doi.org/10.1017/S0950268817003041

Sociodemographic risk factors

Sex of index patients was a non-significant factor in most studies reporting its association with transmissibility [7–17], although two studies found female sex to be associated with greater risk of infection [18, 19]. With one exception [13], included studies considered adult index patients exclusively. Of these, most studies that reported the effect of age found a non-significant association [7, 9, 12, 14, 15, 20], although two studies reported that increasing age was positively associated with *M. tb* transmission [17, 21]. Of these two studies, one was conducted in France among contacts aged under 17 and found that contacts of patients aged 40 years old or above were more likely to be infected than contacts of younger patients [21]. By contrast, a study from the USA reported that contacts of patients aged 64 years or above were less likely to be infected than contacts of younger patients [17].

One study from the Netherlands, which considered both close and casual contacts found that contacts of index patients who were employed or attending school were more likely to be infected [13], while two studies from Peru considering household contacts only found that employment status was not a significant risk factor for *M. tb* transmission [9, 14]. Black race of the index patient was a positive risk factor in one study in the USA which considered all close contacts [15], while another study of contacts aged under 15 years in the USA found race was not associated with contact infection [17] (Table 1).

Behavioural factors and co-morbidities

Index patients' smoking status was a significant risk factor for LTBI among contacts in three studies [12, 22, 23], but was non-significant in four [9, 10, 23, 24]. In one study from Peru, smoking more than one cigarette per day was associated with TB transmission to child contacts, but was not significant for transmission to adults. This study also reported that heavy alcohol consumption (compared with no alcohol intake) was associated with risk of contact infection [23], although other studies [8–10, 12] found no significant association. Similarly, injection drug use was associated with higher TB transmission in a case–control study from Spain [25], but this risk factor was non-significant in others [12, 15] (Table 2).

TB-related clinical risk factors

Differences in the index patients' clinical presentation were considered as potential risk factors for *M. tb* transmission to contacts by several studies (Table 3). Compared with contacts of index patients who did not cough, contacts of index patients with any cough were more likely to be infected in two studies [13, 26], while another two studies found no significant association [7, 27]. Of studies considering the duration of coughing, two studies found that contacts of patients coughing for 2 weeks or longer were more likely to be infected than contacts of patients who coughed for a shorter period of time or who did not cough [13, 19], while a study reported coughing for 5 weeks or longer to be a significant risk factor [28]. In contrast, three studies found that duration of a cough, considered as a continuous variable was not statistically significant [9, 24, 29].

Treatment delay was another risk factor considered for *M. tb* transmission, although studies considered different cut-off points for categorising a number of days of treatment delay. In three studies [7, 19, 30] a treatment delay of 30 days or more (compared with <30 days) was associated with greater risk of infection and disease among contacts. Similarly, compared with no treatment delay, delay for more than 28 days [11] was a risk factor in one study, while in two other studies [15, 30] delay of 90 days or more was significant. Another study reported the duration of illness considered as a continuous variable to be a significant factor for contact LTBI [31].

Sputum smear-positivity was reported to be associated with *M. tb* transmission by ten studies [12, 13, 15, 22, 26, 27, 32]. Other studies compared patients based on the bacilli density per field, although categorisation was inconsistent. Sputum smear bacilli density of 2+ [11, 23, 33] and 3+ [7, 11, 23, 34] compared with scanty were significant risk factors in all studies reporting this variable. Similarly, one study [21] reported very high sputum smear densities ⩾100 bacilli/field as a significant

Table 1. Socio-demographic risk factors for infectiousness of index patients extracted from included studies

Risk factors	Positive association study, AOR (95% CI)	Negative association study, AOR (95% CI)	Non-significant association
Female sex	5.0 (2.0–12.2) [18] 2.8 (1.3–6.1) [19]		[7–17]
All age categories			[9, 7, 12, 15, 14, 20]
Age <64 *vs.* 65 and older		1.2[a] [17]	[12]
Age >40 years *vs.* age under 40 years	9.9 (1.8–53.9) [21]		
Currently employed/working			[9, 14]
Currently studying	1.8 (1.2–2.5)[b] [13]		[9, 14]
Education level			[9, 14, 20]
Immigrant		0.6 (0.5–0.9)[b] [13]	[12]
Black race	3.0 (1.0–9.4) [15]		
All races			[17][c]

[a]Confidence interval not provided.
[b]Adjusted relative risk.
[c]Considered as a categorical variable.
AOR, adjusted odds ratio; CI, confidence interval.

https://doi.org/10.1017/S0950268817003041

Table 2. Behavioural and co-morbidity risk factors for infectiousness of index patients extracted from included studies

Risk factors	Positive association study, AOR (95% CI)	Negative association study, AOR (95% CI)	Non-significant association
Smoking	1.5 (1.3–1.7) [12] 1.4 (1.0–2.0)[a] [23] 2.7 (1.5–4.7) [22]		[9, 10, 23, 24]
Heavy alcohol consumption	1.2 (1.0–1.5)[a] [23]		[8, 9, 10, 12]
Injection drug users	4.2 (1.1–16.0) [25][b]		[12, 15]
BMI ⩾18.5			[10]
BCG vaccination			[10]
Diabetes mellitus		0.2 (0.1–0.7) [9][b]	[10]
History of incarceration			[9]
Comorbidities excluding HIV			[14]
HIV sero-positivity	2.1 (1.2–3.8) [25][b]	0.2 (0.1–0.9) [29] 0.1 (0.03–0.5) [37] 0.5 (0.3–0.9) [33]	[9, 11, 12, 14, 16, 32, 48, 49]
CD4 count			[11]

[a]Adjusted relative risk.
[b]Indicates that these studies take contact active TB disease as an outcome.
AOR, adjusted odds ratio; CI, confidence interval.

risk factor. Sputum production and haemoptysis were found to be significant risk factors in one study [13].

Lung cavity lesion findings were associated with greater infectiousness in eight studies [12, 13, 15, 19, 21, 35, 36], but non-significant in another nine studies [8, 9, 11, 16, 20, 23, 24, 26, 29, 31]. The number of lung zones involved on CXR was also reported as a risk factor in one study, with the involvement of four or more zones associated with more contact infections than those with fewer zones involved [28].

Of the four studies which compared transmission based on drug sensitivity, one [9] reported contacts of multidrug-resistant TB (MDR-TB) patients were less likely to be infected, while three [7, 17, 37] reported no significant difference.

Household risk factors

A history of TB among family members was found to be a risk factor for contact TB disease in two studies [24, 36], although this was non-significant in one more recent study [7]. The type of residence of the index patient was also reported as a significant factor, with those living in slums or apartments associated with more infections among contacts [19]. Household crowding with two or more persons per room was a significant risk factor for *M. tb* transmission in one study [19] although two other studies found no significant effect [24, 29].

Meta-analysis

Three factors were reported in a sufficient number of studies to permit meaningful meta-analysis. Twenty-six studies were included in one or more of these meta-analyses, with three factors of interest being sputum smear-positive disease, lung cavitation and HIV status of the index patient (Table S5).

Contacts of sputum smear-positive pulmonary TB patients were found to be more frequently infected than contacts of smear-negative patients, with a pooled adjusted odds ratio (AOR) of 2.15 (95% CI 1.47–3.17, $I^2 = 38\%$) (Fig. 1). A similar association was observed from the pooled crude odds ratios (2.29, 95% CI 1.84–283, $I^2 = 54\%$) (Fig. S2). Sensitivity analysis indicated that no single study, when excluded, affected the pooled estimates (Table S6), although funnel plots suggested possible publication bias (Fig. S5).

The odds of infection among contacts of patients with cavitation also appeared to be around two, although heterogeneity between study estimates was substantial. Our estimate of the pooled AOR is 1.90 (95% CI 1.26–2.84, $I^2 = 63\%$) (Fig. 2) and 1.93 (95% CI 1.36–2.75, $I^2 = 85\%$) for the pooled crude odds ratio (Fig. S3). In sensitivity analyses, excluding a cross-sectional study from Thailand [19] lowered the pooled AOR to 1.64, while excluding other studies did not markedly change our point estimate. In the crude odds ratios analysis, exclusion of any individual study did not result in a marked change to our pooled estimate (Table S7) and publication bias was not revealed by funnel plots (Fig. S6).

HIV seropositivity of index patients was negatively associated with infectiousness in the AORs analysis (0.45, 95% CI 0.26–0.80, $I^2 = 52\%$) (Fig. 3). However, this association was non-significant in the analysis of crude odds ratio (0.76, 95% CI 0.55–1.07, $I^2 = 80\%$) (Fig. S4). Sensitivity analysis of the adjusted estimate found that excluding any one of the studies did not change the pooled estimate significantly (Table S8) and the funnel plot of studies revealed no evidence of publication bias (Fig. S7).

Discussion

This systematic review and meta-analyses assessed the capacity of index TB patients to transmit the infection to their contacts based on their demographic, behavioural and clinical characteristics. Our meta-analysis showed that sputum smear-positivity, the presence of lung cavitation and HIV seronegativity were the factors most consistently associated with greater infectiousness of index cases. Our results were also broadly consistent with the premise that more severe lung parenchymal involvement being associated with greater infectiousness. We found no sociodemographic

https://doi.org/10.1017/S0950268817003041

Table 3. Index tuberculosis disease-related clinical risk factors for infectiousness of index patients extracted from included studies

Risk factors	Positive association study, AOR (95% CI)	Negative association study, AOR (95% CI)	Non-significant association
Chest radiography			
Cavitary lesion	1.6 (1.4–1.8) [12] 2.0 (1.6–2.6)[a] [13] 3.1 (1.3–7.3) [15] 4.4 (2.4,8.1) [19] 6.3 (1.2–32.9) [21] 1.2 (0.6–2.5) [36][b] 2.2 (1.1–4.4) [35][b]		[8, 10, 11, 16, 20, 23, 24, 26, 29, 31]
4–5 lung zones involved *vs.* 3 or less	2.1 (1.0–4.4) [28]		
Coughing characteristics			
Any coughing	4.2 (2.1–8.6)[a] [13] 2.8 (1.3–6.1) [26]		[7, 27]
Duration of a cough in weeks			[9, 29]
Coughing for ⩾4 weeks			[24]
Coughing for >2 weeks	3.0 (1.8–4.9) [19] 4.2 (1.6–11.2) [36][b]		
Coughing for 10 weeks	2.7 (1.3–5.6) [28]		
Treatment delay			
Treatment delay >30 days	1.8 (1.2–2.7) [7][b] 1.8 (1.1–3.0) [19] 1.9 (1.2–2.99) [30]		[12, 33]
Treatment delay <14 days			[11, 23]
Treatment delay 14–28 days			[11, 23]
Treatment delay >28	1.3 (1.1–1.7)[a] [11]		[23]
Treatment delay 60–90 days	2.4 (1.6–4.1) [30]		
Treatment delay ⩾90 days	2.3 (1.1–5.1) [15] 2.3 (1.5–3.6) [30] 8.0 (3.4–18.6) [19]		
Time sick in days	1.01 (1.001–1.011) [31]		[10] [37]
Sputum characteristics			
Sputum smear positive	1.7 (1.3–2.2) [12] 1.9 (1.2–3.0)[a] [13] 3.3 (1.3–8.3) [15] 2.7 (1.5–4.7) [22] 1.7 (1.3–2.3) [32] 2.2 (1.1–4.9) [26] 6.1 (1.4–26.5) [27]		[16, 29]
Sputum smear density ⩾2+			[24]
Sputum smear density 1+			[9, 11]
Sputum smear density 2+	1.4 (1.2–1.8)[a] [23] 5.9 (1.6–21.3) [33] 1.2 (1.0–1.4)[a] [11]		[19, 20]
Sputum smear density 3+	1.5 (1.2–1.8)[a] [23] 1.2 (1.1–1.4)[a] [11] 1.5 (1.0–2.3) [7][b] 2.8 (1.6–4.8) [34]		[8, 9, 14, 18, 20]
Sputum smear density >100 bacilli/field	5.9 (1.6–22.0) [21]		
Sputum appearance purulent			[8]
Sputum volume ⩾5 ml			[8]
Any sputum production	1.8 (1.3–2.5)[a] [13]		

(*Continued*)

Table 3. (*Continued.*)

Risk factors	Positive association study, AOR (95% CI)	Negative association study, AOR (95% CI)	Non-significant association
Haemoptysis	1.7 (1.2–2.3)[a] [13]		
Sputum culture positive			[24] [33]
Sputum PCR positive	3.8 (1.1–13.6)[a] [13]		
Culture positive after 3 months of treatment			[36]
Broncho-alveolar lavage AFB positive	1.7 (1.1–2.6)[a] [13]		
Type of TB disease			
Multidrug-resistant TB		0.6 (0.3–0.9) [9][b]	[7, 17, 37]
Hospitalisation by current illness			[9]
Side effects of medication			[9]
Past history of active TB			[9, 14]
Prior TB treatment outcome not cured	3.3 (1.2–8.8) [50]		
Pulmonary TB	2.1 (1.1–4.0) [12] 4.5 (2.1–9.5)[a] [13]		

[a]Adjusted relative risk.
[b]Outcome considered was active TB disease in contacts.
AOR, adjusted odds ratio; CI, confidence interval.

characteristics that were consistently associated with the infectiousness of patients. Clinical factors were studied and reported in a variety of ways, but generally consistent with the message that prolonged symptoms such as coughing and delay to diagnosis lead to increased infectiousness. Treatment delays of >28 days were found to be a predictor of a greater proportion of contact infections.

Although most studies found no significant association, smoking and alcohol intake were found to be risk factors for greater infectiousness in some studies. A possible explanation could be that these risk factors are associated with both biological changes in the host [38], as well as differences in social mixing behaviours such as sharing enclosed bars [39], such the association is likely to be setting specific. In addition, exposure to second-hand smoke increases the susceptibility of contacts to infection, particularly among children [40, 41].

Compared with smear-negative patients, smear-positive patients were found to be approximately twice as infectious. This association was not confounded by cavitation since all studies but one included in the pooled analysis had adjusted for this factor [27]. The positive association between smear-positivity and infectiousness is likely to reflect smear-positive patients expelling a greater number of *M. tb* into the environment than smear-negative patients, as *M. tb* bacilli are only detectable with sputum microscopy when the concentration of the bacilli is sufficient [42]. Similarly, index patients presenting cavitation were found to be twice as infectious as those without cavitation. This is consistent with the widely held belief that cavitary TB is the

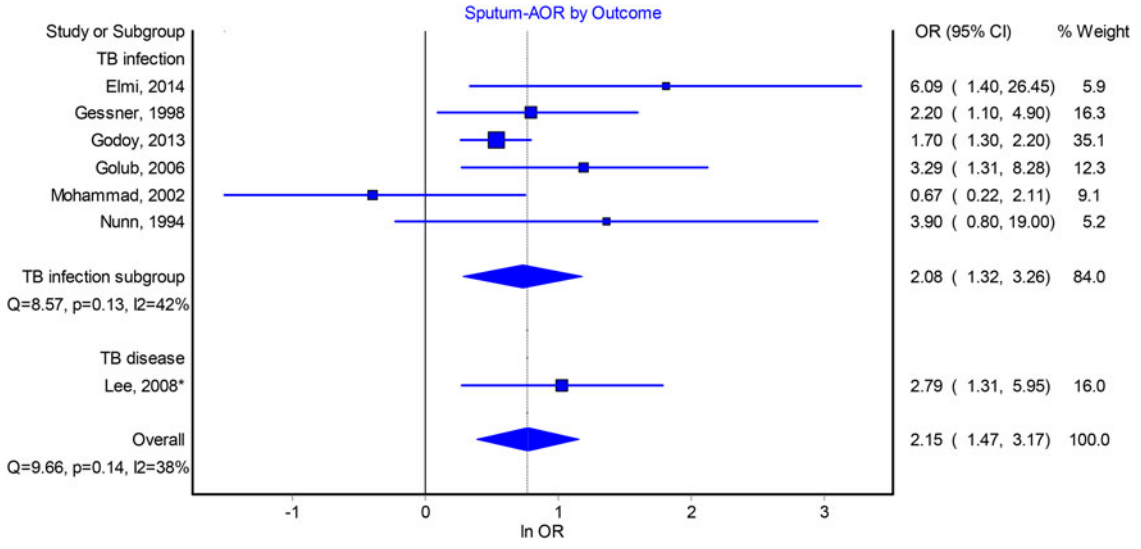


Fig. 1. Forest plot of adjusted odds ratios for the association between index patient sputum smear-positivity and infectiousness.

https://doi.org/10.1017/S0950268817003041

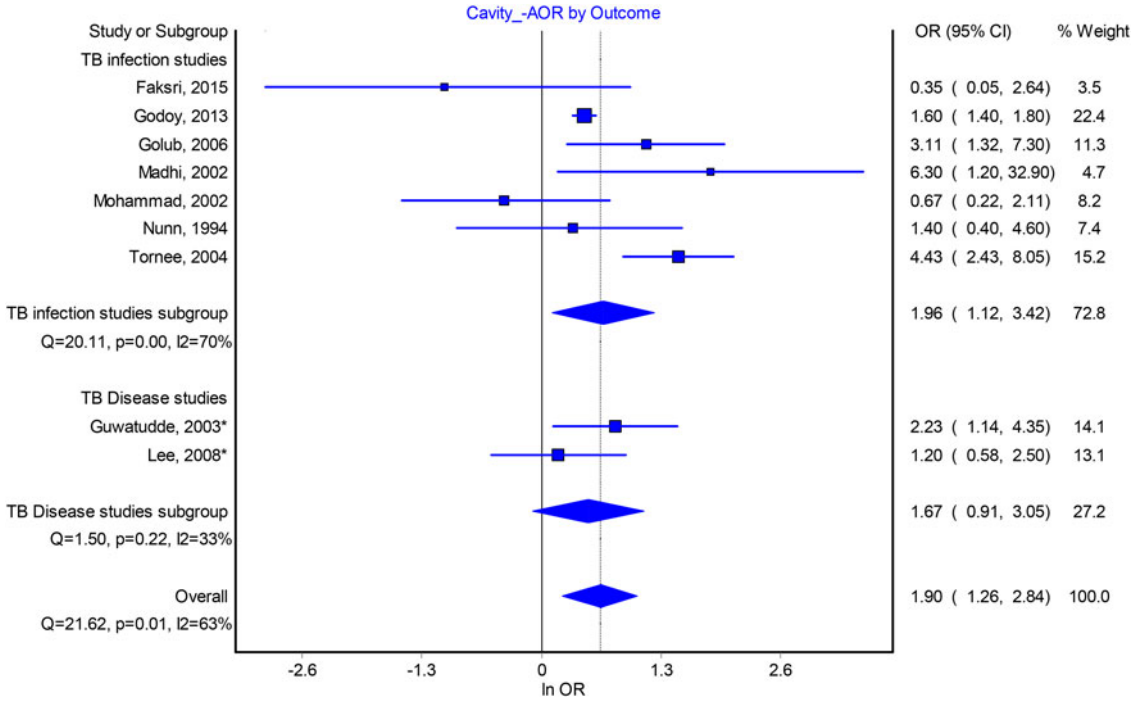


Fig. 2. Forest plot of the association between index patient lung cavitation and infectiousness from adjusted odds ratio analysis.

consequence of lung tissue damage and reflects the development of macroscopic openings between the site of infection and the airways, which facilitate expectoration of the bacteria [43].

In contrast, we found that patients living with HIV were around half as likely to transmit to their contacts, after adjusting for other factors. Notably, all such studies involved patients with untreated HIV. A previous meta-analysis found no significant difference between HIV negative and positive index patients with regards to their infectiousness, although it should be noted that this study only pooled crude odds ratios or relative risks [44], in contrast to our study which considered estimates adjusted for potential confounders. The lower infectiousness of HIV-positive index patients may be due to their less frequent cavitation (due to suppression of the immune response) [43, 45] or patterns of social mixing. Of the five studies included in this analysis, two [16, 29] adjusted for cavitation and one of which found HIV status non-significant [16]. HIV-positive patients may also be diagnosed and treated for active TB more rapidly due to fast disease progress or their closer contact with health care services, which may decrease their infectious period. It has previously been found that HIV-positive TB patients are less likely to be smear positive than HIV-negative TB patients [46, 47]. However, four of the five studies included in our meta-analysis adjusted for sputum smear status, such that the association of HIV seropositivity with TB transmission does not appear to be mediated by sputum smear-positivity. Therefore, HIV-positive patients may have a modestly reduced level of infectiousness but should not be presumed to be non-infectious, which is consistent with extremely high disease burden in high HIV-burden countries. As such, an equal focus of contact investigation has been advised for both HIV-seropositive and negative index patients [48].

The funnel plots suggest possible publication bias, particularly for studies that examined the association of infectivity and smear positivity. Other potential sources of bias result from the observational nature of studies included in this review, including selection bias and information bias. Non-differential misclassification bias

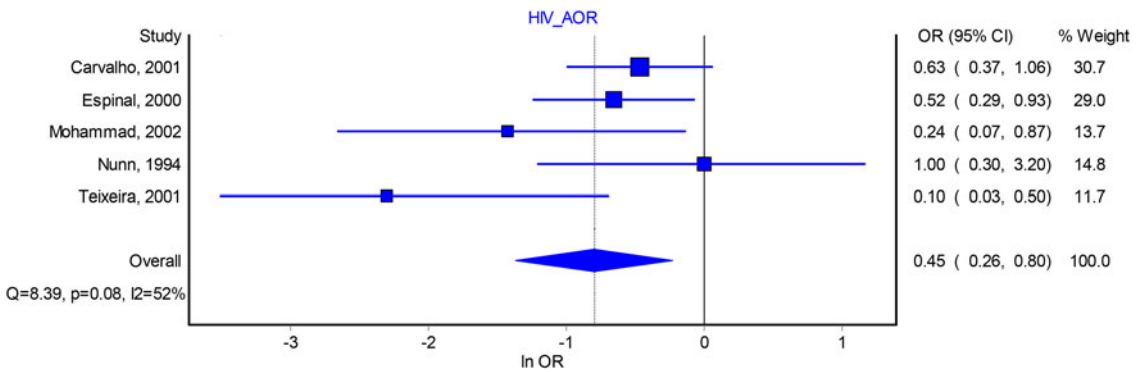


Fig. 3. Forest plot of the association between index patient HIV seropositivity and infectiousness from adjusted odds ratio analysis. *Indicates that these studies take contact active disease as an outcome. AOR, adjusted odds ratio; OR, odds ratio; CI, confidence interval.

https://doi.org/10.1017/S0950268817003041

could arise due to the inability of TST to differentiate recent LTBI attributable to the index exposure from distant past infection. This would dilute the effect, leading to an under-estimate and a bias towards the null. This may particularly be seen in studies with broader definitions of contacts, where distant latent infection might represent a greater proportion of all latent infection. By contrast, studies employing stricter definitions and focusing on those with greater exposure to the index case, such as close contacts only, would provide more accurate estimates, although this may be offset by lower precision due to smaller sample sizes. Studies using stricter definitions of infection (e.g. higher TST cut-offs) would also obtain more accurate effect size estimates, but at the expense of precision. Duration of follow-up differed between studies, with some assessing their study group at only a single point in time, which may affect the development of an infection response and its detection. Such insufficient duration of follow up could similarly lead to underestimation of the outcome measures, through greater dilution from distant infection. Studies considering less infectious index patients, such as those that included smear-negative patients, could also have a greater proportion of disease attributable to remote infection and a bias towards the null.

We have reviewed the outcomes of infection and subsequent disease together under the implicit assumption that both of these outcomes reflect index patient infectiousness. This is only correct if we consider that contacts have a similar risk of reactivation following infection regardless of the characteristics of the index patient. We believe this is reasonable because the risk of reactivation should be primarily attributable to contact characteristics (e.g. age, comorbidities) and because pooling of studies considering active disease with those considering LTBI did not increase statistical heterogeneity, although it could also be influenced by characteristics of the organism and extent of exposure. However, we have carefully indicated which outcome is being considered throughout and presented disaggregated meta-analysis results. Last, it should be noted that we are considering infectiousness as the proportion of contacts infected or diseased and have not considered the absolute number of infections generated, which may also reflect infectiousness to some extent.

Conclusions

We identified three clinical characteristics that can be used to identify TB patients with a greater capacity for transmission; smear positivity, the presence of cavitation on chest X-ray and HIV status. Thus, in an era of increasing use of molecular diagnostics, traditional diagnostic methods, such as sputum smear microscopy and chest radiography, still provide important information on patients' infectiousness. HIV-positive TB patients not under treatment may be modestly less infectious, although such patients should not be presumed to be non-infectious. Sociodemographic and behavioural characteristics were less consistently associated with infectiousness and their importance may be setting-dependent. The public health implication of these findings is in the potential to provide close follow-up and targeted interventions for these highly infectious TB patients.

Supplementary Material. The supplementary material for this article can be found at https://doi.org/10.1017/S0950268817003041

Acknowledgements. The authors thank the Monash University for providing access to electronic databases. Y. A. Melsew is a recipient of a Monash Graduate Scholarship for his PhD study. No specific funding was received for this study.

Declaration of Interest. The authors declare that they have no competing interests.

References

1. **World Health Organization** (2015) *A Global Action Framework for TB Research in Support of the Third Pillar of WHO's end TB Strategy*. Geneva. WHO/HTM/TB/2015.26.
2. **Woolhouse ME, *et al.*** (1997) Heterogeneities in the transmission of infectious agents: implications for the design of control programs. *Proceedings of the National Academy of Sciences of the United States of America* **94**(1), 338–342.
3. **Galvani AP and May RM** (2005) Epidemiology: dimensions of superspreading. *Nature* **438**(7066), 293–295.
4. **Wang C-S, *et al.*** (2008) Predictors for identifying the most infectious pulmonary tuberculosis patient. *Journal of the Formosan Medical Association* **107**(1), 13–20.
5. **Stroup DF, *et al.*** (2000) Meta-analysis of observational studies in epidemiology: a proposal for reporting. *Jama* **283**(15), 2008–2012.
6. **Wells G, *et al.*** (2000) The Newcastle–Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses. Available from: http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp (accessed 4 January 2017).
7. **Otero L, *et al.*** (2016) A prospective longitudinal study of tuberculosis among household contacts of smear-positive tuberculosis cases in Lima, Peru. *BMC Infectious Diseases* **16**(259), 1–8.
8. **Jones-Lopez EC, *et al.*** (2016) Cough Aerosols of *Mycobacterium tuberculosis* in the prediction of incident tuberculosis disease in household contacts. *Clinical Infectious Diseases* **63**(1), 10–20.
9. **Grandjean L, *et al.*** (2015) Transmission of multidrug-resistant and drug-susceptible tuberculosis within households: a prospective cohort study. *PLoS Medicine* **12**(6), e1001843.
10. **Faksri K, *et al.*** (2015) Transmission and risk factors for latent tuberculosis infections among index case-matched household contacts. *Southeast Asian Journal of Tropical Medicine & Public Health* **46**(3), 486–495.
11. **Huang CC, *et al.*** (2014) The effect of HIV-related immunosuppression on the risk of tuberculosis transmission to household contacts. *Clinical Infectious Diseases* **58**(6), 765–774.
12. **Godoy P, *et al.*** (2013) Immigrants do not transmit tuberculosis more than indigenous patients in Catalonia (Spain). *Tuberculosis* **93**(4), 456–460.
13. **Lohmann EM, *et al.*** (2012) Grading of a positive sputum smear and the risk of *Mycobacterium tuberculosis* transmission. *International Journal of Tuberculosis and Lung Disease* **16**(11), 1477–1484.
14. **Grandjean L, *et al.*** (2011) Tuberculosis in household contacts of multidrug-resistant tuberculosis patients. *International Journal of Tuberculosis and Lung Disease* **15**(9), 1164–1169.
15. **Golub JE, *et al.*** (2006) Delayed tuberculosis diagnosis and tuberculosis transmission. *International Journal of Tuberculosis & Lung Disease* **10**(1), 24–30.
16. **Nunn P, *et al.*** (1994) The effect of human immunodeficiency virus type-1 on the infectiousness of tuberculosis. *Tubercle & Lung Disease* **75**(1), 25–32.
17. **Snider DE Jr, *et al.*** (1985) Infection and disease among contacts of tuberculosis cases with drug-resistant and drug-susceptible bacilli. *American Review of Respiratory Disease* **132**(1), 125–132.
18. **Kifai EJ and Bakari M** (2009) Mantoux skin test reactivity among household contacts of HIV-infected and HIV un-infected patients with sputum smear positive TB in Dar es Salaam, Tanzania. *East African Journal of Public Health* **6**(2), 211–218.
19. **Tornee S, *et al.*** (2004) Risk factors for tuberculosis infection among household contacts in Bangkok, Thailand. *Southeast Asian Journal of Tropical Medicine & Public Health* **35**(2), 375–383.
20. **Elliott AM, *et al.*** (1993) The impact of HIV on infectiousness of pulmonary tuberculosis: a community study in Zambia. *AIDS* **7**(7), 981–987.
21. **Madhi F, *et al.*** (2002) Transmission of tuberculosis from adults to children in a Paris suburb. *Pediatric Pulmonology* **34**(3), 159–163.

22. **Singh M, *et al.*** (2005) Prevalence and risk factors for transmission of infection among children in household contact with adults having pulmonary tuberculosis. *Archives of Disease in Childhood* **90**(6), 624–628.
23. **Huang CC, *et al.*** (2014) Cigarette smoking among tuberculosis patients increases risk of transmission to child contacts. *International Journal of Tuberculosis & Lung Disease* **18**(11), 1285–1291.
24. **Nair D, *et al.*** (2016) Household contact screening and yield of tuberculosis cases – a clinic based study in Chennai, South India. *PLoS ONE* **11**(9), e0162090.
25. **Cayla JA, *et al.*** (1996) The influence of intravenous drug use and HIV infection in the transmission of tuberculosis. *AIDS* **10**(1), 95–100.
26. **Gessner BD, Weiss NS and Nolan CM** (1998) Risk factors for pediatric tuberculosis infection and disease after household exposure to adult index cases in Alaska. *Journal of Pediatrics* **132**(3), 509–513.
27. **Elmi OS, *et al.*** (2014) Prevalence and associated factors with transmission of latent tuberculosis among household contacts of multi-drug resistant tuberculosis patients in Malaysia. *World Journal of Medical Sciences* **10**(3), 285–294.
28. **Lienhardt C, *et al.*** (2003) Risk factors for tuberculosis infection in children in contact with infectious tuberculosis cases in the Gambia, West Africa. *Pediatrics* **111**(5), e608–e614.
29. **Mohammad Z, *et al.*** (2002) A preliminary study of the influence of HIV infection in the transmission of tuberculosis. *Southeast Asian Journal of Tropical Medicine & Public Health* **33**(1), 92–98.
30. **Xu L, *et al.*** (2008) Dose-response relationship between treatment delay of smear-positive tuberculosis patients and intra-household transmission: a cross-sectional study. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **102**(8), 797–804.
31. **Mendes MA, *et al.*** (2013) Ontact screening in tuberculosis: can we identify those with higher risk? *European Respiratory Journal.* **41**(3), 758–760.
32. **Loredo C, *et al.*** (2014) Yield of close contact tracing using two different programmatic approaches from tuberculosis index cases: a retrospective quasi-experimental study. *BMC Pulmonary Medicine* **14**(1), 133.
33. **Espinal MA, *et al.*** (2000) Infectiousness of *Mycobacterium tuberculosis* in HIV-1-infected patients with tuberculosis: a prospective study. *Lancet (British Edition)* **355**(9200), 275–280.
34. **Rathi SK, *et al.*** (2002) Prevalence and risk factors associated with tuberculin skin test positivity among household contacts of smear-positive pulmonary tuberculosis cases in Umerkot, Pakistan. *International Journal of Tuberculosis and Lung Disease* **6**(10), 851–857.
35. **Guwatudde D, *et al.*** (2003) Tuberculosis in household contacts of infectious cases in Kampala, Uganda. *American Journal of Epidemiology* **158**(9), 887–898.
36. **Lee MS, *et al.*** (2008) Early and late tuberculosis risks among close contacts in Hong Kong. *International Journal of Tuberculosis & Lung Disease* **12**(3), 281–287.
37. **Teixeira L, *et al.*** (2001) Infection and disease among household contacts of patients with multidrug-resistant tuberculosis. *International Journal of Tuberculosis & Lung Disease* **5**(4), 321–328.
38. **Davies P, *et al.*** (2006) Smoking and tuberculosis: the epidemiological association and immunopathogenesis. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **100**(4), 291–298.
39. **Lönnroth K, *et al.*** (2008) Alcohol use as a risk factor for tuberculosis–a systematic review. *BMC Public Health* **8**(1), 289.
40. **Jubulis J, *et al.*** (2014) Modifiable risk factors associated with tuberculosis disease in children in Pune, India. *The International Journal of Tuberculosis and Lung Disease* **18**(2), 198–204.
41. **Du Preez K, *et al.*** (2011) Environmental tobacco smoke exposure increases *Mycobacterium tuberculosis* infection risk in children. *The International Journal of Tuberculosis and Lung Disease* **15**(11), 1490–1497.
42. **Initiative GL** (2014) *Mycobacteriology Laboratory Manual*. Geneva: WHO Stop TB Partnership.
43. **Ernst JD** (2012) The immunological life cycle of tuberculosis. *Nature Reviews Immunology* **12**(8), 581–591.
44. **Cruciani M, *et al.*** (2001) The impact of human immunodeficiency virus type 1 on infectiousness of tuberculosis: a meta-analysis. *Clinical Infectious Diseases* **33**(11), 1922–1930.
45. **Haramati L, Jenny-Avital E and Alterman D** (1997) Effect of HIV status on chest radiographic and CT findings in patients with tuberculosis. *Clinical Radiology* **52**(1), 31–35.
46. **Elliott A, *et al.*** (1993) Negative sputum smear results in HIV-positive patients with pulmonary tuberculosis in Lusaka, Zambia. *Tubercle and Lung Disease* **74**(3), 191–194.
47. **Klein NC, *et al.*** (1989) Use of mycobacterial smears in the diagnosis of pulmonary tuberculosis in AIDS/ARC patients. *Chest* **95**(6), 1190–1192.
48. **Martinez L, *et al.*** (2016) Infectiousness of HIV-seropositive patients with tuberculosis in a high-burden African setting. *American Journal of Respiratory and Critical Care Medicine* **194**(9), 1152–1163.
49. **Carvalho ACC, *et al.*** (2001) Transmission of *Mycobacterium tuberculosis* to Contacts of HIV-infected Tuberculosis Patients. *American Journal of Respiratory and Critical Care Medicine* **164**(12), 2166–2171.
50. **Baliashvili D, *et al.*** (2016) Contacts of retreatment tuberculosis cases with a prior poor treatment outcome are at increased risk of latent tuberculosis infection. *International Journal of Infectious Diseases* **43**, 49–50.

Supplementary Materials

Contents

Table 1: Complete search strategy

MEDLINE with Ovid software

(“tuberculosis”[Title] OR "tuberculosis"[MeSH Terms] OR "mycobacterium tuberculosis"[MeSH Terms] OR "tuberculosis, pulmonary"[MeSH Terms] OR “TB” [Title])

AND

(("contact$"[All Fields]) OR ("contact tracing"[MeSH Terms]) OR "disease outbreaks"[MeSH Terms] OR "contact*"[Title] OR "spread*"[Title] OR "contact follow*"[All Fields] OR "contact tracing"[Title] OR "disease transmission"[All Fields] OR "case find*"[Title] OR (cluster*[Title] OR “Genotypic cluster*”[All fields] OR “Phylogenetic cluster*”[All Feilds]

AND

(Transmiss*[Title]) OR "index case*"[All Fields] OR ("secondary case"[All Fields]) OR (“case finding”[All Fields]) OR (“infectiousness”[All Fields]) OR (“Molecular epidemiology”[Title]))

EMBASE search with Ovid software

(“tuberculosis”[Title] OR "tuberculosis"[MeSH Terms] OR "mycobacterium tuberculosis"[MeSH Terms] OR "lung tuberculosis"[MeSH Terms] OR “latent tuberculosis”OR “TB” [Title])

AND

(("contact$"[All Fields]) OR ("contact examination"[MeSH Terms]) OR "epidemic"[MeSH Terms] OR "contact tracing"[Title] OR "disease transmission"[All Fields] OR "case find*"[Title] OR (cluster*[Title] OR “Genotypic cluster*”[All fields] OR “Phylogenetic cluster*”[All Feilds]

AND

(Transmission*[Title]) OR "index case*"[All Fields] OR ("secondary case"[All Fields]) OR (“case finding”[All Fields]) OR (“infectiousness”[All Fields]) OR (“Molecular epidemiology”[Title]))

Global Health and Global Archieve with CAB-Direct software

Tuberculosis AND Transmission AND Contact tracing [all feilds]

Mycobacterium tuberculosis AND Transmission AND Genotypic Clustering

Web of Science

Tuberculosis [title] OR Mycobacterium tuberculosis (title) OR TB [title]

AND

Contact tracing [title]

Contact*

Genotypic Clustering [title]

AND

Transmission [title]

Table 2- Quality assessment of 37 included studies with Newcastle-Ottawa Scale (Point scored)

	Studies & reference	**Selection (4 stars)**				**Comparability (2 star)**		**Outcome (3 Star)**			**Total (9 stars)**
		Representativeness of the sample	Sample size	Ascertainment of exposure	Non-respondents	controls for the most important factor (Age)	control for any additional factor	Assessment of outcome	Follow-up long enough for outcomes to occur	Adequacy of follow up of participants	
1.	Baliashvili, 2016	*		*				*			3*
2.	Elliott, 1993			*		*	*	*			4*
3.	Maciel, 2009			*				*	*	*	4*
4.	Madhi, 2002			*		*	*	*			4*
5.	Singh, 2005			*		*	*	*			4*
6.	Teixeira, 2001			*		*	*	*			4*
7.	Elmi, 2014	*		*	*		*	*			5*
8.	Faksri, 2015			*	*	*	*	*			5*
9.	Jones-Lopez, 2016			*	*	*	*	*			5*
10.	Mendes, 2012	*		*		*	*	*			5*
11.	Nair, 2016	*		*		*	*	*			5*
12.	Nunn, 1994	*		*		*	*	*			5*
13.	Carvalho, 2001	*		*		*	*	*	*		6*
14.	Castillo Otero, 1999	*	*	*	*			*		*	6*
15.	Cayla, 1996	*		*	*	*	*	*			6*
16.	Kifai, 2009	*	*	*		*	*	*			6*
17.	Mohammad, 2002	*		*	*	*	*	*			6*
18.	Tornee, 2004		*	*	*	*	*	*			6*
19.	Espinal, 2000	*		*	*	*	*	*		*	7*
20.	Gessner, 1998	*	*	*	*	*	*	*			7*
21.	Godoy, 2013	*	*	*	*	*	*	*			7*
22.	Golub, 2006	*		*	*	*	*	*		*	7*
23.	Grandjean, 2011	*	*	*	*	*	*	*			7*

24.	Lienhardt, 2003	*		*		*	*	*	*	*	7*
25.	Ling, 2011	*	*	*	*	*	*	*			7*
26.	Rathi, 2002	*	*	*	*	*	*	*			7*
27.	Xu, 2008	*	*	*	*	*	*	*			7*
28.	Lohmann, 2012	*	*	*	*		*	*	*	*	8*
29.	Loredo, 2014	*	*	*	*		*	*	*	*	8*
30.	Grandjean, 2015	*	*	*	*	*	*	*	*	*	9*
31.	Guwatudde, 2003	*	*	*	*	*	*	*	*	*	9*
32.	Huang, 2014 (Cigar)	*	*	*	*	*	*	*	*	*	9*
33.	Huang, 2014	*	*	*	*	*	*	*	*	*	9*
34.	Lee, 2008	*	*	*	*	*	*	*	*	*	9*
35.	Martinez, 2016	*		*	*	*	*	*	*	*	9*
36.	Otero, 2016	*	*	*	*	*	*	*	*	*	9*

Table 3- Definitions of index cases and contacts in the 37 included studies.

Definition of index patients	First Author & year	Reference
o Newly diagnosed sputum smear-positive PTB	Otero, 2016 Nair, 2016 Jones-Lopez, 2016 Kifai, 2009 Tornee, 2004 Rathi, 2002 Cayla, 1996 Xu, 2008	[1] [2] [3] [4] [5] [6] [7] [8]
o Confirmed PTB patients	Mendes, 2012 Golub, 2006 Lienhardt, 2003 Mohammad, 2002 Madhi, 2002 Carvalho, 2001 Espinal, 2000 Gessner, 1998 Nunn, 1994 Elliott, 1993 Singh, 2005	[9] [10] [11] [12] [13] [14] [15] [16] [17] [18] [19]
o Re-treatment AFB positive patients	Baliashvili, 2016	[20]
o MDR-TB patients and drug-susceptible TB patients	Grandjean, 2015 Teixeira, 2001 Snider, 1985	[21] [22] [23]
o Newly diagnosed TB patients	Faksri, 2015 Godoy, 2013 Ling, 2011	[24] [25] [26]
o All TB patients	Loredo, 2014 Lohmann, 2012 Lee, 2008 Guwatudde, 2003 Castillo Otero, 1999	[27] [28] [29] [30] [31]
o New culture-positive TB patients	Huang, 2014 Martinez, 2016 Maciel, 2009	[32] [33] [34]
o Adults (15 years and above) with drug-susceptible TB	Huang, 2014	[35]
o Patients with MDR-TB and XDR-TB	Elmi, 2014	[36]
o Adults (14 years and above) with MDR-TB	Grandjean, 2011	[37]
Definition of contacts		
• Household contacts	Otero, 2016 Nair, 2016 Jones-Lopez, 2016 Grandjean, 2015 Faksri, 2015 Huang, 2014 Huang, 2014 Elmi, 2014 Grandjean, 2011 Kifai, 2009 Martinez, 2016	[1] [2] [3] [21] [24] [32] [35] [36] [37] [4] [33]

	Guwatudde, 2003 Rathi, 2002 Mohammad, 2002 Teixeira, 2001 Carvalho, 2001 Espinal, 2000 Nunn, 1994 Elliott, 1993 Xu, 2008	[30] [6] [12] [22] [14] [15] [17] [18] [8]
• Daily contacts (both household and non-household)	Baliashvili, 2016 Cayla, 1996	[20] [7]
• Close contacts who shared an enclosed space	Loredo, 2014 Lee, 2008 Golub, 2006	[27] [29] [10]
• Contacts at different level of concentric circle	Godoy, 2013 Lohmann, 2012	[25] [28]
• Close (household) and casual contacts	Mendes, 2012 Castillo Otero, 1999	[9] [31]
• Children and adolescents (aged 15 or younger) household contacts	Maciel, 2009 Tornee, 2004 Madhi, 2002 Gessner, 1998 Snider, 1985 Ling, 2011	[34] [5] [13] [16] [23] [26]
• Children under the age of 5 years in household contact	Singh, 2005 Lienhardt, 2003	[19] [11]

Table 4 - Methods used to detect TB transmission from index patients to their contacts by 37 included studies

Outcome	Method of screening (cut-off)	First Author	Reference
LTBI	TST (≥10mm)	Baliashvili, 2016	[20]
LTBI	TST (≥10mm)	Huang, 2014	[32]
LTBI	TST (⩾10mm, ⩾5mm HIV+)	Huang, 2014	[35]
LTBI	TST (≥5mm)	Godoy, 2013	[25]
LTBI	TST (≥10mm)	Lohmann, 2012	[28]
LTBI	TST (⩾10mm, ⩾5mm HIV+)	Kifai, 2009	[4]
LTBI	TST (⩾10mm)	Martinez, 2016	[33]
LTBI	TST (≥5mm)	Golub, 2006	[10]
LTBI	TST (⩾10mm)	Singh, 2005	[19]
LTBI	TST (⩾15mm)	Tornee, 2004	[5]
LTBI	TST (⩾5mm)	Lienhardt, 2003	[11]
LTBI	TST (⩾10mm)	Rathi, 2002	[6]
LTBI	TST (⩾10mm)	Mohammad, 2002	[12]
LTBI	TST (⩾10mm)	Teixeira, 2001	[22]
LTBI	TST (≥10mm, ≥5mm HIV+)	Carvalho, 2001	[14]
LTBI	TST (≥5mm)	Espinal, 2000	[15]
LTBI	TST (≥5mm)	Nunn, 1994	[17]
LTBI	TST (≥5mm)	Elliott, 1993	[18]
LTBI	TST (⩾10mm)	Snider, 1985	[23]
LTBI	TST (⩾15mm)	Xu, 2008	[8]
LTBI	TST (≥10mm) & IGRA	Jones-Lopez, 2016	[3]
LTBI	TST (≥10mm) & IGRA	Faksri, 2015	[24]
LTBI	TST & IGRA	Mendes, 2012	[9]
LTBI	IGRA	Elmi, 2014	[36]
TB disease	symptoms and CXR	Otero, 2016	[1]
TB disease	CXR	Nair, 2016	[2]
TB disease	Bacteriological, CXR or clinical	Grandjean, 2015	[21]
TB disease	Bacteriological, CXR or clinical	Grandjean, 2011	[37]
TB disease	Bacteriological, CXR or clinical	Lee, 2008	[29]
TB disease	Bacteriological, CXR or clinical	Guwatudde, 2003	[30]
TB disease	Bacteriological, CXR or clinical	Ling, 2011	[26]
LTBI & TB	TST& bacteriological, CXR or clinical	Loredo, 2014	[27]
LTBI & TB	TST (⩾5mm)& bacteriological, CXR or clinical	Gessner, 1998	[16]
LTBI & TB	TST (≥10mm), CXR or clinical	Maciel, 2009	[34]
LTBI & TB	TST (⩾10mm)& CXR	Madhi, 2002	[13]
LTBI & TB	TST (≥5mm) & CXR	Castillo Otero, 1999	[31]
LTBI & TB	TST & CXR	Cayla, 1996	[7]

Table 5: Characteristics of Studies included in Meta-analyses

	Author, year	Country	Study period	Design	Index no.	Contact no.	Method of contact screening
1.	Carvalho, 2001 [14]	Brazil	1995-1997	Population based	86	360	TST
2.	Castillo Otero, 1999 [31]	Spain	1992-1996	Population based	302	1029	TST
3.	Cayla, 1996 [7]	Spain	1990-1993	Population based	248	1080	TST & CXR
4.	Elliott, 1993 [18]	Zambia	1989	Population based	71	304	TST
5.	Elmi, 2014 [36]	Malaysia	2010-2012	Population based	69	70	IGRA
6.	Espinal, 2000 [15]	Russia	14months	Population based	755	803	TST
7.	Faksri, 2015 [24]	Thailand	2012-2013	Population based	40	70	TST& IGRA
8.	Gessner, 1998 [16]	USA,	1987-1994	Population based		282	TST
9.	Godoy, 2013 [25]	Spain	2005-2006	Population based	1329	6548	TST
10.	Golub, 2006 [10]	USA	2000-2001	Population based	124	703	TST
11.	Grandjean, 2015 [21]	Peru	2010-2013	Population based	700	3417	Bacteriological,x-ray or clinica
12.	Guwatudde, 2003 [30]	Uganda	1995-1999	Population based	302	1206	Bacteriological, CXR or clinical
13.	Jones-Lopez, 2016 [3]	Uganda	2009–2011	Population based	85	369	TST& IGRA
14.	Lee, 2008 [29]	Hong Kong	2000	Population based	1537	4661	Bacteriological, CXR or clinical
15.	Ling, 2011 [26]	Taiwan	2005-2007	Contacts age <20	2952	5879	Sputum, CXR, Clinical TB
16.	Loredo, 2014 [27]	Brazil	2005-2008	Population based	369	1310	TST
17.	Maciel, 2009 [34]	Brazil	2003-2006	Population based		146	TST
18.	Madhi, 2002 [13]	Paris	1997-2000	Children contacts 3months-17yrs		408	TST
19.	Martinez, 2016 [33]	Uganda	1995-2006	Population based	503	1933	TST
20.	Mendes, 2012 [9]	Portugal	2011	Population based	69	397	TST& IGRA
21.	Mohammad, 2002 [12]	Malaysia	1999-2000	Population based	56	138	TST
22.	Nair, 2016 [2]	India,	2007-2011	Population based	280	544	CXR & Symptoms
23.	Nunn, 1994 [17]	Kenya	1989-1990	Population based	82	357	TST
24.	Singh, 2005 [19]	India	18months	Population based	200	281	TST
25.	Teixeira, 2001 [22]	Brazil	1994-1998	Population based	78	408	TST
26.	Tornee, 2004 [5]	Thailand	2002-2003	contacts aged less than 15 years	342	500	TST

Table 6: Sensitivity analysis result for studies included in meta-analyses of sputum smear positivity Crude odds ratios

Excluded study	Pooled OR	LCI 95%	HCI 95%	Cochran Q	p	I^2	I^2 LCI 95%	I^2 HCI 95%
Castillo Otero, 1999	2.44	1.96	3.04	19.33	0.04	48.26	0	74.18
Elmi, 2014	2.24	1.81	2.77	22.31	0.01	55.19	11.76	77.24
Gessner, 1998	2.28	1.82	2.87	23.06	0.01	56.64	14.10	77.88
Godoy, 2013	2.45	1.92	3.14	21.05	0.02	52.48	5.73	76.05
Golub, 2006	2.20	1.79	2.71	20.76	0.02	51.82	4.26	75.76
Lee, 2008	2.19	1.78	2.71	20.67	0.02	51.62	3.82	75.67
Ling, 2011	2.23	1.79	2.78	21.77	0.02	54.07	9.27	76.75
Loredo, 2014	2.44	1.91	3.11	22.89	0.01	56.31	14.27	77.74
Maciel, 2009	2.27	1.82	2.83	23.33	0.01	57.13	16.08	78.10
Mohammad, 2002	2.34	1.87	2.93	24.03	0.01	58.39	18.87	78.66
Nunn, 1994	2.24	1.80	2.78	22.20	0.01	54.96	11.26	77.14
Singh, 2005	2.24	1.78	2.74	20.98	0.02	52.35	5.43	75.99

Table 7: Sensitivity analysis result for studies included in meta-analyses of lung Cavitation Crude odds ratios

Excluded study	**Pooled OR**	**LCI 95%**	**HCI 95%**	**Cochran Q**	**p**	**I^2**	**I^2 LCI 95%**	**I^2 HCI 95%**
Elliott, 1993	1.89	1.29	2.78	84.85	4.86E-13	85.86	77.47	91.12
Faksri, 2015	2.06	1.44	2.93	82.86	1.17E-12	85.52	76.85	90.94
Gessner, 1998	1.91	1.29	2.83	85.62	3.45E-13	85.98	77.70	91.19
Godoy, 2013	1.96	1.28	3.00	66.98	1.17E-09	82.08	70.54	89.10
Golub, 2006	1.85	1.29	2.66	84.01	7.06E-13	85.71	77.21	91.04
Guwatudde, 2003	1.93	1.32	2.81	86.66	2.18E-13	86.15	77.10	91.28
Jones-Lopez, 2016	2.06	1.44	2.93	83.30	9.63E-13	85.59	76.10	90.98
Lee, 2008	1.92	1.32	2.80	86.61	2.23E-13	86.14	77.10	91.28
Madhi, 2002	1.82	1.28	2.60	81.65	2.00E-12	85.30	76.47	90.82
Mendes, 2012	2.03	1.39	2.96	84.52	5.63E-13	85.80	77.37	91.09
Mohammad, 2002	1.97	1.36	2.86	86.82	2.03E-13	86.18	78.04	91.30
Nair, 2016	2.04	1.41	2.95	84.60069	5.43E-13	85.81	77.39	91.10
Nunn, 1994	1.92	1.30	2.82	86.37788	2.47E-13	86.11	77.92	91.26
Tornee, 2004	1.70	1.35	2.16	27.27098	0.01	56.00	18.09	76.36

Table 8: Sensitivity analysis result for studies included in meta-analyses of HIV sero-status Crude odds ratios

Excluded study	Pooled OR	LCI 95%	HCI 95%	Cochran Q	p	I^2	I^2 LCI 95%	I^2 HCI 95%
Carvalho, 2001	0.82	0.57	1.16	34.97	1.13E-05	79.98	61.16	89.68
Cayla, 1996	0.68	0.49	0.93	29.89	9.94E-05	76.58	53.33	88.25
Godoy, 2013	0.71	0.48	1.05	30.31	8.34E-05	76.90	54.08	88.38
Grandjean, 2015	0.73	0.51	1.06	40.56	9.83E-07	82.74	67.35	90.88
Loredo, 2014	0.79	0.56	1.12	39.46	1.60E-06	82.26	66.28	90.67
Martinez, 2016	0.76	0.49	1.17	36.01	7.22E-06	80.56	62.47	89.93
Mohammad, 2002	0.81	0.57	1.15	37.60	3.61E-06	81.38	64.33	90.28
Nunn, 1994	0.72	0.50	1.04	39.64	1.47E-06	82.34	66.46	90.70
Teixeira, 2001	0.84	0.61	1.17	32.05	3.98E-05	78.16	56.99	88.91

Figure 1-Flowchart of the identification, screening, and inclusion/exclusion of studies

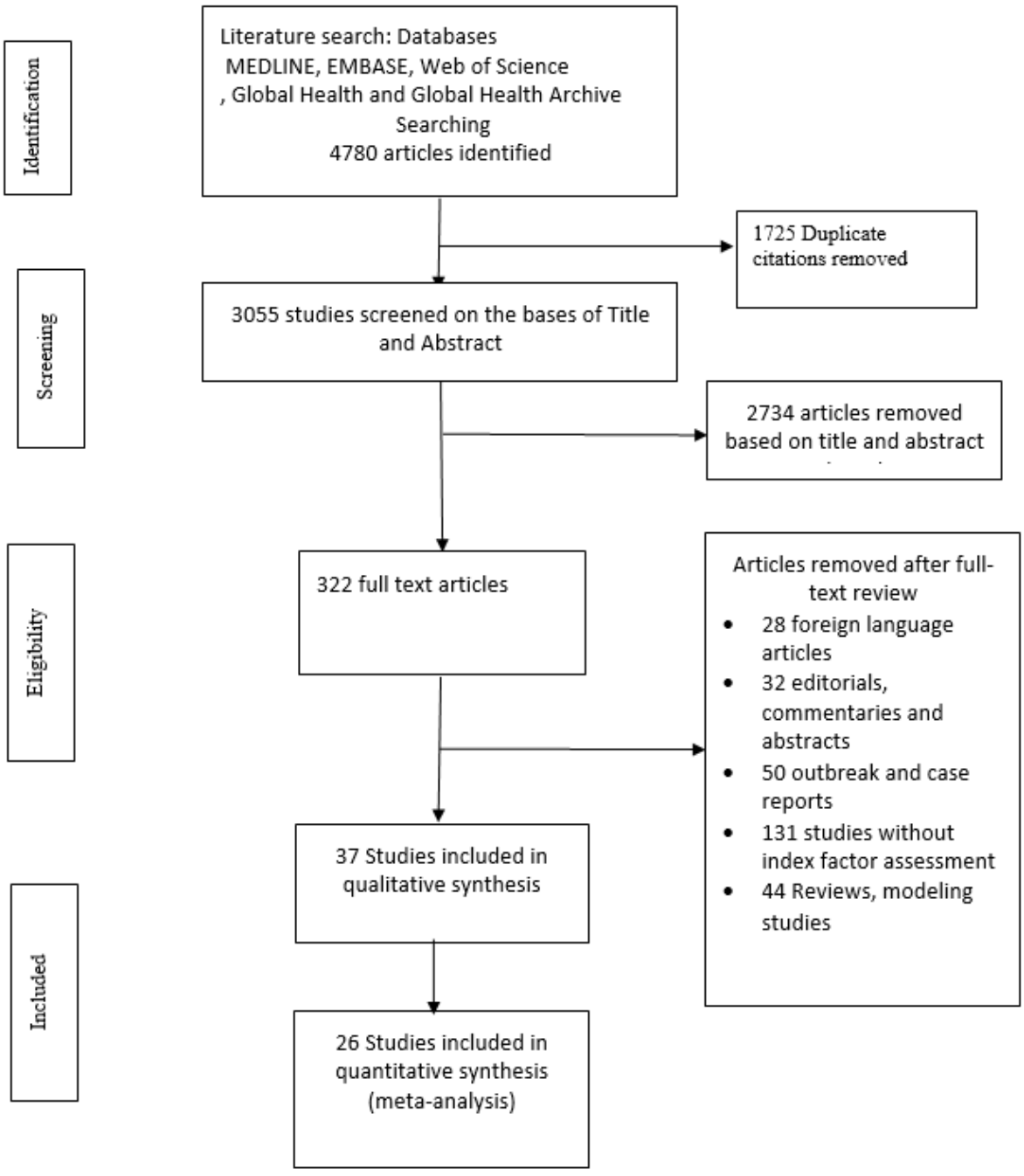


Figure 2- Forest plot of crude odds ratios for the association between index patient sputum smear-positivity and infectiousness. * indicates that these studies take contact active disease as an outcome

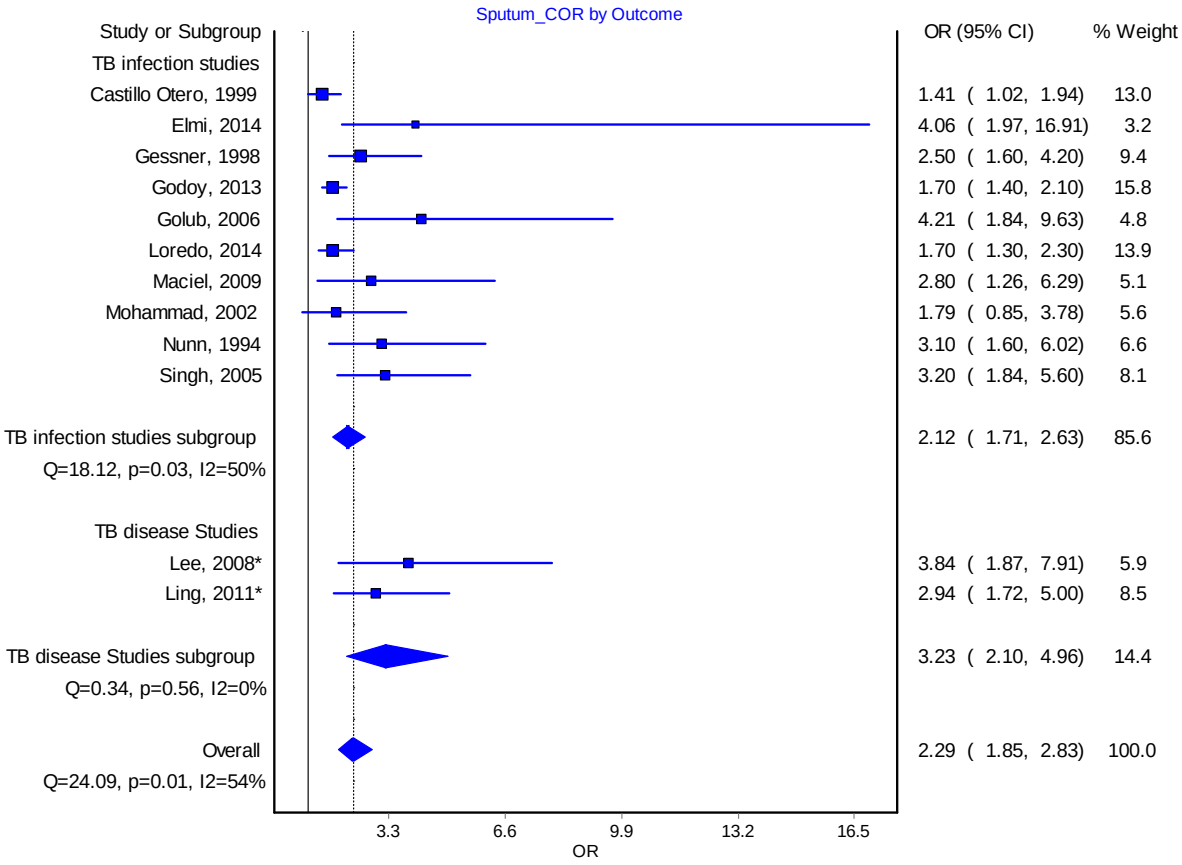


Figure 3-Forest plot of the association between index patient lung cavitation and infectiousness from crude odds ratio analysis. * indicates that these studies take contact active disease as an outcome

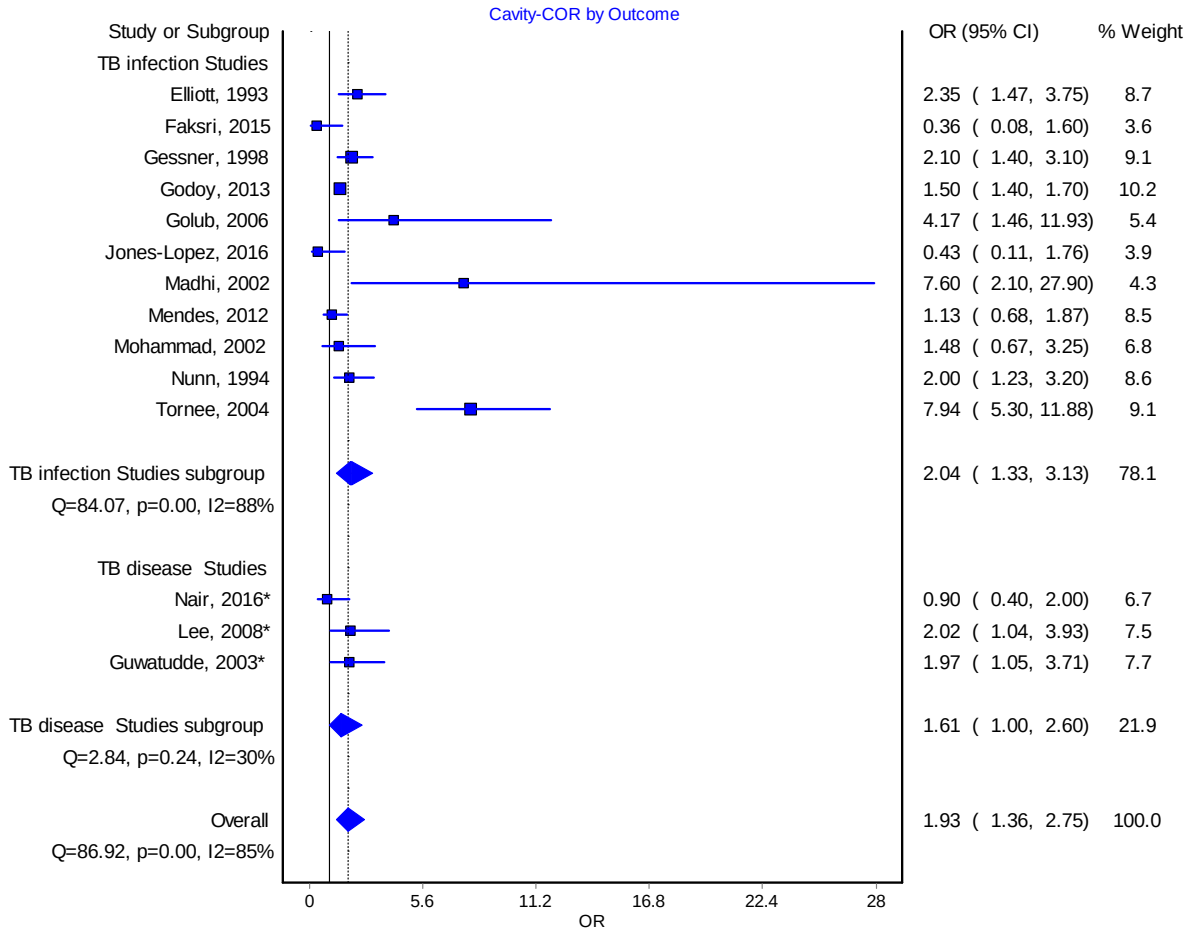


Figure 4-Forest plot of the association between index patient HIV sero-positivity and infectiousness from crude odds ratio analysis. * indicates that these studies take contact active disease as an outcome

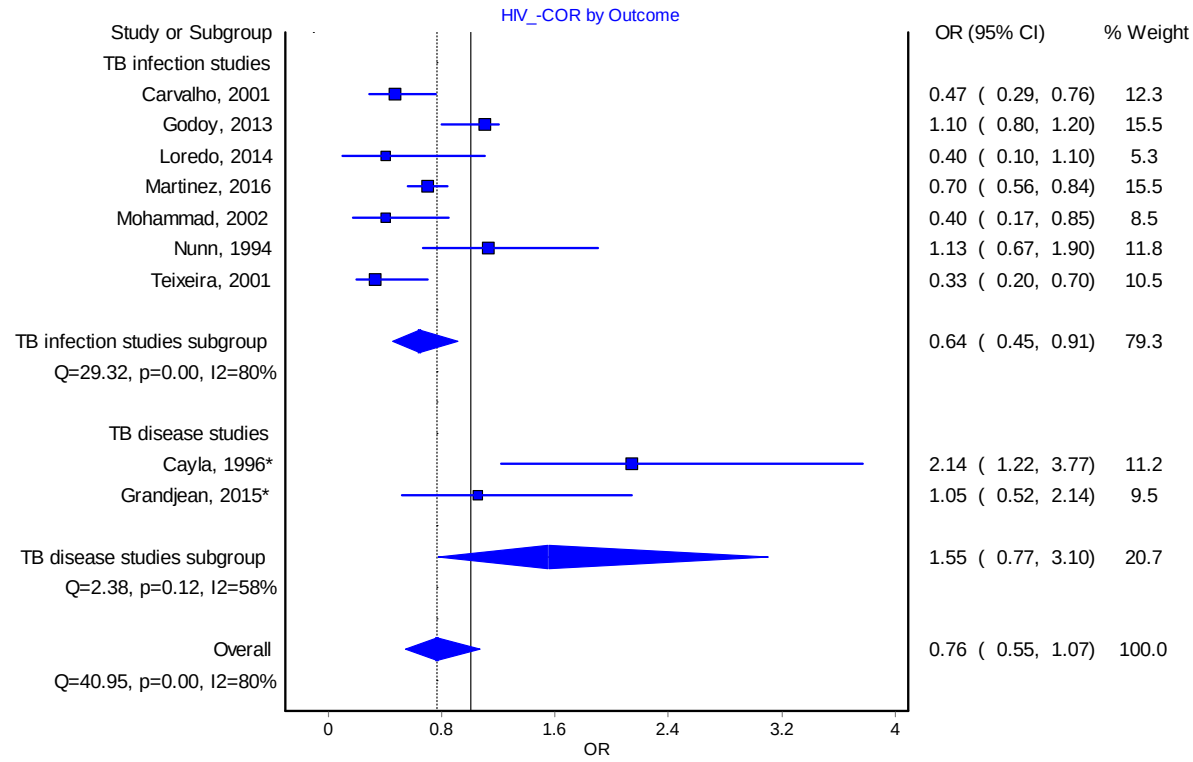


Figure 5-Funnel plot of results for the crude odds ratio of the association between sputum smear-positivity and infectiousness

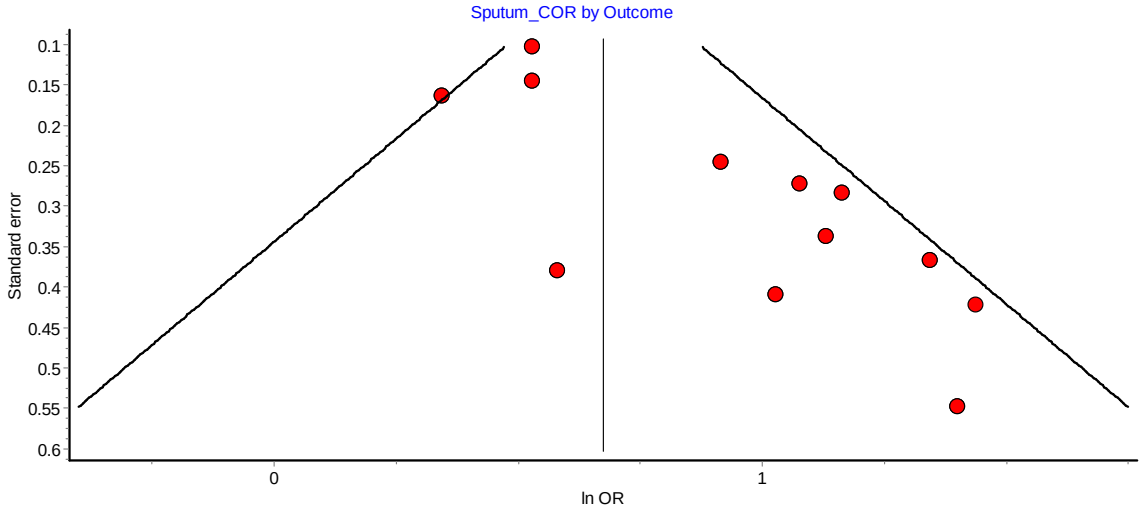


Figure 6-Funnel plot of results for the crude odds ratio of the association between cavitation and infectiousness

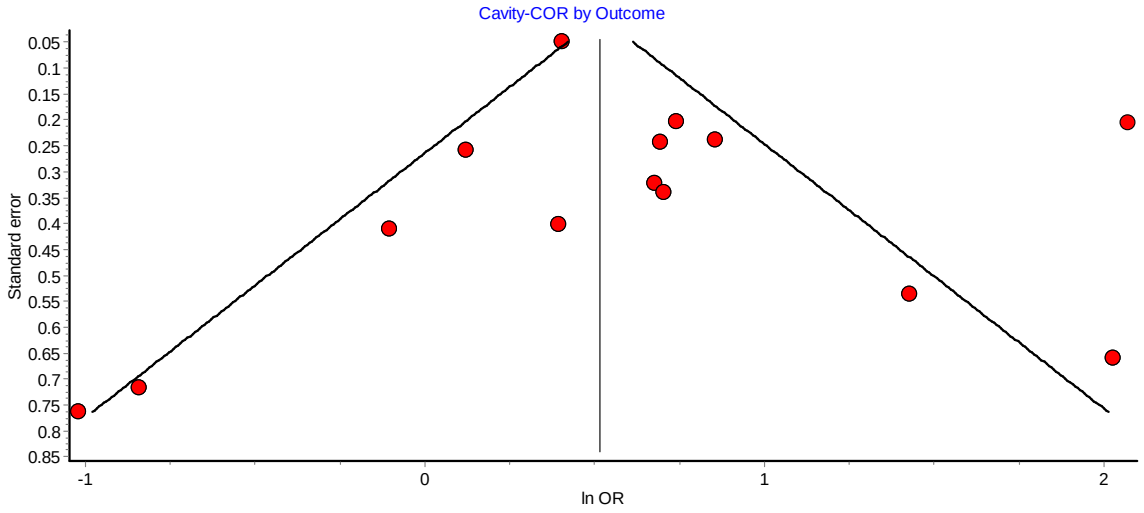
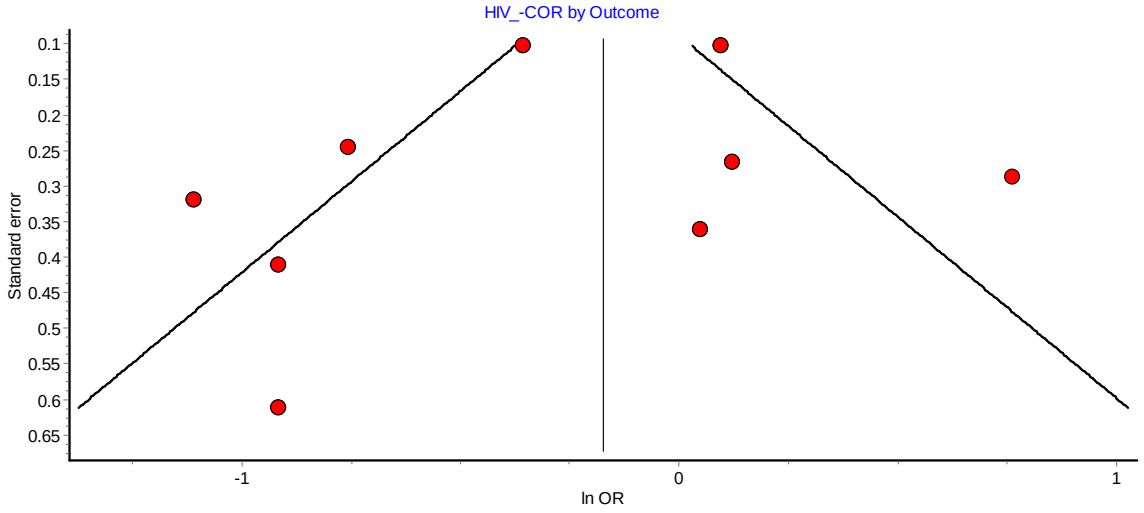


Figure 7-Funnel plot of results for the crude odds ratio of the association between HIV sero-positivity and infectiousness



References

(1) **Otero L, et al.** A prospective longitudinal study of tuberculosis among household contacts of smear-positive tuberculosis cases in Lima, Peru. *BMC Infectious Diseases* 2016; **16 (**259).

(2) **Nair D, et al.** Household contact screening and yield of tuberculosis cases-a clinic based study in Chennai, South India. *PLoS ONE* 2016; **11 (9)**.

(3) **Jones-Lopez EC, et al.** Cough Aerosols of mycobacterium tuberculosis in the prediction of incident tuberculosis disease in household contacts. *Clinical Infectious Diseases* 2016; **63**(1): 10-20.

(4) **Kifai EJ, Bakari M.** Mantoux skin test reactivity among household contacts of HIV-infected and HIV un-infected patients with sputum smear positive TB in Dar es Salaam, Tanzania. *East African Journal of Public Health* 2009; **6**(2): 211-218.

(5) **Tornee S, et al.** Risk factors for tuberculosis infection among household contacts in Bangkok, Thailand. *Southeast Asian Journal of Tropical Medicine & Public Health* 2004; **35**(2): 375-383.

(6) **Rathi SK, et al.** Prevalence and risk factors associated with tuberculin skin test positivity among household contacts of smear-positive pulmonary tuberculosis cases in Umerkot, Pakistan. *International Journal of Tuberculosis and Lung Disease* 2002; **6**(10): 851-857.

(7) **Cayla JA, et al.** The influence of intravenous drug use and HIV infection in the transmission of tuberculosis. *AIDS* 1996; **10**(1): 95-100.

(8) **Xu L, et al.** Dose-response relationship between treatment delay of smear-positive tuberculosis patients and intra-household transmission: a cross-sectional study. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 2008; **102**(8): 797-804.

(9) **Mendes MA, et al.** Contact screening in tuberculosis. Can we identify those with higher risk? *European Respiratory Journal Conference: European Respiratory Society Annual Congress* 2012; **40**.

(10) **Golub JE, et al.** Delayed tuberculosis diagnosis and tuberculosis transmission. *International Journal of Tuberculosis & Lung Disease* 2006; **10**(1): 24-30.
(11) **Lienhardt C, et al.** Risk factors for tuberculosis infection in children in contact with infectious tuberculosis cases in the Gambia, West Africa. *Pediatrics* 2003; **111**(5): e608-614.
(12) **Mohammad Z, et al.** A preliminary study of the influence of HIV infection in the transmission of tuberculosis. *Southeast Asian Journal of Tropical Medicine & Public Health* 2002; **33**(1): 92-98.
(13) **Madhi F, et al.** Transmission of tuberculosis from adults to children in a Paris suburb. *Pediatric Pulmonology* 2002; **34**(3): 159-163.
(14) **Carvalho ACC, et al.** Transmission of Mycobacterium tuberculosis to Contacts of HIV-infected Tuberculosis Patients. *American Journal of Respiratory and Critical Care Medicine* 2001; **164**(12): 2166-2171.
(15) **Espinal MA, et al.** Infectiousness of Mycobacterium tuberculosis in HIV-1-infected patients with tuberculosis: a prospective study. *Lancet (British edition)* 2000; **355**(9200): 275-280.
(16) **Gessner BD, Weiss NS, Nolan CM.** Risk factors for pediatric tuberculosis infection and disease after household exposure to adult index cases in Alaska. *Journal of Pediatrics* 1998; **132**(3): 509-513.
(17) **Nunn P, et al.** The effect of human immunodeficiency virus type-1 on the infectiousness of tuberculosis. *Tubercle & Lung Disease* 1994; **75**(1): 25-32.
(18) **Elliott AM, et al.** The impact of HIV on infectiousness of pulmonary tuberculosis: a community study in Zambia. *AIDS* 1993; **7**(7): 981-987.
(19) **Singh M, et al.** Prevalence and risk factors for transmission of infection among children in household contact with adults having pulmonary tuberculosis. *Archives of Disease in Childhood* 2005; **90**(6): 624-628.
(20) **Baliashvili D, et al.** Contacts of retreatment tuberculosis cases with a prior poor treatment outcome are at increased risk of latent tuberculosis infection. *International Journal of Infectious Diseases* 2016; **43**: 49-50.
(21) **Grandjean L, et al.** Transmission of Multidrug-Resistant and Drug-Susceptible Tuberculosis within Households: A Prospective Cohort Study. *PLoS Medicine* 2015; **12 (6)** (e1001843).
(22) **Teixeira L, et al.** Infection and disease among household contacts of patients with multidrug-resistant tuberculosis. *International Journal of Tuberculosis & Lung Disease* 2001; **5**(4): 321-328.
(23) **Snider DE, Jr., et al.** Infection and disease among contacts of tuberculosis cases with drug-resistant and drug-susceptible bacilli. *American Review of Respiratory Disease* 1985; **132**(1): 125-132.
(24) **Faksri K, et al.** Transmission and Risk Factors for Latent Tuberculosis Infections among Index Case-Matched Household Contacts. *Southeast Asian Journal of Tropical Medicine & Public Health* 2015; **46**(3): 486-495.
(25) **Godoy P, et al.** Immigrants do not transmit tuberculosis more than indigenous patients in Catalonia (Spain). *Tuberculosis* 2013; **93**(4): 456-460.
(26) **Ling DL, et al.** Contact investigation for tuberculosis in Taiwan contacts aged under 20 years in 2005. *International Journal of Tuberculosis & Lung Disease* 2011; **15**(1): 50-55.
(27) **Loredo C, et al.** Yield of close contact tracing using two different programmatic approaches from tuberculosis index cases: A retrospective quasi-experimental study. *BMC Pulmonary Medicine* 2014; **14 (1),** 133.
(28) **Lohmann EM, et al.** Grading of a positive sputum smear and the risk of Mycobacterium tuberculosis transmission. *International Journal of Tuberculosis and Lung Disease* 2012; **16**(11): 1477-1484.
(29) **Lee MS, et al.** Early and late tuberculosis risks among close contacts in Hong Kong. *International Journal of Tuberculosis & Lung Disease* 2008; **12**(3): 281-287.
(30) **Guwatudde D, et al.** Tuberculosis in household contacts of infectious cases in Kampala, Uganda. *American Journal of Epidemiology* 2003; **158**(9): 887-898.
(31) **Castillo Otero Dd, et al.** Investigation of tuberculosis contacts in a nonhospital pneumology practice. *European Journal of Clinical Microbiology & Infectious Diseases* 1999; **18**(11): 790-795.

(32) **Huang CC, et al.** The Effect of HIV-Related Immunosuppression on the Risk of Tuberculosis Transmission to Household Contacts. *Clinical Infectious Diseases* 2014; **58**(6): 765-774.
(33) **Martinez L, et al.** Infectiousness of HIV-seropositive patients with tuberculosis in a high-burden African setting. *American journal of respiratory and critical care medicine* 2016 Nov 1; **194**(9): 1152-1163.
(34) **Maciel ELN, et al.** Juvenile household contacts aged 15 or younger of patients with pulmonary TB in the greater metropolitan area of Vitoria, Brazil: A cohort study. . *Jornal Brasileiro de Pneumologia* 2009; **35**(4): 359-366.
(35) **Huang CC, et al.** Cigarette smoking among tuberculosis patients increases risk of transmission to child contacts. *International Journal of Tuberculosis & Lung Disease* 2014; **18**(11): 1285-1291.
(36) **Elmi OS, et al.** Prevalence and associated factors with transmission of latent tuberculosis among household contacts of multi-drug resistant tuberculosis patients in Malaysia. *World Journal of Medical Sciences* 2014; **10**(3): 285-294.
(37) **Grandjean L, et al.** Tuberculosis in household contacts of multidrug-resistant tuberculosis patients. *International Journal of Tuberculosis and Lung Disease* 2011; **15**(9): 1164-1169.

Chapter 3

3. The role of super-spreading events in *Mycobacterium tuberculosis* transmission: Evidence from contact tracing

3.1. Chapter overview

Having provided evidence concerning the variation among patients with TB in their capacity to transmit infection and identified associated factors in Chapter 2, Chapter 3 addresses thesis question 2 by quantifying heterogeneous infectiousness and the impact of super-spreading in *Mtb* transmission. This Chapter analyses the number of secondary infections per index TB patient from an extensive population-based TB contact investigation data from the Victorian TB Program, Australia. By constructing empirical offspring distributions, the Chapter shows that heterogeneous infectiousness was an essential feature of *Mtb* transmission: very few (10% of TB patients) individuals are highly infectious, with super-spreaders account for more than three-quarters of secondary infections. Further, this Chapter identifies factors associated with the number of secondary infections that each TB patient produces.

In order to optimise TB control interventions, a recommendation is made on the need to consider the heterogeneity of transmission and super-spreading. This Chapter also provides data that will be used to inform novel models able to capture TB transmission dynamics and capture heterogeneous infectiousness, which will be described in more detail in the modelling phase of the research.

Citation: **Melsew YA**, Gambhir M, Cheng AC, McBryde ES, Denholm JT, Tay EL, Trauer JM. The role of super-spreading events in *Mycobacterium tuberculosis* transmission: evidence from contact tracing. *BMC infectious diseases*. 2019 Dec;19(1):244.

Melsew et al. BMC Infectious Diseases (2019) 19:244
https://doi.org/10.1186/s12879-019-3870-1

BMC Infectious Diseases

RESEARCH ARTICLE Open Access

The role of super-spreading events in *Mycobacterium tuberculosis* transmission: evidence from contact tracing

Yayehirad A. Melsew[1,2*], Manoj Gambhir[1], Allen C. Cheng[1], Emma S. McBryde[3,4], Justin T. Denholm[5,6], Ee Laine Tay[7] and James M. Trauer[1,5]

Abstract

Background: In current epidemiology of tuberculosis (TB), heterogeneity in infectiousness among TB patients is a challenge, which is not well studied. We aimed to quantify this heterogeneity and the presence of "super-spreading" events that can assist in designing optimal public health interventions.

Methods: TB epidemiologic investigation data notified between 1 January 2005 and 31 December 2015 from Victoria, Australia were used to quantify TB patients' heterogeneity in infectiousness and super-spreading events. We fitted a negative binomial offspring distribution (NBD) for the number of secondary infections and secondary active TB disease each TB patient produced. The dispersion parameter, *k*, of the NBD measures the level of heterogeneity, where low values of *k* (e.g. $k < 1$) indicate over-dispersion. Super-spreading was defined as patients causing as many or more secondary infections as the 99th centile of an equivalent homogeneous distribution. Contact infection was determined based on a tuberculin skin test (TST) result of ≥10 mm. A NBD model was fitted to identify index characteristics that were associated with the number of contacts infected and risk ratios (RRs) were used to quantify the strength of this association.

Results: There were 4190 (2312 pulmonary and 1878 extrapulmonary) index TB patients and 18,030 contacts. A total of 15,522 contacts were tested with TST, of whom 3213 had a result of ≥10 mm. The dispersion parameter, *k* for secondary infections was estimated at 0.16 (95%CI 0.14–0.17) and there were 414 (9.9%) super-spreading events. From the 3213 secondary infections, 2415 (75.2%) were due to super-spreading events. There were 226 contacts who developed active TB disease and a higher level of heterogeneity was found for this outcome than for secondary infection, with *k* estimated at 0.036 (95%CI 0.025–0.046). In regression analyses, we found that infectiousness was greater among index patients found by clinical presentation and those with bacteriological confirmation.

Conclusion: TB transmission is highly over dispersed and super-spreading events are responsible for a substantial majority of secondary infections. Heterogeneity of transmission and super-spreading are critical issues to consider in the design of interventions and models of TB transmission dynamics.

Keywords: Tuberculosis, Super-spreading, Negative binomial distribution, Victoria

* Correspondence: 078yayu@gmail.com
[1]Department of Epidemiology and Preventive Medicine, School of Public Health and Preventive Medicine, Monash University, Melbourne, Victoria, Australia
[2]Department of Epidemiology and Biostatistics, Institute of Public Health, University of Gondar, Gondar, Ethiopia
Full list of author information is available at the end of the article

Background

The Global Tuberculosis (TB) Strategy looks towards the ultimate vision of elimination of the TB epidemic, although the disease still causes more than 10 million cases and 1.8 million deaths each year [1, 2]. The epidemic is not homogeneously distributed, but is a collection of heterogeneous local micro-epidemics [3]. The existence of heterogeneity in transmission has the potential to disrupt elimination strategies, many of which assume broadly similar transmission potential of infectious people. Hence, it is essential to understand and quantify the degree of heterogeneity that exists in TB transmission.

Several studies have reported heterogeneity in the capacity of individual source patients to transmit various pathogens to their contacts [4–7]. Variation between infectious individuals in their capacity to transmit infectious agents is well described, with some super-spreaders infecting large number of contacts while others may only infect very few or none [4, 5, 7–10]. The contribution of super-spreading has previously been quantified for directly transmitted infections such as SARS, measles, smallpox, monkey-pox and pneumonic plague [4].

TB patients are diverse in their capacity to transmit infection to their contacts, with a systematic review that included several contact tracing studies showing that clinical, demographic and behavioural characteristics of TB patients were associated with their ability to transmit *Mycobacterium tuberculosis* (*M. tb*) infection [11]. Quantification of heterogeneity in *M. tb* transmission will help to understand better how its transmission is sustained. A study in the Netherlands using genotypic clustering data quantified *M. tb* transmission heterogeneity and reported signs of super-spreading [10]. However, how TB patients vary with respect to their capacity to produce secondary infection is not well understood, including the extent to which super-spreading events exist and are responsible for driving transmission. Understanding transmission heterogeneity and characterising those with greater capacity to spread the infection is critical to better target interventions and predict their likely impact. As the global TB response transitions towards ending TB and the epidemic becomes more localised [3], understanding heterogeneities of transmission is increasingly important. We aimed to characterise transmission heterogeneity in a well-resourced setting using high-quality surveillance data including detailed information on contacts and their infection status.

Methods

Setting

Victoria is a state of Australia with approximately 5.6 million people and a single centralised tuberculosis program (the Victorian Tuberculosis Program; VTP). Notification of all confirmed or suspected cases of TB disease is mandatory for both laboratories and clinicians and culture confirmation of *M. tb* is routine in this setting. While hospitalisation of cases is not mandatory, those with pulmonary disease are typically maintained in isolation until they are considered non-infectious (> 2 weeks of effective therapy and/or smear-negativity) [12, 13]. On receipt of a notification, a public health nurse from the VTP is allocated to the patient to provide support, assist with treatment compliance, and assess the extent of contact tracing required. Household contacts and others with greater than an estimated 8 h of contact are considered eligible for screening, with individualised assessment of screening recommendations for higher-risk contacts performed (e.g. immunosuppressed or those with high-intensity exposure).

Contact investigation initially consists of clinical assessment and serial testing for *M. tb* infection. Testing of contacts is conducted by either tuberculin skin testing (TST) using the Mantoux procedure or an interferon gamma release assay (IGRA), although during the study period the large majority of testing was undertaken with TST. Those negative on initial testing are tested again 8–12 weeks following exposure. Contacts with either symptoms suggestive of active disease or a positive test for TB infection undergo chest x-ray (CXR) and further clinical assessment, with isoniazid preventive treatment offered for those where active disease is excluded [12].

Data

Data from the VTP which are stored by the Victorian Department of Health and Human Services (DHHS) were used for the following analysis. Index patients were classified as confirmed cases of TB notified from 1 January 2005 to 31 December 2015 in residents of Victoria. The data set includes contact tracing information and results of testing for *M. tb* infection, with cases of subsequent active TB disease linked to these contact episodes now extending to March 2017 (see [14] for earlier publication of linkage process). We constructed empirical offspring distributions from the detailed contact tracing data set of the VTP.

Ethical approval was obtained from Monash University, Human Research Ethics Committee (Project Number: 7776) and permission was given by the VTP and DHHS.

Definitions

A microbiologically confirmed case of TB requires culture or polymerase chain reaction (PCR) diagnosis of *Mycobacterium tuberculosis*, while clinical/radiological diagnosis may also be made by a medical practitioner experienced in TB management. Approximately 90% of TB cases in Victoria are bacteriologically confirmed [15]. All cases diagnosed with active TB in this dataset also underwent secondary case review by a TB

specialist to ensure that guidelines for confirming TB disease were met.

Contacts were individuals identified as having had personal contact with index patients by the VTP, through school, workplace, household and other settings. Contact latent TB infection (LTBI) was defined as a TST result of ≥10 mm in an identified contact. Where contacts had had multiple TSTs, we used the value of the latest TST result performed within three months of exposure. Any contact developing active TB during the stated period until 21 March 2017 was considered to have secondary TB. We define a super-spreading event as the number of secondary infections per index case that was greater than the 99th centile of the equivalent Poisson distribution (with distribution mean equal to the mean number of infections per index).

Fitting distributions to data

We were primarily interested in the distribution of the number of secondarily infected contacts from each index patient. Although super-spreading is usually defined in terms of the number of secondary cases of active disease produced by each index patient, we wished to estimate parameters in the absence of preventive therapy. As isoniazid prophylaxis is used widely in our setting and its use may be clustered according to index patients (e.g. family members electing together to undertake preventive treatment), we anticipated this could artefactually inflate our estimates of over-dispersion. Although there are contact factors that are likely to affect progression to active disease after infection, these factors may not be differentially distributed by index patient and the distribution of infections would be unaffected by use of preventive therapy. Therefore, the number of secondarily infected cases is our primary analysis throughout the remainder of the paper, although analogous analyses are presented for the distribution of contacts (close contacts and all contacts) and secondary cases of active disease to facilitate epidemiological interpretation.

Our primary outcome can be described using a probability distribution termed an offspring distribution, defined as the probability of the number of transmission events across the range of index TB patients. This process can be modelled by a negative binomial distribution (NBD), which has the advantage of being able to accommodate over-dispersed count data [16–18]. The NBD permits sufficient flexibility with only two parameters (the shape parameter and the mean) [4] and subsumes the Poisson distribution while also allowing for "over-dispersion", where the variance (of the offspring distribution) may be greater than the mean [18].

We denote the individual reproductive number by v and, the distribution of individual reproductive numbers (offspring distribution) by Z. To incorporate individual infectious histories, v follows a negative binomial offspring distribution with dispersion parameter k and mean m, such that $Z \sim NegB(m, k)$. The dispersion parameter k quantifies the extent of over-dispersion in the count data. If there is extra-heterogeneity between index patients in the number of secondary infections produced, dispersion increases and the parameter k approaches zero ($k \to 0$). In the absence of over-dispersion, $k \to \infty$ and the mean and the variance approach parity, with the negative binomial distribution reducing to the Poisson distribution. If $k = 1$, the negative binomial distribution reduces to the geometric distribution, such that the negative binomial model can accommodate Poisson, geometric and over-dispersed distributions [16].

The probability of observing index patients with v number of infected contacts is given by:

$$P(Z = v) = \frac{(k+v-1)!}{v!(k-1)!}.\left(\frac{m}{m+k}\right)^{v}\left(1+\frac{m}{k}\right)^{-k}, m > 0, k > 0 \quad (1)$$

As the variance $m(1 + (m/k))$ approaches the mean (m), over-dispersion decreases, i.e. $k \to \infty$.

Interpretation of transmission parameters

The parameters, k and m were estimated by maximum likelihood estimation (MLE), which provides unbiased estimates, especially for large sample sizes [16]. The MLE of the mean of the offspring distribution, m is the sample mean of Z or the mean number of secondary infections. The dispersion parameter, k, was estimated after fitting the data to the negative binomial distribution using the MASS package [19] of the R environment for statistical computing [20], with a value of k less than one interpreted as evidence of super-spreading [10].

Super-spreading events

We used the protocol proposed by Lloyd-Smith et al [4] in which: first, we calculated the mean number of secondary infections per index or effective reproductive number, R_n; second, we constructed a Poisson distribution with mean R_n, representing the range of Z (offspring distribution) due to stochasticity without individual variation; third, we define a super-spreading event as any patient who infected more than $Z(i)$ contacts, where $Z(i)$ is i^{th} centile of the offspring Poisson distribution. Arbitrarily but as in this previous study for SARS, we considered the 99th percentile of this distribution as the cut-point to determine super-spreading events. For prediction of the expected proportion of super-spreading event, we produce a negative binomial distribution with dispersion parameter, k, and the mean number of secondary infections per index, R_n [4, 21].

Identifying associations with index characteristics

The outcome variable was the number of secondary infections per index patient, which was found to be over-dispersed as described below. Therefore, we fitted a negative binomial regression model with both bivariate and multivariate regression with the MASS package [19] of the R environment for statistical computing [20]. The logarithmic scale coefficients were exponentiated to give ratios and are presented with their 95% confidence intervals (CI).

We evaluated the need for a negative binomial regression model (because of inequality of the conditional mean and conditional variance) with the likelihood ratio test. We also evaluated the predictive accuracy of the model with rootograms [22].

Results

The Victorian TB program data had a total of 4190 confirmed TB index patients and 18,030 contacts within the period from 1 January 2005 to 31 December 2015. The mean age of index patients was 33.0 years and 54.9% were male. Among index patients, 1878 (44.8%) were extrapulmonary, while 1757 (42.0%) were pulmonary and 555 (13.2%) had both pulmonary plus other site involvement.

The average age of contacts was 28.4 years, 9276 (51.4%) were female and 8510 (47.2%) were male (with 244 (1.4%) contacts sex not stated). The majority of contacts (15,031; 83.4%) were contacts of pulmonary only TB patients, while (2988; 16.6%) were contacts of patients of TB at pulmonary plus other sites and only 11 contacts were identified from EPTB patients (although all EPTB were considered to have produced zero secondary infections and secondary cases). Henceforward we use the term "pulmonary" to refer to any patient with pulmonary involvement, i.e. both the "pulmonary only" and the "pulmonary plus other sites" categories. There were five categories for the types of contacts, 8059 close contacts, 5484 school contacts, 2366 work contacts, 1286 casual contacts and 824 contacts from other congregate setting such as hospitals, nursing homes, airlines and childcare facilities.

Secondary infection distribution

A total of 15,522 of 18,019 contacts of PTB index patients were tested with TST. Based on our cut-off for diagnosing infection as those with a TST result of ≥10 mm as positive, 3213 of contacts were infected and 12,309 were not. Of 3213 infected contacts 2050 (63.8%) were close contacts. Of the 4190 index patients (1878 extrapulmonary and 2312 pulmonary) 3166 (75.6%) did not produce any secondary infection, with all extrapulmonary patients assumed not to have produced any secondary infections (Additional file 1). There were 26 cluster sizes of secondary infection, ranging from zero infections to 41 infections per index. The mean and variance of the number of secondary infections was 0.77 and 5.06 infections per index respectively. The median number of secondary infections per index patient was zero, the 95th centile was two and the 99th centile was 10 infections per index. By fitting the NBD to the observed distribution of secondary infection, we found evidence of over-dispersion (k = 0.16, 95%CI 0.14–0.17) for all types of contacts.

Restricting our analysis to index patients with pulmonary involvement only, the mean and variance of the number of infections per index was 1.4 and 8.3 infections per index, respectively with evidence of over-dispersion (k= 0.36, 95%CI 0.33–0.40) (Fig. 1a). In another restricted analysis of close contacts only, there were 2042 secondary infections, of which 771 (37.8%) were super-spreading events, with less dispersion compared to all contact types (k= 0.98, 95%CI 0.84–1.12).

Super-spreading events

We constructed a Poisson distribution with the mean number of infections per index (i.e. 0.77) to establish a cut-off number of secondary infections per index for defining super-spreading events. The 99th centile was three infections per index. Therefore, we classified transmission events where index patients produced three or more secondary infections as super-spreading events. Accordingly, there were 414 (9.9% index patients) associated with super-spreading events, which accounted for a total of 2415 (75.2%) of the 3213 secondary infections. For predicting the number of super-spreading events in TB transmission, we estimated the expected proportion of index patients, with confidence intervals considering the dispersion parameter, $k = 0.16$ and the effective reproductive number,$R_n = 0.77$. With this approach, the expected proportion of TB super-spreading events was 9.8% (95% CI: 8.9–10.6%).

Secondary active TB disease distribution

We further analysed infectiousness heterogeneity from the number of secondary active TB cases per index TB patient. There were 226 secondary active TB cases identified among 18,030 contacts. Among these secondary TB cases, approximately half (116; 51.3%) were the sole secondary case identified among the contacts of a specific index patient. Among 4190 index patients, only 137 (3.3%) were responsible for all 226 secondary TB cases. Two of the secondary cases were contacts of extrapulmonary index patients (maternal to foetal transmission in utero). The largest cluster of secondary TB disease was 12 cases. The distribution of secondary TB cases per index, was over dispersed with the dispersion parameter estimated at k = 0.036 (95%CI 0.025–0.046) (Fig. 1b).

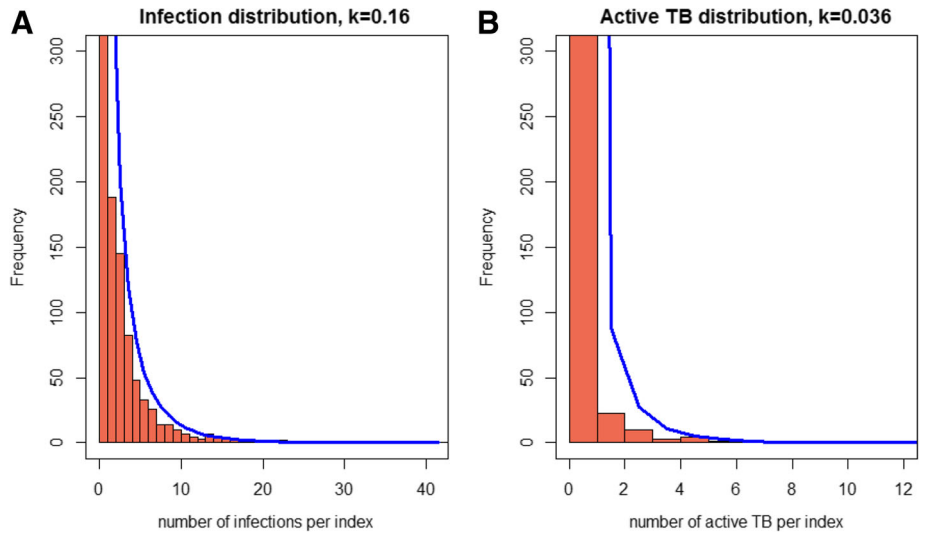


Fig. 1 a. Distribution of number of secondary infections per index TB patient with negative binomial distribution fitted to count data in Victoria, for the period 2005–2015. The number of index patients with zero infections was 3166 (beyond limit of vertical axis). **b.** Distribution of secondary TB disease by index patients in Victoria, for the period 2005–2015, with negative binomial distribution fitted to count data. The number of index patients with zero secondary active TB was 4054 (beyond limit of vertical axis). Both panels include data for both PTB and EPTB index patients and all types of contacts

Contact distribution

We also investigated the individual variation among index patients with respect to the number of contacts identified to determine how the transmission heterogeneity could be related to contact patterns. There was evidence of heterogeneity (k= 0.38, 95%CI 0.36–0.41) for the distribution all contact types (Additional file 2). Similar but lesser heterogeneity was found (k = 0.63, 95%CI 0.59–0.68) after restricting our analysis to close contacts only (Additional file 3).

Associations of index characteristics with number of infections

Because of the extent of heterogeneity described above, we fitted a negative binomial regression model for index patients with pulmonary involvement to determine the characteristics of index patients that were associated with the number of secondary infections. The index characteristics included were age, sex, site of disease, patient detection pathway, CXR result, method of diagnosis, whether the patient was new or relapse and number of contacts per index.

From this multivariate model, TB in pulmonary and additional sites (compared to pulmonary only) was independently associated with a 42% decrease in the number of secondary infections. Identification through contact tracing and the Australian post-migration follow up program ("health undertakings" [23]) (compared to clinical presentation) was associated with a lower number of secondary infections (71 and 46% respectively). Diagnosis by PCR, histology or clinical signs (compared to culture) was associated with a 70% decrease in secondary infections, while diagnosis by radiological techniques was associated with a 55% decrease. The number of contacts identified for each index patient showed a positive association with the number of secondary infections produced, with the identification of one additional contact associated with an increase in the number of secondary infections by 4 % (Table 1).

The likelihood ratio test showed that assuming equality of the conditional mean and variance was not safe (P-value < 0.001), whereas model evaluation indicated that the negative binomial model fitted well to the observed data. Compared to the equivalent Poisson model, the negative binomial model was well-fitted and predicted counts well with minimal residuals (Additional file 4).

Discussion

To our knowledge, this study is the first to use programmatic epidemiological observations to formally quantify *M. tb* transmission heterogeneity. We found evidence of super-spreading events as constituting the large majority of *M. tb* transmission events. We also demonstrated considerable variability between index TB patients in three important respects: in the number of contacts identified, the number of contacts infected and the number of cases of secondary active TB subsequently occurring. Therefore, assuming a homogeneous population of infective patients in TB transmission modelling may be highly unrealistic. From a programmatic viewpoint, although the effective reproduction number is less than one in our setting,

Table 1 Associations between number of secondary infections and patient characteristics among pulmonary TB index patients in Victoria, 2005–2015. Negative binomial regression model

Variable		Number of index patients	Crude RR (95%CI)	Adjusted RR (95%CI)
Age category (years)	0–14	140	0.43 (0.30–0.62)*	0.84 (0.56–1.26)
	15–24	486	1.60 (1.32–1.95)*	0.94 (0.80–1.11)
	25–44	890	Referent	
	45–64	378	0.77 (0.61–0.96)*	1.02 (0.85–1.22)
	65 and above	417	0.93 (0.75–1.15)	0.91 (0.77–1.09)
Sex	Female	954	Referent	
	Male	1357	0.95 (0.81–1.10)	0.93 (0.82–1.05)
Site of disease	Pulmonary	1757	Referent	
	Pulmonary plus additional sites	555	0.51 (0.43–0.62)*	0.58 (0.50–0.68)*
Patient detection pathway	Clinical Presentation	1839	Referent	
	Contact tracing	131	0.132 (0.08–0.21)*	0.29 (0.18–0.46)*
	Screening	24	0.49 (0.23–1.11)	0.64 (0.33–1.22)
	Health undertaking	317	0.37 (0.29–0.46)*	0.54 (0.44–0.67)*
CXR	Normal	31	Referent	
	Abnormal	879	1.79 (0.84–3.60)	1.18 (0.66–2.12)
	Unknown	1401	1.95 (0.91–3.91)	1.27 (0.71–2.26)
Diagnosis	Culture	1967	Referent	
	PCR/NAT/Ht/Cs	130	0.15 (0.09–0.23)*	0.30 (0.19–0.45)*
	Radiological	214	0.21 (0.15–0.29)*	0.45 (0.33–0.62)*
New or relapse	New case	2177	Referent	
	Relapse following full treatment	83	1.23 (0.84–1.86)	1.07 (0.78–1.47)
	Relapse following partial treatment	30	0.53 (0.26–1.14)	0.70 (0.38–1.25)
	Unknown	21	1.38(0.67–3.24)	1.02(0.56–1.91)
Number of contacts identified		18019[a]	1.06(1.05–1.06)*	1.04(1.04–1.05)*

RR rate ratio, *CI* confidence interval, *CXR* Chest X-Ray, *PCR/NAT/Ht/Cs* Polymerase chain reaction/nucleic acid test/histology/clinical signs, Health undertaking is post migration follow-up program

[a]=total number of contacts, * = statistically significant at *P* value < 0.05

significant transmission may still occur due to a very small number of super-spreading events (Fig. 2).

We found that *M. tb* transmission is heterogeneous, with the distribution of secondary infections per index varying by more than simple random variation between individuals. The level of over-dispersion was comparable with the estimate from a previous study that employed genotypic data (k=0.1) [10]. Although the distribution of secondarily infected contacts per index patient is probably a better marker for *M. tb* transmission than genotypic clusters, our approach may even underestimate heterogeneity (overestimate k) due to dilution of differences between patients from more homogeneously distributed distant past infection (i.e. prior to the index exposure identified). Our estimate of k was slightly higher (i.e. less heterogeneous) than an estimate for SARS transmission heterogeneity (k=0.1) [4], though our finding may under-estimate the true level of heterogeneity. The distribution of secondary cases was even more heterogeneous (k< 0.04) than secondary infections. This

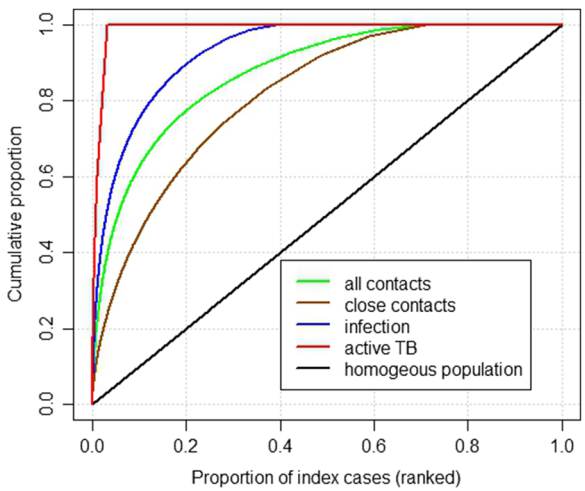


Fig. 2 Proportion of all contacts, close contacts, proportion of secondary infections and secondary active TB disease due to a given proportion of infectious patients in Victoria, for the period 2005–2015

finding was also more heterogeneous compared to the previous TB estimate from genotypic data (k = 0.1) [10] and other infectious diseases such as SARS [4]. However, the effective use of preventive therapy in this setting and long incubation period could lead to an underestimate of the risk of active TB, as there is a risk of missing late reactivations of TB, although late reactivation are relatively rare in Victoria [24]. The effective use of preventive therapy could also overestimate the true level of heterogeneity in cases of secondary active TB, by differentially reducing the number of secondary cases produced by some index cases. The analyses of contacts demonstrated that these heterogeneities are not solely driven by heterogeneity in contact patterns. Therefore, we believe that the true value of the *M. tb* transmission dispersion parameter is likely to fall somewhere between the *k* estimate for secondary infections and secondary active TB distributions.

Based on our definition [4], three-quarters of all secondary *M. tb* infections occurred as a result of super-spreading events. Although the average number of secondary infections per index (R_n) was less than one (0.77), TB rates may not decline as expected in the population due to this high heterogeneity. Moreover, TB transmission goes well beyond the 20/80 rule-of-thumb for infectious disease transmission which states that 80% of transmission is due to only 20% of the population [25], since in our results 20% of index patients produce 90% of secondary infections (Fig. 3).

Our regression analysis identified several important associations between index patients' characteristics and the number of secondary infections they produced, rather than considering just the proportion of contacts infected as typically done in previous studies [26, 27].

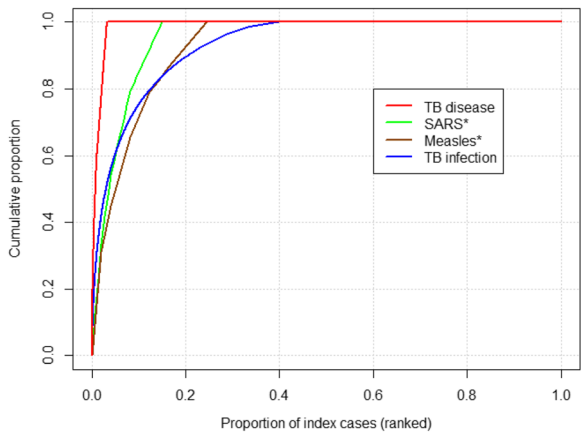


Fig. 3 Comparison of proportion of secondary TB infections and secondary active TB disease with SARS and Measles due to a given proportion of index cases. *estimates taken from: Lloyd-Smith JO, Schreiber SJ, Kopp PE, Getz WM. Superspreading and the effect of individual variation on disease emergence. Nature. 2005;438(7066):355–9

Compared to index patients with pulmonary TB only, index patients with TB involving pulmonary and non-pulmonary sites were less infectious. This may be explained by those with extrapulmonary TB but minor CXR abnormalities often being classified as "pulmonary plus other sites", with these patients tending to have a low bacillary load and smear-negative pulmonary disease (given that it is well-established that pulmonary patients are more infectious than extrapulmonary who have virtually zero infectiousness [28, 29]). The method of patient identification was also an important predictor of infectiousness, with index patients found by clinical presentation being more infectious than those found through contact tracing or post-migration follow-up. This could be explained by the fact that those patients identified through the passive process of relying upon clinical presentation spend longer infectious, whereas those identified through the more active approaches of contact tracing and post-migration follow-up allows earlier identification and treatment. This explanation is consistent with delayed diagnosis and treatment being a major predictor of TB patients' infectiousness [30–33]. Similarly, index patients diagnosed by culture produced a higher number of secondary infections, which is consistent with patients with culture-positive results having a higher bacillary load [34].

The most important limitation of our study is that some secondary infections might be the result of distant past infections, rather than relating to the contact episode. Our definition of super-spreading events is based on contact infection rather than active disease in contacts, since there is no standard definition in the case of TB and the difficulty in interpreting active disease in a setting of widespread use of preventive therapy. However, we argue that contact infection is the best available measure of true *M. tb* transmission in our setting and this definition could be used for future studies on the disease.

Conclusions

We conclude that *M. tb* transmission is a highly heterogeneous process in our population and super-spreading events are a major driver of transmission. Therefore, it is essential to consider this heterogeneity when modelling TB transmission dynamics and considering control strategies. Future observational studies should characterise super-spreading in different epidemiological settings to further characterise this phenomenon.

Additional files

Additional file 1: Figure S1. Schematic presentation of number of index TB patient and contacts in Victoria, for the period 2005–2015. (DOCX 52 kb)

Additional file 2: Figure S2. Distribution of number of contacts per index TB patient in Victoria, for the period 2005–2015. A. All contacts (with negative binomial distribution fitted to count data) strategies, the number of index patients with zero contacts was 639 (beyond limit of vertical axis). B. Subset of contacts (0–40 contacts per index only). (DOCX 34 kb)

Additional file 3: Figure S3. Distribution of number of Close contacts per index TB patient in Victoria, for the period 2005–2015. A. All Close contacts (with negative binomial distribution fitted to count data). B. Subset of close contacts (0–40 close contacts per index), the number of index patients with zero close contacts was 653 (beyond limit of vertical axis). (DOCX 34 kb)

Additional file 4: Figure S4. Hanging rootograms for a Poisson model (upper panel) and negative binomial model (lower panel) count data. (DOCX 37 kb)

Abbreviations
BCG: Bacillus Calmette–Guérin; CI: Confidence Interval; CXR: Chest X-Ray; DHHS: Department of Health and Human Services; IGRA: Interferon Gamma Release Assay; LTBI: Latent Tuberculosis Infection; *M. tb*: Mycobacterium tuberculosis; PCR/NAT: Polymerase Chain Reaction/Nucleic Acid Test; RR: Rate Ratio; TB: Tuberculosis; TST: Tuberculin Skin Test; VTP: Victorian Tuberculosis Program; WHO: World Health Organization

Acknowledgments
The authors would like to thank Monash University for providing the PhD scholarship to YAM. We are also thankful the Victorian Department of Health and Human Services for providing data.

Funding
Y. A. Melsew is a recipient of a Monash Graduate Scholarship for his PhD study. J. M. Trauer is a recipient of an Early Career Fellowship from the National Health and Medical Research Council. No specific funding was received for this study.

Availability of data and materials
The data that support the findings of this study are available from the Victorian Department of Health and Human Services but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Data are however available from the authors upon reasonable request and with permission of the Victorian Department of Health and Human Services.

Authors' contributions
YAM, MG and JMT conceptualised the study. YAM developed the protocol and, JMT and MG reviewed it. ET and YAM performed data extraction and cleaning. YAM, ACC, ESM, JTD and JMT undertook the analysis. YAM drafted the manuscript and all authors involved in consecutive revisions of the manuscript. All authors reviewed and approved the final manuscript.

Ethics approval and consent to participate
Ethical approval was obtained from Monash University, Human Research Ethics Committee (Project Number: 7776) and permission was given by the Victorian Department of Health and Human Services and the Victorian Tuberculosis Program. Since we used secondary data that were stored by the Victorian Department of Health and Human Services, we did not seek individual consent.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Publisher's Note

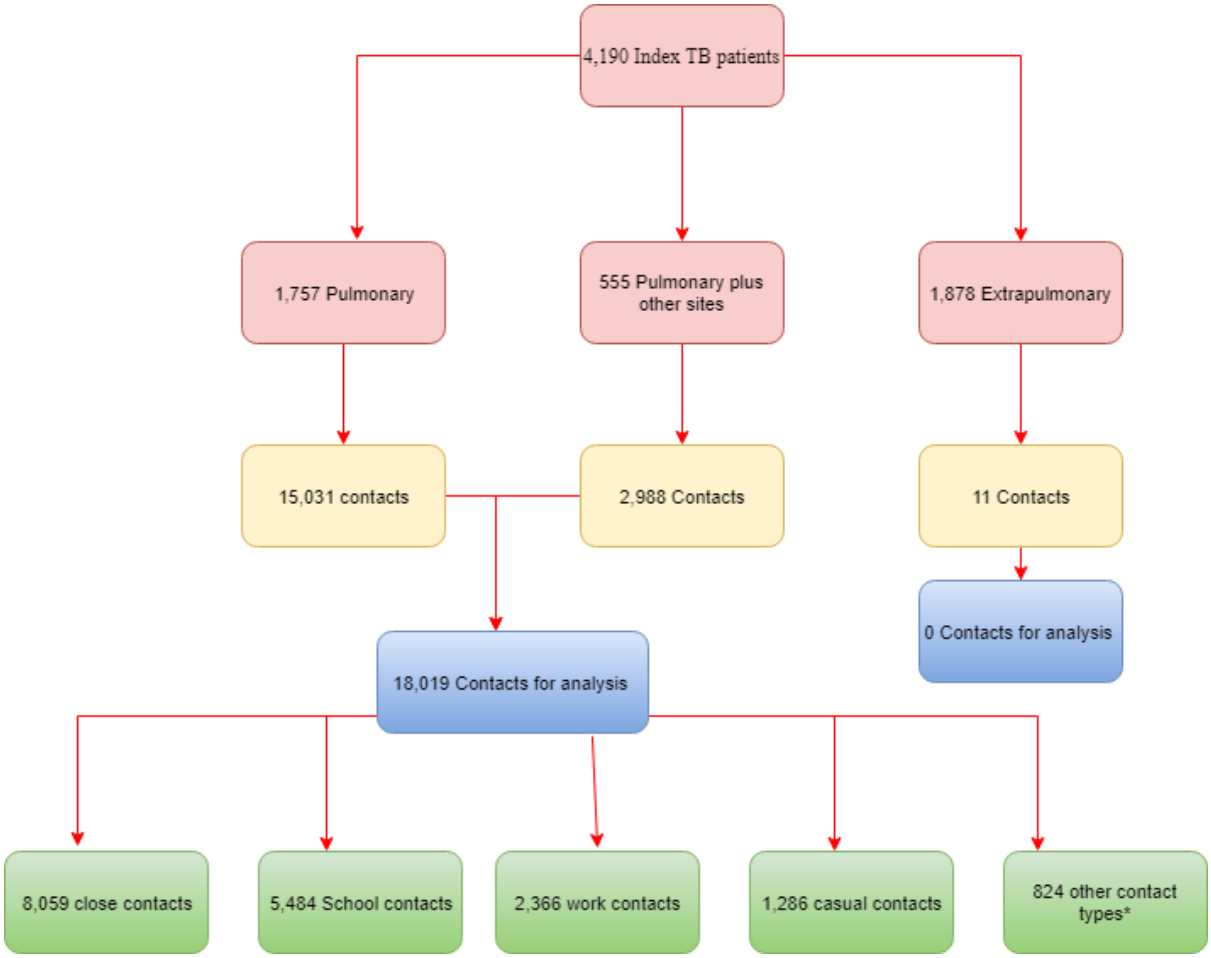
Author details
[1]Department of Epidemiology and Preventive Medicine, School of Public Health and Preventive Medicine, Monash University, Melbourne, Victoria, Australia. [2]Department of Epidemiology and Biostatistics, Institute of Public Health, University of Gondar, Gondar, Ethiopia. [3]Australian Institute of Tropical Health and Medicine, James Cook University, Townsville, Queensland, Australia. [4]Department of Medicine at Royal Melbourne Hospital, University of Melbourne, Melbourne, Victoria, Australia. [5]The Victorian Tuberculosis Program at the Peter Doherty Institute, Melbourne, Victoria, Australia. [6]Department of Microbiology and Immunology, University of Melbourne, Melbourne, Victoria, Australia. [7]Department of Health and Human Services, Health Protection branch, Melbourne, Victoria, Australia.

Received: 4 November 2017 Accepted: 4 March 2019
Published online: 12 March 2019

References
1. WHO. Global tuberculosis report 2016. Switzerland: World Health Organization; 2016.
2. WHO. The End TB Strategy: Global strategy and targets for tuberculosis prevention, care and control after 2015. Geneva: World Health Organization; 2014.
3. Pai M, Behr MA, Dowdy D, Dheda K, Divangahi M, Boehme CC, Ginsberg A, Swaminathan S, Spigelman M, Getahun H, et al. Tuberculosis. Nat Rev Dis Primers. 2016;2:16076.
4. Lloyd-Smith JO, Schreiber SJ, Kopp PE, Getz WM. Superspreading and the effect of individual variation on disease emergence. Nature. 2005;438(7066):355–9.
5. Shen Z, Ning F, Zhou W, He X, Lin C, Chin DP, Zhu Z, Schuchat A. Superspreading SARS events, Beijing, 2003. Emerg Infect Dis. 2004;10(2):256–60.
6. Small M, Tse C, Walker DM. Super-spreaders and the rate of transmission of the SARS virus. Physica D: Nonlinear Phenomena. 2006;215(2):146–58.
7. Stein RA. Super-spreaders in infectious diseases. Int J Infect Dis. 2011;15(8):e510–3.
8. Curtis AB, Ridzon R, Vogel R, McDonough S, Hargreaves J, Ferry J, Valway S, Onorato IM. Extensive transmission of mycobacterium tuberculosis from a child. N Engl J Med. 1999;341(20):1491–5.
9. Galvani AP, May RM. Epidemiology: dimensions of superspreading. Nature. 2005;438(7066):293–5.
10. Ypma RJ, Altes HK, van Soolingen D, Wallinga J, van Ballegooijen WM. A sign of superspreading in tuberculosis: highly skewed distribution of genotypic cluster sizes. Epidemiology. 2013;24(3):395–400.
11. Melsew YA, Doan TN, Gambhir M, Cheng AC, McBryde E, Trauer JM. Risk factors for infectiousness of patients with tuberculosis: a systematic review and meta-analysis. Epidemiology & Infection. 2018;146(3):345-53.
12. Department of Health & Human Services. Management, control and prevention of tuberculosis: Guidelines for health care providers; Victorian Government; 2015.
13. Mycobacterial infections (tuberculosis) [https://www2.health.vic.gov.au/public-health/infectious-diseases/disease-information-advice/tuberculosis]. Accessed 7 Apr 2017.
14. Moyo N, Tay E, Denholm J. Evaluation of tuberculin skin testing in tuberculosis contacts in Victoria, Australia, 2005–2013. Public Health Action. 2015;5(3):188–93.
15. Dale K, Tay E, Trevan P, Denholm J. Mortality among tuberculosis cases in Victoria, 2002–2013: case fatality and factors associated with death. Int J Tuberc Lung Dis. 2016;20(4):515–23.
16. Lloyd-Smith JO. Maximum likelihood estimation of the negative binomial dispersion parameter for highly overdispersed data, with applications to infectious diseases. PLoS One. 2007;2(2):e180.
17. Fisher RA. The negative binomial distribution. Ann Eugenics. 1941;11(1):182–7.
18. Bliss CI, Fisher RA. Fitting the negative binomial distribution to biological data. Biometrics. 1953;9(2):176–200.
19. Venables WN, Ripley BD. Modern Applied Statistics with S, Fourth edn. New York: Springer; 2002.
20. RCoreTeam. R: A Language and Environment for Statistical Computing. Vienna: R Foundation for Statistical Computing; 2014.
21. Lloyd-Smith JO, Schreiber SJ, Getz WM. Moving beyond averages: Individual-level variation in. In: Mathematical Studies on Human Disease Dynamics: Emerging Paradigms and Challenges: AMS-IMS-SIAM Joint Summer Research Conference on Modeling the Dynamics of Human Diseases: Emerging Paradigms and Challenges, July 17-21, 2005, Snowbird, Utah: 2006: American Mathematical Soc; 2006. p. 235.

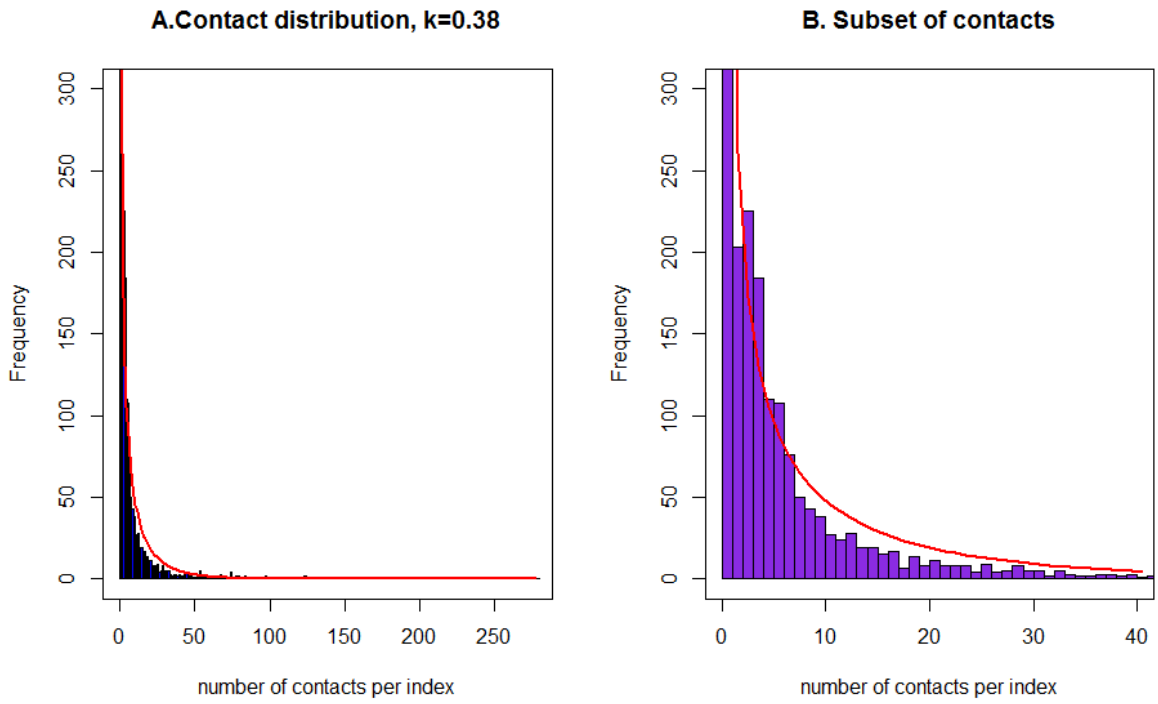
22. Kleiber C, Zeileis A. Visualizing count data regressions using rootograms. Am Stat. 2016;70(3):296–303.
23. Flynn M, Brown L, Tesfai A, Lauer T. Post-migration screening for active tuberculosis in Victoria, Australia. Int J Tuberc Lung Dis. 2012;16(1):50–4.
24. Trauer JM, Moyo N, Tay E-L, Dale K, Ragonnet R, McBryde ES, Denholm JT. Risk of active tuberculosis in the five years following infection... 15%? Chest J. 2016;149(2):516–25.
25. Woolhouse ME, Dye C, Etard J-F, Smith T, Charlwood J, Garnett G, Hagan P, Hii J, Ndhlovu P, Quinnell R. Heterogeneities in the transmission of infectious agents: implications for the design of control programs. Proc Natl Acad Sci. 1997;94(1):338–42.
26. Carvalho AC, Deriemer K, Nunes ZB, Martins M, Comelli M, Marinoni A, KRITSKI AL. Transmission of mycobacterium tuberculosis to contacts of HIV-infected tuberculosis patients. Am J Respir Crit Care Med. 2001;164(12):2166–71.
27. Faksri K, Reechaipichitkul W, Pimrin W, Bourpoern J, Prompinij S. Transmission and risk factors for latent tuberculosis infections among index case-matched household contacts. Southeast Asian J Trop Med Public Health. 2015;46(3):486.
28. Godoy P, Cayla JA, Carmona G, Camps N, Alvarez J, Rodes A, Altet N, Pina JM, Barrabeig I, Orcau A, et al. Immigrants do not transmit tuberculosis more than indigenous patients in Catalonia (Spain). Tuberculosis. 2013;93(4):456–60.
29. Lohmann EM, Koster BFPJ, Le Cessie S, Kamst-van Agterveld MP, Van Soolingen D, Arend SM. Grading of a positive sputum smear and the risk of mycobacterium tuberculosis transmission. Int J Tuberc Lung Dis. 2012;16(11):1477–84.
30. Tornee S, Kaewkungwal J, Fungladda W, Silachamroon U, Akarasewi P, Sunakorn P. Risk factors for tuberculosis infection among household contacts in Bangkok, Thailand. Southeast Asian J Trop Med Public Health. 2004;35(2):375–83.
31. Golub J, Bur S, Cronin W, Gange S, Baruch N, Comstock G, Chaisson R. Delayed tuberculosis diagnosis and tuberculosis transmission. Int J Tuberc Lung Dis. 2006;10(1):24–30.
32. Lin X, Chongsuvivatwong V, Lin L, Geater A, Lijuan R. Dose–response relationship between treatment delay of smear-positive tuberculosis patients and intra-household transmission: a cross-sectional study. Trans R Soc Trop Med Hyg. 2008;102(8):797–804.
33. Mendes MA, Gaio R, Reis R, Duarte R. Contact screening in tuberculosis: can we identify those with higher risk? Eur Respir J. 2013;41(3):758–60.
34. O'Shea MK, Koh GC, Munang M, Smith G, Banerjee A, Dedicoat M. Time-to-detection in culture predicts risk of mycobacterium tuberculosis transmission: a cohort study. Clin Infect Dis. 2014;59(2):177–85.

Supplementary Material

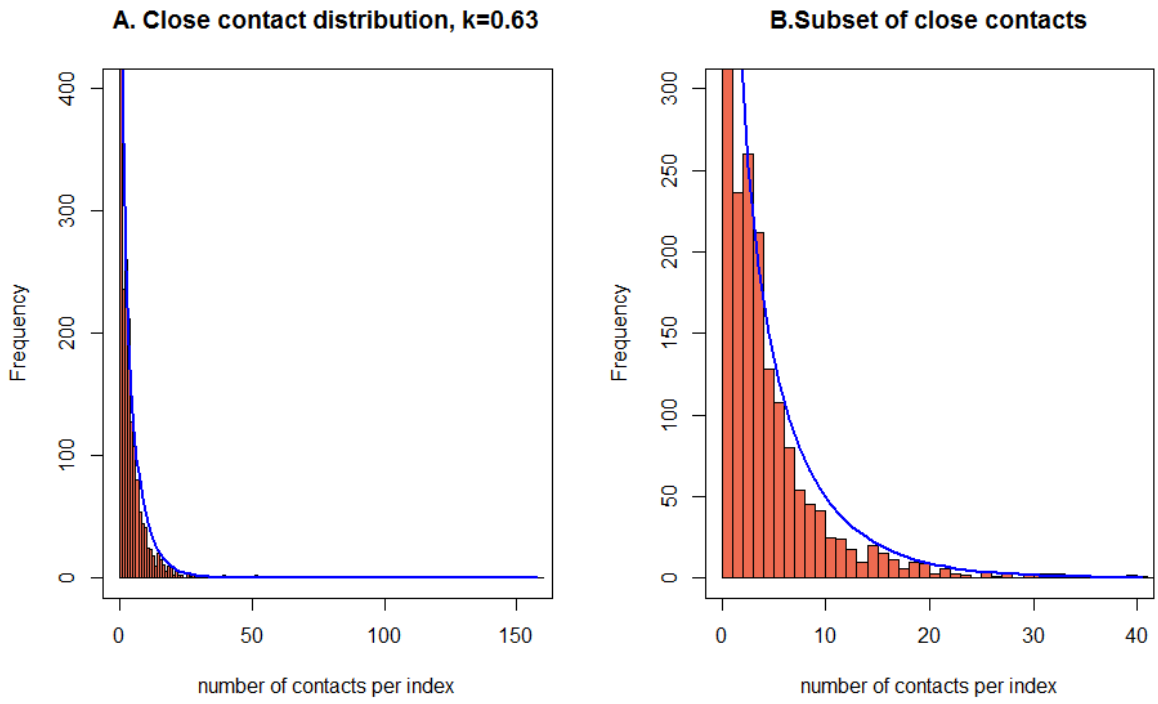


Other contact types* =includes contacts from hospitals, nursing homes, airlines and childcare facilities.

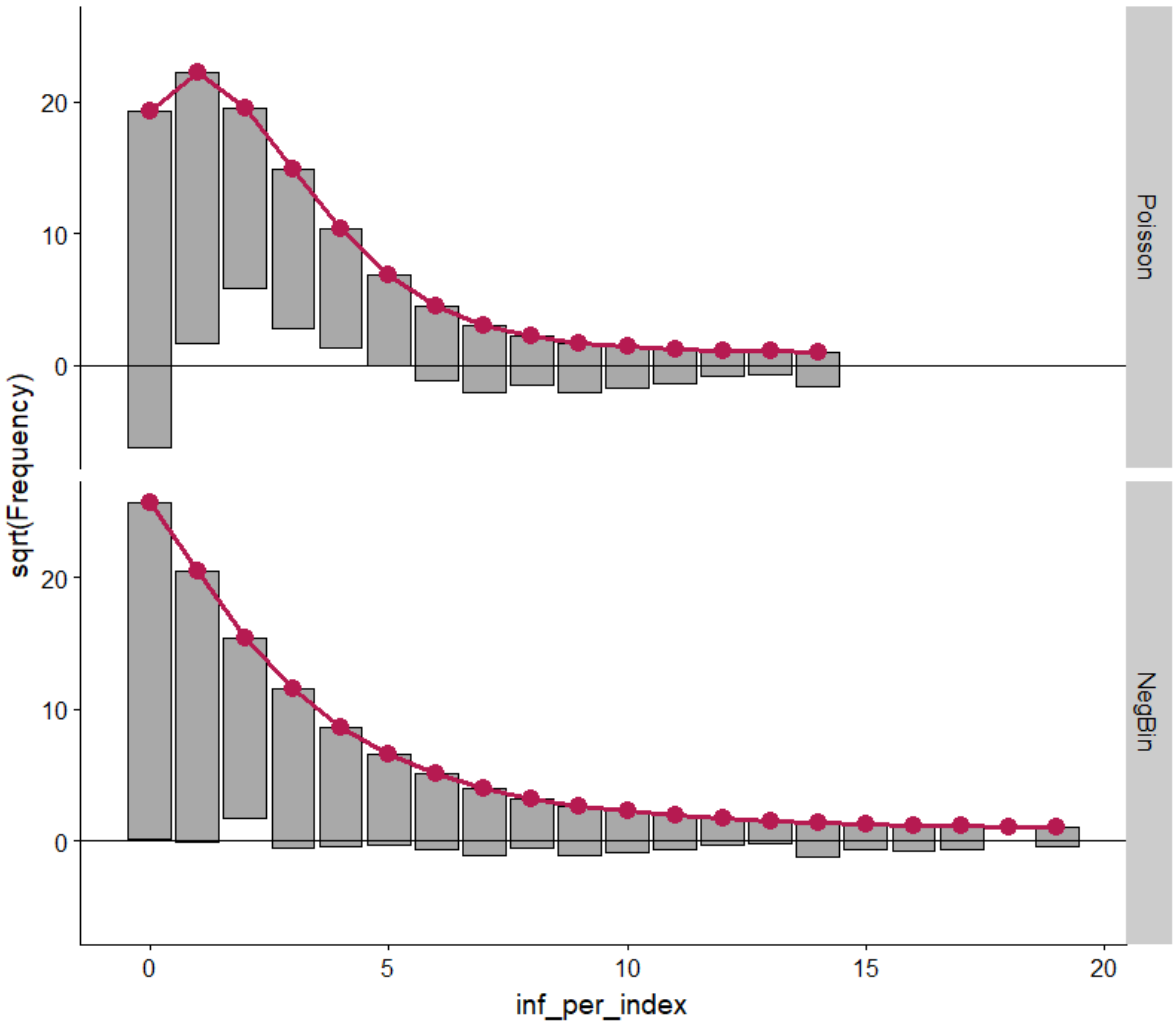
Supplementary figure 1: Schematic presentation of number of index TB patient and contacts in Victoria, for the period 2005-2015.



Supplementary figure 2: Distribution of number of contacts per index TB patient in Victoria, for the period 2005-2015. A. All contacts (with negative binomial distribution fitted to count data) strategies, the number of index patients with zero contacts was 639 (beyond limit of vertical axis). B. Subset of contacts (0-40 contacts per index only).



Supplementary figure 3: Distribution of number of Close contacts per index TB patient in Victoria, for the period 2005-2015. A. All Close contacts (with negative binomial distribution fitted to count data). B. Subset of close contacts (0-40 close contacts per index), the number of index patients with zero close contacts was 653 (beyond limit of vertical axis).



Supplementary figure 4: Hanging rootograms for a Poisson model (upper panel) and negative binomial model (lower panel) count data.

The figure shows the difference between predicted counts given the model (thick red line with dots) and observed counts (shown as bars hanging from the red dots). The x-axis is the outcome variable (number of secondary infections per index) while the y-axis is square root of the observed or expected count. The reference line is drawn at a height of zero, such that bars hanging below this reference line indicate under-prediction while bars hanging above zero indicate over-prediction. The Poisson model shows evidence of under-prediction for zero counts and most counts above five while the negative binomial model had less under prediction and predicted zero counts well. The Poisson model also over predicted the 1-4

counts, while in the negative binomial model over prediction was observed only for two counts. Therefore, the negative binomial model predicted counts well and considerably better than the Poisson model.

Chapter 4

4. Profiles of tuberculosis disease activation among contacts of patients with tuberculosis in a low-transmission setting

4.1. Chapter overview

In addition to analysing heterogeneity in infection transmission, analysis of the heterogeneity in progression to active TB after being exposed to an infectious TB patient can provide a more complete picture of *Mtb* transmission and disease progression. Using contact investigation data from the Victorian TB program, this Chapter explores the profiles of contacts' active TB progression and identifies factors associated with the rate of active TB disease progression among contacts. This Chapter documents the patterns of disease activation among contacts of TB patients in detail, highlighting those at higher risk of TB after exposure and considering implications for interventions. The Chapter shows that TB disease progression among contacts of TB patients was found to be substantially heterogeneous, provides recommendations on the follow-up and preventive treatment to be based on detailed characteristics of individual contacts and their associated index patients.

This Chapter addresses thesis question 3, and it is published in shorter form as a research letter, which is attached as an ***Appendix B*** to this thesis.

Citation: **Melsew YA**, Cheng AC, McBryde ES, Denholm JT, Tay E, Ragonnet R, Trauer JM. Profiles of tuberculosis disease activation among contacts of patients with tuberculosis. European Respiratory Journal. 2019 Jan 1:1900353.

4.2. Abstract

Background: The risk of a person acquiring *Mycobacterium tuberculosis* infection and then progressing to tuberculosis (TB) disease depends both on the level of the index patient's infectiousness and on the susceptibility of the exposed contact. The aim was to assess patterns and identify risk factors for TB disease activation among contacts of pulmonary TB patients in Victoria, Australia.

Methods: A cohort study was performed using records on contacts of pulmonary TB (PTB) patients notified from 1 January 2005 to 31 December 2015 in Victoria, Australia. The main outcomes of interest were primary progression (development of active TB within the first six months of exposure) and late reactivation (development of active TB after six months of exposure) among contacts. Survival analysis was performed to compare the rates of TB disease based on various contact characteristics. Logistic regression was used to identify factors associated with primary progression, while a Cox-proportional hazard model was applied to identify factors associated with late TB reactivation.

Results: A total of 17,740 contacts, identified from a total of 1,672 index patients with PTB, were included in the analyses. There were 224 cases of active TB diseases among contacts, and the total person-years of follow-up was 118,276 person-years, giving a crude incidence rate of 189 cases of active TB per 100,000 person-years of follow-up. Of those who developed active TB, 140 (62.5%) were primary progression. Higher rates of TB disease were found among males, children and TST-positive contacts (log-rank $p<0.05$). The vast majority of TB disease among children were primary progressions, although high rates of late reactivation were observed as children reached 15 years of age. From regression analyses, it was clear that younger age, close contact, high TB burden country of birth and exposure to index patients with recurrent TB were all associated with a higher risk of TB disease activation. BCG vaccination history appeared to be protective for primary progression but a

risk factor for late reactivation, with the large majority of adult late reactivations occurring in BCG-vaccinated contacts.

Conclusions: The pattern of TB disease activation among children is mainly primary progression, but late reactivations emerge as children reach age ≥15years. The possibility that BCG increases the rates of late reactivation disease in adults should be further explored. Preventive therapy and lack of accounting for it may have biased the results in unpredictable ways.

Keywords: Tuberculosis, primary progression, late reactivation, Australia

4.3. Introduction

With an estimated 10.4 million cases and 1.7 million deaths, tuberculosis (TB) is the ninth leading cause of death worldwide [1]. The World Health Organization (WHO) devised the ambitious *End TB Strategy* with targets for 2035, including a 90% reduction in incidence compared with 2015. One of the pillars of this strategy is integrated patient-centred care and prevention that can be achieved through intensified contact investigation [2].

The risk of a person acquiring *Mycobacterium tuberculosis* (*Mtb*) infection and then progressing to TB disease varies due to factors related to index patients' infectiousness and the level of susceptibility of the exposed contact [2, 3]. Some contacts develop TB in the early period following exposure, while others take many years to progress to disease or may never do so [4]. Factors related to the susceptibility of contacts have been widely studied; however, the risk associated with the index case is not well defined other than a few factors such as sputum smear status and cavitation [5-7].

Previous studies to identify factors associated with TB disease among contacts have found that social and behavioural factors play an important role in *Mtb* transmission [8]. Behaviours such as alcohol use and smoking are known risk factors for subsequent TB disease among contacts [9, 10], while the risk of TB disease may also vary with age, partly due to an accumulation of these risk factors [2]. However, it is also essential to distinguish the factors that are associated with different progression profiles (such as early progression and late reactivation) after exposure to an index patient.

In the state of Victoria in Australia, contact investigation is routinely performed, and active TB arising in contacts is linked to index cases. A previous study from this linked data process estimated the absolute risk of TB disease in the years after infection [3]. The current study advances on this earlier work; in addition to having both a longer follow-up and a larger

cohort, it estimates profiles, patterns and risk factors for TB disease. Identifying and quantifying these risk factors will help to target those contacts who are at greatest risk of developing active TB disease for follow-up and early intervention. Thus the aim of this study was to determine the rate of active TB and the risk factors for primary progression and late reactivation, including factors associated with the index case, among contacts of TB patients in Victoria, Australia.

4.4. Methods

Study design

A cohort study was constructed using data from contacts of pulmonary TB (PTB) patients notified from 1 January 2005 to 31 December 2015 to the Victorian Tuberculosis Program (VTP), Australia, linked to subsequent episodes of active TB in the same jurisdiction. We looked at the contacts of TB patients from the point of their respective index patient's diagnosis and compared the time to active TB disease development.

Setting

The study was conducted in Victoria, a state of Australia, with approximately 5.9 million people and a single centralised tuberculosis program (the Victorian Tuberculosis Program; VTP). Medical practitioners and diagnostic laboratories are required under Victorian public health legislation to report cases of TB to health authorities, and the management of all notified cases of TB and their contacts is coordinated by the VTP [11-14]. Contact investigation initially consists of clinical assessment and serial testing for *Mtb* infection. Testing of contacts is conducted by either tuberculin skin testing (TST) using the Mantoux method or an interferon-gamma release assay (IGRA), although during the study period the vast majority of testing was undertaken with TST [15]. Contacts with either symptom suggestive of active disease or a positive test for *Mtb* infection undergo a chest x-ray and

further clinical assessment, with isoniazid preventive treatment generally recommended for those where the active disease has been excluded [11].

Study population

The study population includes all contacts of pulmonary TB (PTB) patients recorded and followed by the VTP. All documented contacts of PTB patients notified from 1 January 2005 to 31 December 2015 were included. Contacts' follow-up information was linked to their respective index TB patients' information by the index identification number. We excluded contacts from analysis if their age was unavailable in the dataset but did not exclude records if other variables were absent.

Definitions

Index patients were pulmonary TB patients defined as confirmed cases with definitive laboratory evidence or clinical evidence of TB according to the Australian standard national case definition for TB [16]. Contacts were defined as individuals identified as having been exposed to index patients by the VTP through households, schools, workplaces, and other settings. In Victoria, close contacts are defined as "people who have had frequent, prolonged and close contact in an enclosed environment with an infectious cause such as all people living in the same dwelling; relatives and friends who have frequent, prolonged and close contact; and work colleagues who share the same indoor work areas on a daily basis, following an individual risk assessment" [11]. Contacts with a TST response ≥10mm were classified as TST-positive. The outcome event of interest was the occurrence of active TB disease among contacts. The contact's country of birth was categorised as high (≥100 cases per 100,000 population) or low burden (<100 cases per 100,000 population) according to WHO definition [17]. Active TB in contacts was defined in the same way as active TB in index patients [16].

Primary progression was defined as the occurrence of active TB among contacts within the first six months of exposure, while late reactivation was defined as the occurrence of active TB after six months of exposure. Time to event was the time in days from the date of index patients' notification to the date of contact active TB notification. Records were censored if contacts had not developed active TB by the date of data extraction (21 March 2017).

Variables

The main outcomes of interest were primary progression (development of active TB within the first six months following exposure) and late TB reactivation (development of TB after six months of exposure). Risk factors that were obtained from the database included baseline characteristics of both the exposed individual or contact and the index patients. Index patients' available characteristics included age, sex, country of birth, method of case finding (i.e. whether the case was identified through active case finding or passive presentation), site of TB disease (pulmonary only/pulmonary plus other sites), sputum smear status, type of case (new/relapse), chest x-ray result, method of TB diagnosis and drug susceptibility testing results (where available). Contacts' available characteristics included age, sex, country of birth, type of contact, tuberculin skin test (TST) size and previous Bacille Calmette-Guérin (BCG) vaccination.

Statistical Methods

The incidence rate of active TB among infected contacts was calculated per person-year of follow-up. To compare survival times between exposure groups, Kaplan-Meier curves were plotted and the significance of differences was tested with log-rank tests, with p-values less than 0.05 considered significant. We fitted a logistic regression model for all the included contacts to identify factors associated with primary progression. Crude and adjusted odds ratios (OR) were provided with their 95% confidence intervals (95% CI) for bivariate and multivariate analyses respectively.

For those adult contacts (age ≥15years) who did not have primary progression, Cox-proportional hazard model was fitted to identify risk factors for late TB reactivation. All variables fulfilling the assumption of proportionality of hazard functions with Schoenfeld residuals test were included. The crude and adjusted hazard ratios (HR) were provided with their 95% CI. We did not include TST results in the multivariate analysis due to its expected correlation with variables reflecting the infectiousness of the index case.

Findings were reported consistent with the "Reporting of studies Conducted using Observational Routinely-collected health Data (RECORD) Statement" [18].

Ethical considerations

Ethical approval was obtained from Monash University, Human Research Ethics Committee (Project Number: 7776) and approval for data access and support for the project given by the VTP and DHHS. Data were extracted from a centralised notifiable disease database stored by the Victorian Department of Health and Human Services (DHHS). Non-identifiable data were used for analysis. The data set includes linked contact tracing information, along with the results of their testing for *Mtb* infection and subsequent active TB during the study period.

4.5. Results

Demographic and clinical characteristics

A total of 18,019 contacts were identified from 1,672 PTB index patients notified within the period 1 January 2005 to 31 December 2015. Of these contacts, 17,740 (98.5%) were included in the analysis, while 279 (1.5%) contacts were excluded because age information was unavailable. The majority (14,543 or 82.0%) of contacts were adults aged 15 years and above. Considering the type of contact, the greatest proportion were close contacts (45.1%), followed by school contacts (30.4%). At the end of the follow-up period, 224 (1.3%) contacts had developed active TB (while none of the excluded contacts developed active TB).

Comparing contacts based on their respective index patient characteristics, 53.6% had contact with male index patients, and 41.0% had contact with index patients aged 15-24 years (**Table 1**).

Information about preventive therapy was not routinely available in the dataset used for this study. Data regarding preventive therapy use in 448 contacts were available from a previous study in a sample of this cohort [3]. Based on these data, 83.3% (35/42) of those who developed active TB did not have had preventive treatment, while a smaller proportion of disease-free contacts were not treated (37.4%, **Table 2**).

Table 1: Number of contacts by their own and by their index patients' demographic and clinical characteristics in Victoria, Australia 2005-2015.

Contact characteristics	**Category**	**Number of contacts, n (%)**	**TB, n (%)**	**No TB, n (%)**
Sex	Female	9194(51.8)	101(1.1)	9093(98.9)
	Male	8427(47.5)	123(1.5)	8304(98.5)
	Unknown	119(0.7)	0(0)	119(100)
Age in years	0-4	1197(6.7)	51(4.3)	1146(95.7)
	5-14	2000(11.3)	60(3.0)	1940(97.0)
	15 and above	14543(82.0)	113(0.8)	14430(99.2)
Country of birth	High TB burden country	5130(28.9)	84(1.6)	5046(98.4)
	Low TB burden country	11924(67.2)	139(1.2)	11785(98.8)
	Unknown	686(3.87)	1(0.1)	685(99.9)
Type of contact	Casual contact	1281(7.2)	18(1.4)	1263(98.6)
	Close contact	7994(45.1)	190(2.4)	7804(97.6)
	Environmental contact*	761(4.3)	2(0.3)	759(99.7)
	School contact	5388(30.4)	7(0.1)	5381(99.9)
	Work contact	2316(13.1)	7(0.3)	2309(99.7)
TST	Positive	3213(12.9)	148(4.6)	3065(95.4)
	Negative	12233(69.0)	29(0.2)	12204(99.8)
	No TST	2294(18.1)	47(2.0)	2247(98.0)
BCG	Yes	7542(42.5)	86(1.7)	4978(98.3)
	No	5064(28.6)	54(1.1)	5080(98.9)
	Not stated/Unknown	5134(28.9)	84(1.1)	7458(98.9)
Contacts by their index patient's characteristics				
Sex	Female	8226(46.4)	93(1.1)	8133(98.9)
	Male	9514(53.6)	131(1.4)	9383(9383)
Age in years	1-14	531(3.0)	10(1.9)	521(98.1)
	15-24	7286(41.0)	82(1.1)	7204(98.9)
	25-44	5254(29.6)	100(1.9)	5154(98.1)

	45-64	1730(9.8)	23(1.3)	1707(98.7)
	65 and above	2939(16.6)	9(0.3)	2930(99.7)
Country of birth	High TB burden country	8766(49.4)	115(1.3)	8859(98.7)
	Low TB burden country	8974(50.6)	109(1.2)	8657(98.8)
Method of case ascertainment	Clinical presentation	16760(94.5)	205(1.2)	16555(98.8)
	Active case finding±	980(5.5)	19(1.9)	961(98.1)
Site of TB disease	Pulmonary	14777(83.3)	197(1.3)	14580(98.7)
	Pulmonary plus other sites	2963(16.7)	27(0.9)	2936(99.1)
Type of case	New case	16714(94.2)	201(1.2)	16513(98.8)
	Relapse	784(4.4)	21(2.7)	763(97.3)
	Unknown	242(1.4)	2(0.8)	240(99.2)
Chest x-ray	Abnormal, no cavity	10251(57.8)	86(1.4)	5928(98.6)
	Abnormal, cavity	6014(33.9)	125(1.2)	10126(98.8)
	Abnormal, abnormality not specified	266(1.5)	1(0.4)	265(99.6)
	Normal	341(1.92)	3(0.9)	338(99.1)
	Unknown	868(4.89)	9(1.0)	859(99.0)
Sputum smear positive	Yes	11873(66.9)	154(1.3)	11719(98.7)
	No	2227(12.6)	24(1.1)	2203(98.9)
	Unknown	3640(20.5)	46(1.3)	3594(98.7)
Diagnosis	Culture	17229(97.1)	217(1.3)	17012(98.7)
	Radiological	297(1.7)	6(2.0)	291(98.0)
	Clinical signs	17(0.1)	0(0)	17(100.0)
	Histology	26(0.15)	0	0(100.0)
	Microscopic examination	26(0.15)	0	0(100.0)
	PCR/NAT	145(0.8)	1(0.7)	144(99.3)
Drug susceptibility	Fully sensitive	15513(87.4)	190(1.2)	15323(98.8)
	MDR-TB	462(2.6)	10(2.2)	452(97.8)
	XDR-TB	13(0.1)	0(0)	13(100.0)
	Other resistance profile	1219(6.9)	20(1.6)	1199(98.4)
	Unknown	533(3.0)	4(0.8)	529(99.2)

*This classification includes other congregate settings such as airline, aged care, hospital, ± includes the health undertaking program, contact tracing, screening and routine follow up, TST: tuberculin skin test (If TST≥10mm positive), BCG: bacille Calmette-Guérin, PCR/NAT: polymerase chain reaction, MDR: multi-drug resistant, XDR: extensively drug-resistant

Table 2: Treatment of latent infection status and its relation with contact characteristics among contacts of PTB patients in Victoria, Australia 2005-2015. (n=448)

Contact characteristics		**Preventive treatment status**				
		No treatment	Commenced not completed	Commenced treatment	Completed treatment	Chi-square p-value
Sex	Female	87(40.6%)	4(1.9%)	83(38.8%)	40(18.7%)	0.6409
	Male	100(42.7%)	10(4.3%)	79(33.8%)	45(19.2%)	
Age	0-4 years	26 (48.2%)	1 (1.8%)	18 (33.3%)	9(16.7%)	0.8862
	15 years and above	120(39.9%)	9(3.0%)	113(37.5%)	59-19.6	
	5-14 years	41(44.1%)	4(4.3%)	31(33.3%)	17(18.3%)	

Type of contact	Close contact	186(41.6%)	14(3.13%)	162(36.2%)	85(19.0%)	
	Other contacts	1	0	0	0	
Country of birth	High burden country	32(36.4%)	4(4.5%)	39(44.3%)	13(14.8%)	0.126
	Low burden country	155(43.1%)	10(2.8%)	123(34.2%)	72(20.0%)	
TST	Negative	36(52.2%)	0	22(31.9%)	11(15.9%)	
	Positive	146(39.9%)	13(3.5%)	135(36.9%)	72(19.7%)	
	No result	5(38.5%)	1(7.7%)	5(38.5%)	2(15.4%)	
BCG	No	66(46.8%)	4(2.84%)	45(31.9%)	26(18.4%)	
	Yes	96(36.9%)	10(3.8%)	104(40.0%)	50(19.2)	
	Unknown	25(53.2%)	0	13(27.7%)	9(19.2%)	
TB disease	No	152(37.4%)	13(3.2%)	157(38.7%)	84(20.7%)	<0.0001*
	Yes	35(83.3%)	1(2.4%)	5(11.9%)	1(2.4%)	

Rate of TB disease

Over a median duration of follow up of 7.1 years (95% CI: 6.3-7.2), there were 224 episodes of active TB during the follow-up period while the remainder (n=17516, 98.7%) remained disease-free until the end of the follow-up period. Of those who developed active TB, 140 (62.5%) were notified in the first six months of follow-up (primary progressions), while 153 (68.3%) and 176 (78.6%) of the TB cases were found in the first year and second years of follow up, respectively (**Figure 1**). The total number of person-years of follow up was 118,276, giving an incidence rate of 189 cases of active TB per 100,000 person-years of follow-up among contacts. Stratifying by age group of contacts, the incidence rates among contacts aged under 15 years and 15 years and above were 533 and 116 cases of TB per 100,000 person-years of follow-up, respectively.

Patterns of TB disease

There were differences in the incidence of active TB among contacts based on their demographic and clinical characteristics and those of their index TB patient. With respect to contact characteristics, the exposure variables of sex, age, and TST showed significant differences (**Figure 2**). Male contacts had a slightly higher rate of TB than their female

counterparts (log-rank p=0.03), while children aged <5 years at exposure had the greatest rate of active TB development, followed by children aged 5-14 years (log-rank p<0.0001). All TB cases among children aged <5 years were observed in the first 18 months of follow-up, of which 44/51 (86.3%) were primary progressions while a similar proportion 49/60 (81.6%) of primary progressions was found among TB cases in children aged 5-14 years at exposure. However, in the 5-14 years age group, seven TB activation episodes (all born in low burden settings) occurred after two years of follow-up (up to nine years) and predominantly occurred as the children aged into adolescence and reached approximately 15 years of age. Among TB cases in adult contacts (age ≥15 years), a greater proportion 66/113 (58.4%) were late reactivation, while the remainder were primary progressions.

The TST response was also a significant variable associated with the rate of TB activation among contacts, with a positive TST result (≥10mm) being associated with a greater rate of TB (4.6%, 148/3,213) than those who were TST-negative (0.2%, 29/12,233), despite the fact that TST positives were more likely to receive preventive therapy (**Figure 2**).

Overall, a history of previous BCG vaccination in contacts was associated with a lower incidence of TB activation than the BCG-naïve (log-rank p=0.004), although BCG had different effects according to the age category of contacts (**Figure 3**) with a greater TB incidence observed in adults. To fully illustrate this effect, **Figure 4** shows the distribution of cases by time of disease activation for BCG vaccinated and unvaccinated contacts stratified by their age at exposure. Most TB disease activation among adult contacts (aged 15 years and above) occurred over a longer follow-up period, and only three of the late reactivations in adults were known to be BCG-unvaccinated.

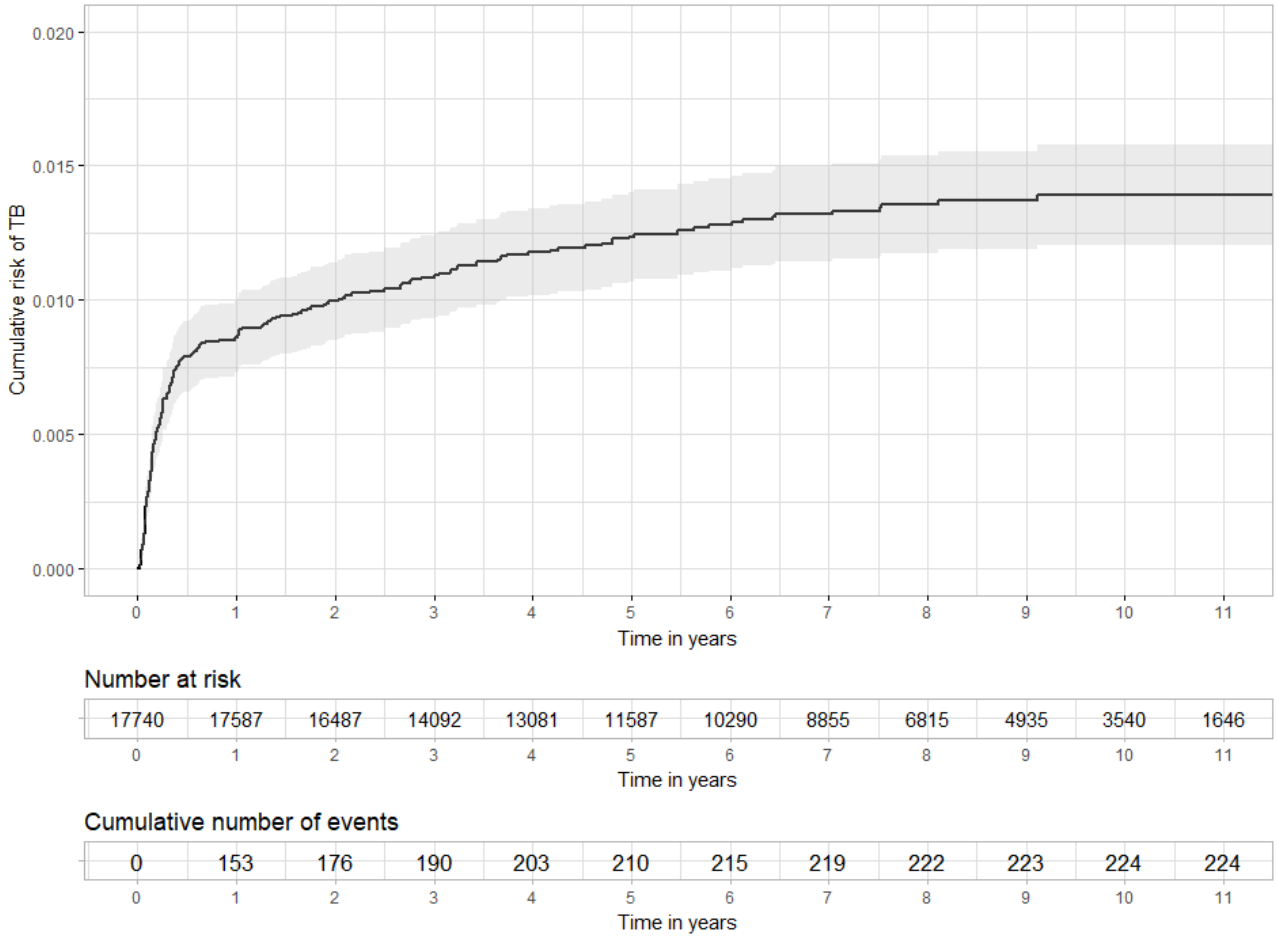


Figure 1: Kaplan-Meier curve for the TB-disease free probability of contacts of TB patients in Victoria, Australia 2005-2015.

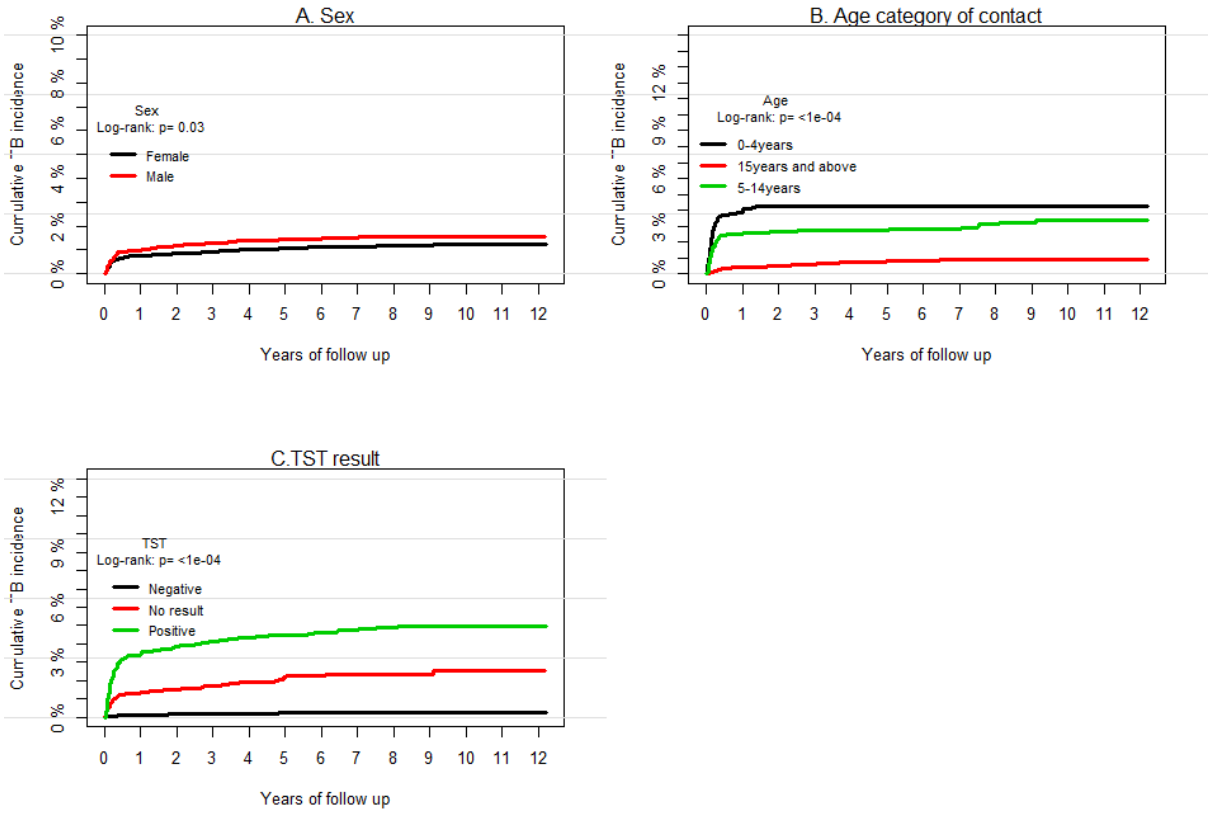


Figure 2: Kaplan-Meier curves for by contact characteristics among contacts of TB patients in Victoria, Australia 2005-2015

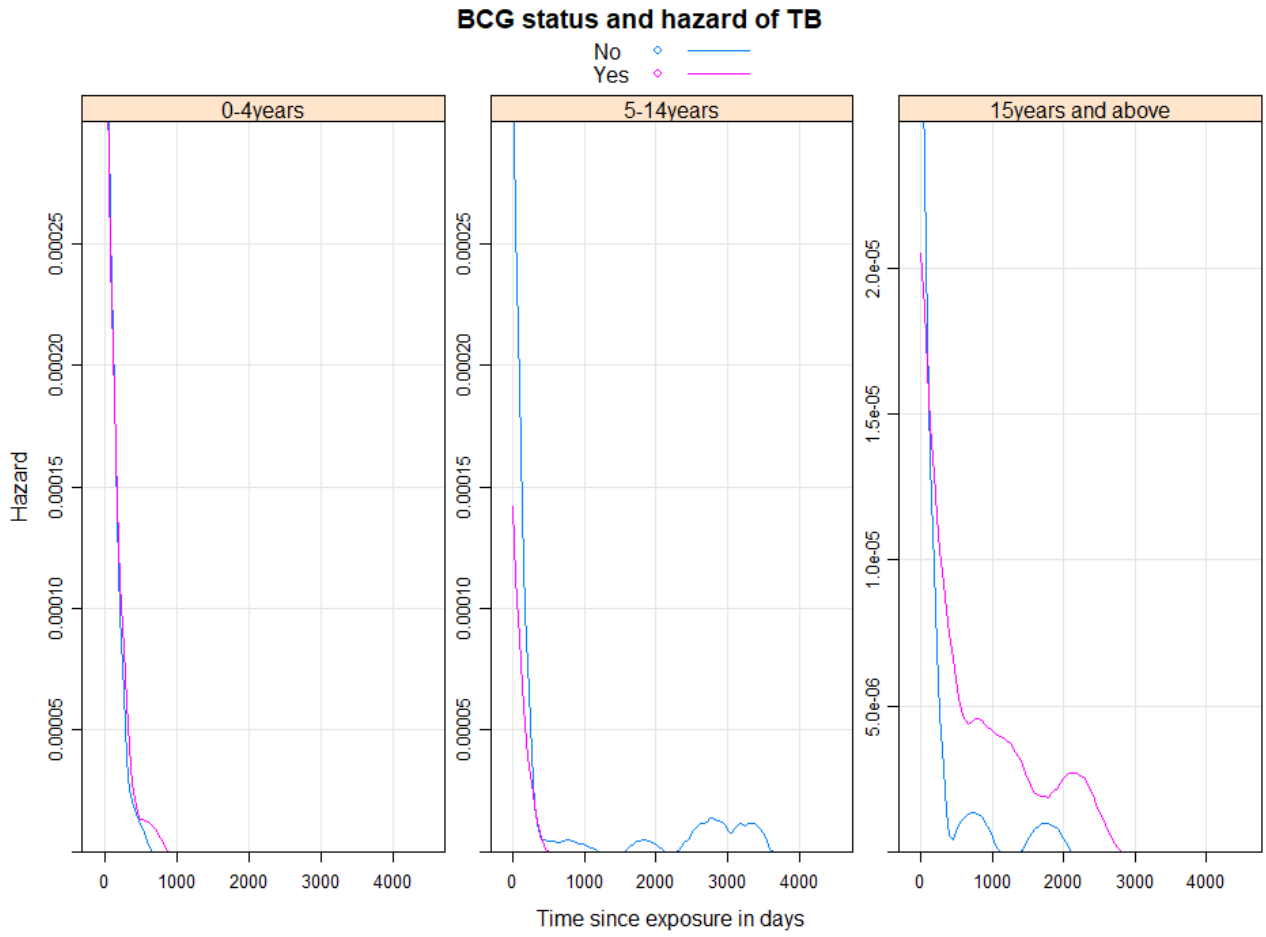


Figure 3: Smoothed hazard estimates by contact BCG vaccination history (No means unvaccinated and Yes is vaccinated) stratified by age among contacts of TB patients in Victoria, Australia 2005-2015 (excluding those with unknown BCG status)

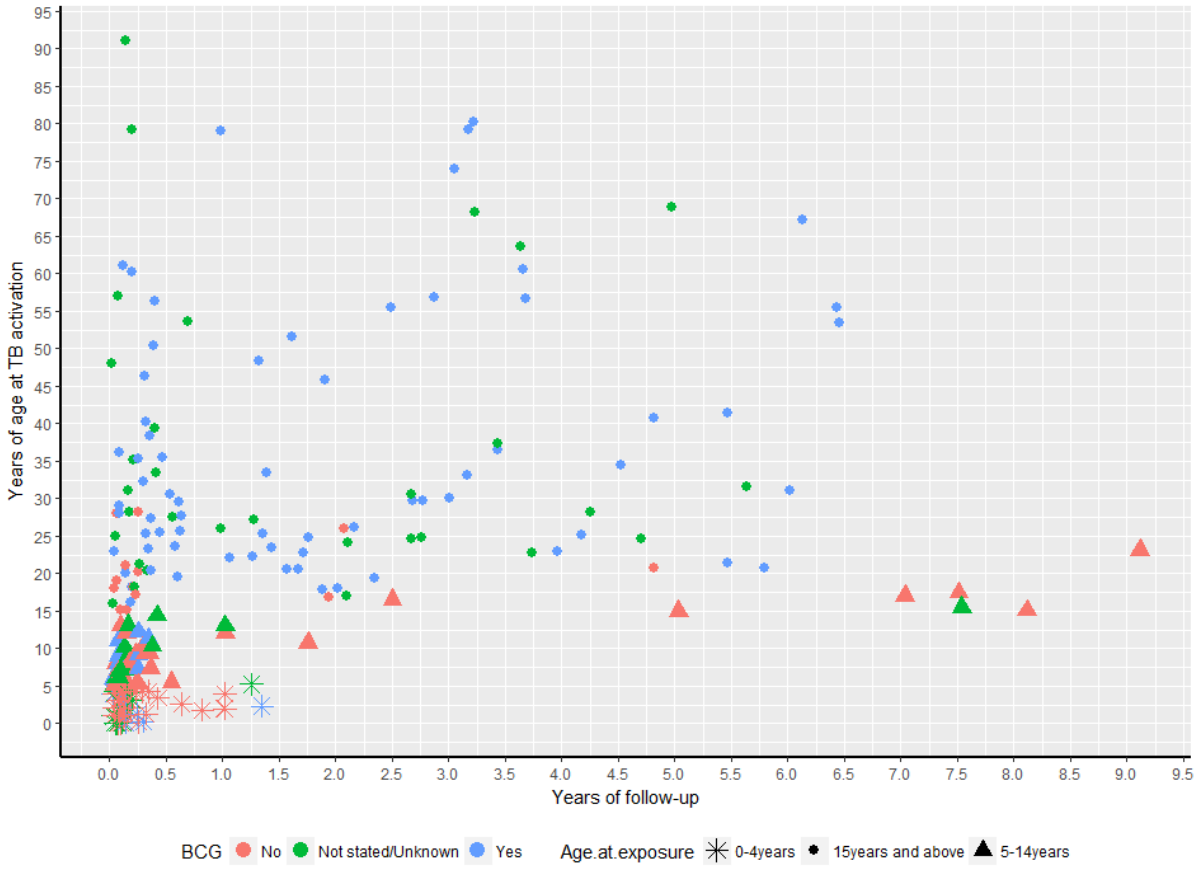


Figure 4: Patterns of contact active TB disease cases throughout the follow-up categorised by age at exposure (shapes) and BCG vaccination history (colour) among contacts of TB patients in Victoria, Australia 2005-2015.

Risk factors for TB disease activation

From the logistic regression model for primary progression, younger age of contact, close contact, high-burden country of birth, absence of BCG vaccination and index patients with relapsed TB were independently associated with primary progression among contacts. On bases of years of age at exposure, compared to adult contacts aged 15 years and above, contacts aged <5 years (AOR=7.79, 95%CI: 4.80-12.62) and 5-14 years (AOR= 5.23, 95%CI: 3.39-8.07) were more likely to experience primary progression. Compared to other contact types, close contacts had a 5.7-fold higher odds of primary progression (AOR=5.73, 95% CI: 3.39-9.69), while contacts born in a high TB burden country were also more likely to progress (AOR=1.76, 95%CI: 1.11-2.81). BCG vaccination was a protective factor against

primary TB progression, with contacts who had been previously BCG vaccinated showing a 53% lower odds of primary progression than those who were BCG-naïve (AOR=0.47, 95% CI: 0.29-0.77). Contacts exposed to patients with relapse TB were at a 2.68-fold greater odds of primary progression than contacts exposed to patients with new TB (AOR=2.68, 95% CI: 1.42-5.06, **Table 2**).

In the Cox-proportional hazard regression analysis for late TB reactivation, close contact, high-burden country of birth and presence of BCG vaccination history were independently associated with contacts' risk of late reactivation. Accordingly, compared to other contact types, close contacts had a considerably higher risk of late TB reactivation (AHR= 3.25, 95%CI: 1.81-5.82). Contacts who were born in high-burden countries also had a greater risk of late reactivation compared to contacts born in low-burden countries (AHR=2.17, 95%CI: 1.26-3.75). BCG vaccination history was a risk factor for late reactivation, with contacts who were BCG vaccinated showing a much higher risk of late reactivation than unvaccinated contacts (AHR= 4.54, 95% CI: 1.35-15.28). **(Table 3)**

Table 3: Risk factor analysis for primary TB progression among contacts of TB patients in Victoria, Australia 2005-2015 (n=17,740).

Contact characteristics	**Category**	**Contact TB disease**		**Crude OR (95%CI)**	**Adjusted OR (95%CI)**	**p-value**
		No	Yes			
Sex	Female	9132	62	Ref		
	Male	8349	78	1.38 (0.98-1.93)	1.23 (0.88-1.73)	0.227
	Not stated	119	0	3.46 (6.53-0.30)	0.00 (0.00-Inf)	0.987
Age	0-4 years	1153	44	11.78 (7.75-17.84)	7.79 (4.80-12.62)	<0.001*
	5-14 years	1951	49	7.75 (5.17-11.61)	5.23 (3.39-8.07)	<0.001 *
	15 years and above	14496	47	Ref		
Type of contact	Close contacts	7871	123	8.94(5.54-15.40)	5.73(3.39-9.69)	<0.001*
	Other contacts	9729	17	Ref	Ref	
Country of birth	High TB burden	5090	40	0.73 (0.52-1.02)	1.76 (1.11-2.81)	0.017*
	Other countries	11824	100	Ref		
	Not stated	686	0	0.00 (0.00-Inf)	0.00 (0.00-Inf)	0.969
TST	Negative	12222	11	Ref	Not included	
	Positive	3112	101	36.06 (20.24-71.31)*		
	Not performed	2266	28	13.73 (7.02-28.83)*		
BCG	Yes	7505	37	0.36 (0.24-0.53)	0.47 (0.29-0.77)	0.002*
	No	4995	69	Ref		
	Not stated/ unknown	5100	34	0.48 (0.32-0.73)	0.79 (0.50-1.23)	0.288
Index characteristics						
Sex	Female	8161	65	1.00 (0.72-1.40)	1.14 (0.80-1.62)	0.485
	Male	9439	75	Ref		
Age	1-14 years	521	10	3.76 (1.76-7.31)	0.20 (0.03-1.45)	0.110
	15-24 years	7249	37	Ref		
	25-44 years	5181	73	2.76 (1.87-4.15)	1.04 (0.69-1.55)	0.860
	45-64years	1713	17	1.94 (1.06-3.40)	1.57 (0.91-2.71)	0.101
	65 years and above	2936	3	0.20 (0.05-0.55)	0.60 (0.32-1.13)	0.111
Country of birth	High TB burden	8897	77	Ref	ref	
	Other countries	8703	77	0.84 (0.60-1.17)	1.15 (0.81-1.63)	0.446
Method of case finding	Clinical presentation	16628	132	Ref	ref	
	Other methods	972	8	1.04 (0.46-1.99)	1.71 (0.91-3.21)	0.096
Site of TB	Pulmonary	14655	122	Ref		
	Pulmonary plus other site	2945	18	0.73 (0.43-1.17)	1.20 (0.76-1.90)	0.435
Type of TB y	New	16584	130	Ref		
	Relapse	774	10	1.65 (0.81-2.99)	2.68 (1.42-5.06)	0.002*
	Unknown	242	0		0.00 (0.00-Inf)	0.988

Chest X-ray result	Abnormal-cavity	5959	55	1.19 (0.84-1.68)	1.20 (0.82-1.75)	0.347
	Abnormal-no cavity	10172	79	Ref		
	Abnormal-no details	265	1	0.49 (0.07-3.51)	0.30 (0.04-2.21)	0.236
	Normal	339	2	0.76 (0.19-3.10)	0.72 (0.17-3.04)	0.655
	Not stated/done	865	3	0.45 (0.14-1.42)	0.42 (0.13-1.36)	0.146
Sputum smear	Yes	2212	15	Ref		
	No	3609	31	0.85 (0.49-1.47)	0.88 (0.47-1.62)	0.677
	Not done/recorded	11779	94	1.08 (0.72-1.62)	1.29 (0.83-2.01)	0.264

Table 4: Risk factor analysis for late TB reactivation in adults among contacts of PTB patients in Victoria, Australia 2005-2015 (n=14,496 the number of contacts disease-free for the first six months of follow up).

Contact characteristics	**Category**	**Contact TB disease**		**Crude HR (95%CI)**	**Adjusted HR (95%CI)**	**P-value**
		No	Yes			
Sex	Female	7585	31	Ref	Ref	
	Male	6733	35	1.26 (0.77-2.06)	1.20 (0.74-1.96)	0.460
	Not stated	112	0	NA	NA	
Type of contact	Close contacts	5785	50	4.43 (2.49-7.86)	3.25(1.81-5.82)	<0.001*
	Other contacts	8645	16	Ref		
Country of birth	High burden country	4602	43	3.73 (2.23-6.24)	2.17 (1.26-3.75)	0.005*
	Low burden country	9246	23	Ref	Ref	
	Not stated	582	0	NA	NA	
TST	Negative	9629	16	Ref	Not included	
	Positive	2729	35	6.94 (3.78-2.74)*		
	Not performed	2072	15	5.06 (2.45-0.43)*		
BCG	Yes	6794	46	7.89 (2.43-25.65)	4.54 (1.35-15.28)	0.014*
	No	3374	3	Ref	Ref	
	Not stated/unknown	4262	17	4.48 (1.30-15.42)	3.52 (1.01-12.26)	0.048*
Index characteristics						
Sex	Female	6798	42	2.08 (1.14-3.82)	0.55 (0.29-1.03)	0.06

	Male	7632	24	Ref	Ref	
Country of birth	High TB burden	7672	36	Ref	Ref	
	Others	6758	30	1.39 (0.76-2.54)	1.43 (0.78-2.60)	0.24
Method of case finding	Clinical presentation	13533	64	Ref	Ref	
	Other methods	897	2	0.50 (0.11-2.27)	0.42 (0.09-1.89)	0.26
Type of TB	New case	13553	64	Ref	Ref	
	Relapse	704	2	0.72 (0.14-3.69)	0.82 (0.16-4.11)	0.81
	Unknown	173	0	NA	NA	
Chest X-ray	Abnormal, cavity	5285	25	0.92 (0.56-1.52)	0.62 (0.16-4.11)	0.17
	Abnormal, no cavity	8025	40	Ref	Ref	
	Abnormal, no details	185	0	NA	NA	
	Normal	291	0	NA	NA	
	Not done/stated	644	1	0.27 (0.04-1.99)	0.40 (0.04-3.89)	0.43

*=statistical significance of $p<0.05$, HR= hazard ratio, CI= confidence interval, Un/CT/Scr/Rtf = Undertaking/contact tracing/screening/routine follow up

4.6. Discussion

The aim of this study was to determine the profile and pattern of contacts' TB disease using an extensive prospectively collected contact investigation data set. We also quantified the risk of primary progression (development of any type of TB within the first six months) and late reactivation (development of any type of TB after six months). Our definition for primary progression was supported by previous findings that the risk of TB to be highest within the first three to nine months after infection [4, 5, 19]. Over more than seven years of median follow-up, we found a total of 1.3% of contacts were diagnosed with active TB disease. The majority, 62.5% of all contact TB disease, was seen in the first six months of the follow-up period. Almost all active TB among children (<5 years) was comprised of primary progression while late reactivation occurred predominantly in adults (≥15years). BCG

vaccination was found to be protective against primary progression, but it was associated with a higher rate of late TB reactivation among adult contacts. Our regression analyses identified risk factors significantly associated with both primary progression and late reactivation.

Our findings showed that a greater proportion of TB disease activation occurred among children aged under 15 years, which is in agreement with previous findings that children are at higher risk of TB disease than adults [20-22]. In addition, almost all active TB among children were a primary progression, which is also consistent with results from other studies [3, 4]. Our finding supports the recommendations of the WHO that contact investigation and provision of preventive treatment to children aged <5 years should be prioritised in resource-limited settings, because of their very high risk of active TB disease [23]. However, we also highlight that those older contacts, especially children aged 5-14 years, are at significant risk of primary progression, and so warrant a similar programmatic response.

We found that late reactivation among children aged 5-14 years was more common after they reached adolescence, or from the age of approximately 15 years. This finding is supported by a previous study which reported that when the first exposure occurred in children (aged 1-14 years), TB did not develop until after the age of 15 years [24]. Globally, there is a large burden of TB in adolescents and young adults of age 10-24 years with an estimate of 1.78 million young people with TB [25].

Consistent with previous studies, we found that BCG vaccination was protective for primary progression and in contacts aged 5-14 years (age at exposure). However, unexpectedly, the risk of late TB reactivation was greater among adults who were BCG vaccinated than unvaccinated ones. Previous studies have consistently reported that BCG is effective in preventing TB disease among children and that its effectiveness is reduced among adults [26-

28]. A previous systematic review and meta-analysis found that protection from BCG is higher in trials of school-age vaccination than neonatal vaccination, while there was no significant protection found in studies that recruited from all age groups [29]. We note similar findings from the largest ever trial estimating the efficacy of BCG, which reported non-significantly higher rates of TB among every age group of vaccinated adults than in the unvaccinated over 15 years of follow-up [30], and a study from the Netherlands that reported findings consistent with those described here [31]. One possible interpretation of this evidence is that BCG is effective in reducing TB-related child morbidity and mortality, but may increase late reactivation rates in adults. These phenomena may lead to perverse effects on the overall epidemic because the adult forms of TB are typically most infectious, and they are critical when considering BCG revaccination programs and interpreting recent trials of novel vaccines whose outcomes consider immunological responses to *Mtb* rather than incidence of TB disease [32]. Although we found no evidence of an association between BCG vaccination and the absence of preventive therapy, this or other unmeasured factors could confound these findings. Furthermore, adults with unknown BCG status were also associated with higher rates of late reactivation, such that ensuring a comprehensive collection of BCG status will be a future focus of our research.

Our study found that contacts who were born in high TB burden countries had a higher risk of TB than contacts born in low TB burden countries (including Australia). This may reflect that contacts born in high TB burden countries may have had additional TB exposures other than the index exposure we identified. An alternative reason for this association may be lower socioeconomic status and poorer general health; migrants from high-burden countries are likely to have poorer health status [33, 34].

Contacts identified as close contact by the VTP had a much higher risk of TB disease, providing reassurance around our ability to discern high-risk patients and an opportunity to

prioritise cases for future chemoprophylaxis. Although the development of active TB in contacts after the infection is mainly dependent on contacts' endogenous factors, we also found associations with index patients' characteristics. Specifically, index patients with recurrent TB appeared more infectious (i.e. their contacts are more likely to have primary progression) than index patients with new TB. This may be explained by patients with recurrent TB having more extended infectious periods compared to those with new TB disease. However, this index characteristic did not show an effect on contacts' risk of late reactivation. The association between shorter time to disease and greater index infectiousness may explain this, perhaps because the infectious dosage is a more important determinant of early progression whereas late reactivation is more dependent on host factors.

Our study has strengths arising from the fact that it is a population-based study from a centralised database. However, we had limited ability to determine any use of preventive therapy, and lack of accounting for it may have biased the results in unpredictable ways. Also, the transmission was inferred based on epidemiologic links due to the absence of genotyping data. This limitation may affect the late reactivation more than primary progression. Another possible limitation arises from the fact that we did not exclude co-prevalent cases in which the direction of transmission cannot be definitively determined. However, in such co-prevalent cases, the VTP determines the source of patients based on detailed epidemiologic investigation [12].

4.7. Conclusions

In conclusion, child contacts were at highest risk of early progression, but high rates of late reactivation among infected children as they reached adolescence imply that children should be strongly considered for preventive interventions even after their initial high-risk period has elapsed. BCG seems to have an important effect on TB epidemiology in preventing early

activation but increasing rates of late reactivation in adults. This effect requires further investigation with more comprehensive data on contacts' receipt of preventive therapy.

Authors' contribution

YAM and JMT conceptualised the study. YAM developed the protocol and, JMT reviewed it. ET and YAM performed data extraction and cleaning. YAM, ACC, ESM, JTD, RR and JMT undertook the analysis. YAM drafted the manuscript and all authors involved in consecutive revisions of the manuscript. All authors reviewed and approved the final manuscript.

Acknowledgement

The authors would like to thank Monash University for providing the PhD scholarship to YAM. We are also thankful for the Victorian Department of Health and Human Services for providing data.

4.8. References

1. WHO. GLOBAL TUBERCULOSIS REPORT 2017. Geneva; 2017.
2. Vynnycky E, Fine PE. Lifetime risks, incubation period, and serial interval of tuberculosis. *Am J Epidemiol* 2000: 152(3): 247-263.
3. Trauer JM, Moyo N, Tay E-L, Dale K, Ragonnet R, McBryde ES, Denholm JT. Risk of Active Tuberculosis in the Five Years Following Infection... 15%? *CHEST Journal* 2016: 149(2): 516-525.
4. Reichler MR, Khan A, Sterling TR, Zhao H, Moran J, McAuley J, Bessler P, Mangura B. Risk and Timing of Tuberculosis Among Close Contacts of Persons with Infectious Tuberculosis. *The Journal of infectious diseases* 2018: 40: 1-9.
5. Tornee S, Kaewkungwal J, Fungladda W, Silachamroon U, Akarasewi P, Sunakorn P. Risk factors for tuberculosis infection among household contacts in Bangkok, Thailand. *Southeast Asian J Trop Med Public Health* 2004: 35(2): 375-383.
6. Teixeira L, Perkins MD, Johnson JL, Keller R, Palaci M, do Valle Dettoni V, Canedo Rocha LM, Debanne S, Talbot E, Dietze R. Infection and disease among household contacts of patients with multidrug-resistant tuberculosis. *Int J Tuberc Lung Dis* 2001: 5(4): 321-328.
7. Melsew Y, Doan T, Gambhir M, Cheng A, McBryde E, Trauer J. Risk factors for infectiousness of patients with tuberculosis: a systematic review and meta-analysis. *Epidemiol Infect* 2018: 1-9.
8. Talarico S, Ijaz K, Zhang X, Mukasa LN, Zhang L, Marrs CF, Cave MD, Bates JH, Yang Z. Identification of factors for tuberculosis transmission via an integrated multidisciplinary approach. *Tuberculosis* 2011: 91(3): 244-249.

9. Herrero MB, Ramos S, Arrossi S. Determinants of non adherence to tuberculosis treatment in Argentina: barriers related to access to treatment. *Revista Brasileira de Epidemiologia* 2015: 18(2): 287-298.
10. Davies P, Yew W, Ganguly D, Davidow A, Reichman L, Dheda K, Rook G. Smoking and tuberculosis: the epidemiological association and immunopathogenesis. *Trans R Soc Trop Med Hyg* 2006: 100(4): 291-298.
11. DepartmentofHealth&HumanServices. Management, control and prevention of tuberculosis: Guidelines for health care providers. State of Victorias, Victoria, Australia, 2015.
12. Services DoHH. Mycobacterial infections (tuberculosis). 2017 [cited 2017 April 07]; Available from: https://www2.health.vic.gov.au/public-health/infectious-diseases/disease-information-advice/tuberculosis
13. Dale K, Tay E, Trevan P, Denholm J. Mortality among tuberculosis cases in Victoria, 2002–2013: case fatality and factors associated with death. *The International Journal of Tuberculosis and Lung Disease* 2016: 20(4): 515-523.
14. Statistics ABo. 2016 Census: Victoria. 2017 [cited 2018 26 July]; Available from: http://www.abs.gov.au/ausstats/abs@.nsf/MediaRealesesByCatalogue/C508DD213FD43EA7CA258148000C6BBE?OpenDocument
15. Moyo N, Tay E, Denholm J. Evaluation of tuberculin skin testing in tuberculosis contacts in Victoria, Australia, 2005–2013. *Public health action* 2015: 5(3): 188-193.
16. Health AGDo. Tuberculosis case definition. 2010 22 December 2010 [cited 2017 20 May]; Available from: http://www.health.gov.au/internet/main/publishing.nsf/Content/cda-surveil-nndss-casedefs-cd_tb.htm
17. WHO. Global Tuberculosis Report 2015. Geneva; 2015.
18. Benchimol EI, Smeeth L, Guttmann A, Harron K, Moher D, Petersen I, Sørensen HT, von Elm E, Langan SM, Committee RW. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) statement. *PLoS Med* 2015: 12(10): e1001885.
19. Behr MA, Edelstein PH, Ramakrishnan L. Revisiting the timetable of tuberculosis. *BMJ* 2018: 362: k2738.
20. Moran-Mendoza O, Marion S, Elwood K, Patrick D, FitzGerald J. Risk factors for developing tuberculosis: a 12-year follow-up of contacts of tuberculosis cases. *The International Journal of Tuberculosis and Lung Disease* 2010: 14(9): 1112-1119.
21. Saunders MJ, Wingfield T, Tovar MA, Baldwin MR, Datta S, Zevallos K, Montoya R, Valencia TR, Friedland JS, Moulton LH. A score to predict and stratify risk of tuberculosis in adult contacts of tuberculosis index cases: a prospective derivation and external validation cohort study. *The Lancet Infectious Diseases* 2017: 17(11): 1190-1199.
22. Erkens CG, Slump E, Verhagen M, Schimmel H, Cobelens F, van den Hof S. Risk of developing tuberculosis disease among persons diagnosed with latent tuberculosis infection in the Netherlands. *Eur Respir J* 2016: ERJ-01157-02016.
23. WHO. Recommendations for investigating contacts of persons with infectious tuberculosis in low- and middle-income countries. World Health Organization, Geneva, 2012.
24. Zeidberg LD, Gass R, DILLON A, Hutcheson R. The Williamson County Tuberculosis Study. A twenty-four-year epidemiologic study. *Am Rev Respir Dis* 1963: 87(3, Pt. 2): 1-88.
25. Snow KJ, Sismanidis C, Denholm J, Sawyer SM, Graham SM. The incidence of tuberculosis among adolescents and young adults: a global estimate. *Eur Respir J* 2018: 51(2): 1702352.
26. Rodrigues LC, Diwan VK, Wheeler JG. Protective effect of BCG against tuberculous meningitis and miliary tuberculosis: a meta-analysis. *Int J Epidemiol* 1993: 22(6): 1154-1158.

27. Roy A, Eisenhut M, Harris R, Rodrigues L, Sridhar S, Habermann S, Snell L, Mangtani P, Adetifa I, Lalvani A. Effect of BCG vaccination against Mycobacterium tuberculosis infection in children: systematic review and meta-analysis. *BMJ* 2014: 349: g4643.
28. Soysal A, Millington KA, Bakir M, Dosanjh D, Aslan Y, Deeks JJ, Efe S, Staveley I, Ewer K, Lalvani A. Effect of BCG vaccination on risk of Mycobacterium tuberculosis infection in children with household tuberculosis contact: a prospective community-based study. *The Lancet* 2005: 366(9495): 1443-1451.
29. Mangtani P, Abubakar I, Ariti C, Beynon R, Pimpin L, Fine PE, Rodrigues LC, Smith PG, Lipman M, Whiting PF. Protection by BCG vaccine against tuberculosis: a systematic review of randomized controlled trials. *Clin Infect Dis* 2013: 58(4): 470-480.
30. Radhakrishna S, Frieden TR, Subramani R. Influence of sex, age & nontuberculous infection at intake on the efficacy of BCG: re-analysis of 15-year data from a double-blind randomized control trial in south India. *Indian J Med Res* 2006: 123(2): 119.
31. Sloot R, Schim van der Loeff MF, Kouw PM, Borgdorff MW. Risk of tuberculosis after recent exposure. A 10-year follow-up study of contacts in Amsterdam. *Am J Respir Crit Care Med* 2014: 190(9): 1044-1052.
32. Nemes E, Geldenhuys H, Rozot V, Rutkowski KT, Ratangee F, Bilek N, Mabwe S, Makhethe L, Erasmus M, Toefy A. Prevention of M. tuberculosis infection with H4: IC31 vaccine or BCG revaccination. *N Engl J Med* 2018: 379(2): 138-149.
33. Bauer TK, Cobb‐Clark DA, Hildebrand VA, Sinning MG. A comparative analysis of the nativity wealth gap. *Econ Inq* 2011: 49(4): 989-1007.
34. Abouzeid M, Philpot B, Janus ED, Coates MJ, Dunbar JA. Type 2 diabetes prevalence varies by socio-economic status within and between migrant groups: analysis and implications for Australia. *BMC Public Health* 2013: 13(1): 252.

Chapter 5

5. Heterogeneous infectiousness in mathematical models of tuberculosis: a systematic review

5.1. Chapter overview

Infectious disease models are mathematical descriptions of the spread of infection that are simplified descriptions of a biological system and can be used to understand the system's behaviour or to inform interventions through predictions. These models should incorporate accurate disease transmission characteristics that arise from epidemiological evidence in their assumptions.

Having presented empirical evidence for the importance of heterogeneity in infectiousness and progression in previous chapters, this Chapter explores the methods employed to capture infectiousness heterogeneity in past transmission dynamic models of TB. By systematically reviewing past dynamic transmission models of TB, this Chapter highlights how previous TB modelling studies have approached heterogeneous infectiousness and which factors were used to stratify patients into various infectivity levels. Also, the Chapter evaluates whether these approaches are supported by epidemiological evidence. As described in the Chapter, the vast majority of TB models did not consider heterogeneous infectiousness at all, and of those that assumed heterogeneous infectiousness the approaches employed did not capture the heterogeneity that was found in previous chapters. The Chapter also provides a recommendation for a new data-driven transmission dynamic model of TB that can optimally capture heterogeneous infectiousness. This Chapter addresses thesis question 4.

Citation: Melsew YA, Adekunle AI, Cheng AC, McBryde ES, Ragonnet R, Trauer JM, Heterogeneous infectiousness in mathematical models of tuberculosis: a systematic review. *Epidemics*. 2019 Oct 17:100374.

Epidemics xxx (xxxx) xxxx

Contents lists available at ScienceDirect

Epidemics

journal homepage: www.elsevier.com/locate/epidemics

Heterogeneous infectiousness in mathematical models of tuberculosis: A systematic review

Yayehirad A. Melsew[a,c,*], Adeshina I. Adekunle[b], Allen C. Cheng[a], Emma S. McBryde[b], Romain Ragonnet[a], James M. Trauer[a]

[a] Department of Epidemiology and Preventive Medicine, School of Public Health and Preventive Medicine, Monash University, 553 St Kilda Road, Melbourne, VIC 3004, Australia
[b] Australian Institute of Tropical Health and Medicine, James Cook University, Townsville, QLD 4811, Australia
[c] Department of Epidemiology and Biostatistics, Institute of Public Health, University of Gondar, Gondar, Ethiopia

ARTICLE INFO

Keywords:
Mathematical modelling
Tuberculosis
Heterogeneous infectiousness
Systematic review

ABSTRACT

TB mathematical models employ various assumptions and approaches in dealing with the heterogeneous infectiousness of persons with active TB. We reviewed existing approaches and considered the relationship between them and existing epidemiological evidence.

We searched the following electronic bibliographic databases from inception to 9 October 2018: MEDLINE, EMBASE, Biosis, Global Health and Scopus. Two investigators extracted data using a standardised data extraction tool. We included in the review any transmission dynamic model of *M. tuberculosis* transmission explicitly simulating heterogeneous infectiousness of person with active TB. We extracted information including: study objective, model structure, number of active TB compartments, factors used to stratify the active TB compartment, relative infectiousness of each active TB compartment and any intervention evaluated in the model.

Our search returned 1899 unique references, of which the full text of 454 records were assessed for eligibility, and 99 studies met the inclusion criteria. Of these, 89 used compartmental models implemented with ordinary differential equations, while the most common approach to stratification of the active TB compartment was to incorporate two levels of infectiousness. However, various clinical characteristics were used to stratify the active TB compartments, and models differed as to whether they permitted transition between these states. Thirty-four models stratified the infectious compartment according to sputum smear status or pulmonary involvement, while 18 models stratified based on health care-related factors. Variation in infectiousness associated with drug-resistant *M. tuberculosis* was the rationale for stratifying active TB in 33 models, with these models consistently assuming that drug-resistant active TB cases were less infectious.

Given the evidence of extensive heterogeneity in infectiousness of individuals with active TB, an argument exists for incorporating heterogeneous infectiousness, although this should be considered in light of the objectives of the study and the research question.

PROSPERO Registration: CRD42019111936.

1. Introduction

Tuberculosis (TB) is one of the top ten causes of death and the leading cause from a single infectious agent (WHO, 2018). Although the disease burden caused by TB is falling globally, we are unlikely to see the first milestones of the *End TB Strategy* achieved in 2020 (WHO, 2014). The current global burden of TB is not homogeneously distributed across populations, but rather is an aggregate of localised micro-epidemics, with this heterogeneous distribution likely to become more prominent as disease burden decreases (Pai et al., 2016). One of several factors that contributes to the heterogeneity of *Mycobacterium tuberculosis (Mtb)* transmission in a given population is the intrinsic variation in the infectiousness of individuals with active TB (Trauer et al., 2018). The heterogeneous infectiousness of such individuals has

* Corresponding author at: Department of Epidemiology and Preventive Medicine, School of Public Health and Preventive Medicine, Monash University, 553 St Kilda Road, Melbourne, VIC 3004, Australia.
E-mail addresses: yayehirad.melsew@monash.edu (Y.A. Melsew), adeshina.adekunle@jcu.edu.au (A.I. Adekunle), allen.cheng@monash.edu (A.C. Cheng), emma.mcbryde@jcu.edu.au (E.S. McBryde), romain.ragonnet@monash.edu (R. Ragonnet), james.trauer@monash.edu (J.M. Trauer).

https://doi.org/10.1016/j.epidem.2019.100374
Received 6 August 2019; Received in revised form 9 October 2019; Accepted 13 October 2019

been established from both epidemiologic contact investigations and genotypic data (Ypma et al., 2013; Melsew et al., 2019). These studies have demonstrated that a very few highly infectious individuals with active TB are responsible for a large proportion of onward transmission, while a much larger proportion of individuals have very low or negligible infectiousness. As this heterogeneity has effects on the basic reproduction number, *Ro*, and the impact of interventions, incorporating heterogeneous infectiousness assumptions in TB transmission dynamic models may be critical depending on the modelling objectives (Trauer et al., 2018).

Transmission dynamic models typically have one of two broad purposes: either to improve understanding of the behaviour of the epidemic, or to make predictions of disease burden, including under counterfactual intervention strategy scenarios. Any such models should represent reality as accurately as possible, while also considering the need for model parsimony, although the optimal balance of these factors is dependent on judgement and epidemiological understanding (McLean, 2013; Cohen and White, 2016). Though the first TB model was published more than half a century ago (Waaler et al., 1962), TB models continue to differ in their assumptions due to imperfect understanding of the complex natural history of TB and lack of available data on the clinical progression of individuals through their stages of disease.

By systematically reviewing previous TB transmission modelling studies, we describe existing methods used to capture heterogeneity in infectiousness of individuals with active TB. Specifically, we aimed to identify all TB transmission models that explicitly stratified the active TB compartment by levels of infectiousness, and to understand their modelling assumptions and parameter choices.

2. Methods

We reviewed all published TB mathematical modelling studies that considered heterogeneity in infectiousness, with infectiousness conceptually defined as the number of secondary infections resulting from an individual with active disease per unit time. The review protocol was prospectively registered with PROSPERO (International Prospective Register of Systematic Reviews, registration number: CRD42019111936) and can be accessed at https://www.crd.york.ac.uk/prospero/display_record.php?RecordID = 111,936). We adhered to the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) guidelines throughout (Liberati et al., 2009) and the PRISMA checklist is provided in Appendix A in Supplementary material.

2.1. Search strategy

Before starting the review, we searched for systematic reviews that assessed methods used to capture heterogeneity in infectiousness of individuals with active TB in transmission dynamic models, with no such reviews found.

We searched the following electronic bibliographic databases from inception to 9 October 2018: MEDLINE, EMBASE, Biosis, Global Health and Scopus for studies in human subjects published in English. All search terms were “exploded” to capture all resources and consisted of “TB”, “Tuberculosis”, “Mycobacterium tuberculosis”, “Dynamic model”, “Transmission dynamics”, “Simulation”. Full search strategy is provided in Appendix B in Supplementary material.

2.2. Study selection

Search results were exported to EndNoteX8.2 (Clarivate Analytics, NY, USA) and duplicates were removed. Titles and abstracts were screened to identify potentially relevant articles. After initial screening for any TB modelling studies, the full texts of eligible articles were collected and assessed for eligibility. Included studies consisted of any transmission dynamic TB modelling study explicitly incorporating the assumption of heterogeneous infectiousness of individuals with active TB. Studies were excluded if they were systematic reviews, did not employ mathematical models, did not model *Mtb*, were intra-host models or assumed homogeneous infectiousness.

2.3. Assessment of quality

As we aimed to assess the diverse methods used to capture heterogeneous infectiousness of individuals with active TB and there was no epidemiological pooling, risk of bias assessment was not performed.

2.4. Data extraction and analysis

Data were extracted using a standardised data extraction tool developed, tested and approved by four investigators (YAM, AIA, RR and JMT). Full manuscripts and supplementary material accompanying the main texts of each article were reviewed. YAM and AIA extracted data and any controversies in interpretation were resolved by consensus.

We extracted general information that included: year of publication, study setting, study objective, model structure, and information about our main objective, including: the number of infectious compartments, factors used to stratify the infectious compartment and the relative infectiousness of each active compartment.

3. Results

Our search strategy returned 1899 unique references, of which the full text of 454 records were assessed for eligibility, and 99 studies met the inclusion criteria of explicitly capturing heterogeneous infectiousness (Fig. 1). Of the 99 included TB models, 89 were compartmental models implemented using ordinary differential equations, one study used both ordinary and partial differential equations and the remaining nine were individual-based models (IBM). Included studies spanned many different epidemiological settings: 25 studies were from Asia or the Asia-Pacific, 21 were from sub-Saharan Africa, 12 studies were from Europe or North America. Nineteen studies represented hypothetical settings based on TB burden, economic development or HIV prevalence. Eight of these models were from high TB-burden settings (or “low/middle income”), two were from low TB burden settings (or “the developed world”), two were from high HIV prevalence settings and one aimed to represent global TB epidemiology.

3.1. Aims of models

The broad objectives of most included TB models were to evaluate the impact of various intervention strategies and make predictions of future disease burden based on various settings and scenarios. Forty-four of the 99 models were designed to estimate the likely impact of currently available interventions (Tuite et al., 2017; Vynnycky et al., 2015; Trauer et al., 2016a; Okuonghae and Aihie, 2008; Okuonghae and Ikhimwin, 2016; Okuonghae and Korobeinikov, 2007; Okuonghae and Omosigho, 2011; Rayhan and Bakhtiar, 2017; Sachdeva et al., 2015; Salje et al., 2014; Salomon et al., 2006; Sharomi et al., 2008; Suen et al., 2014; Thomas et al., 2010; Hickson et al., 2011; Hickson et al., 2012; Hill et al., 2012; Hughes et al., 2006; Huynh et al., 2015; Jung et al., 2002; Kendall et al., 2015; Knight et al., 2015; Liao and Lin, 2012; Mellor et al., 2011; Menzies et al., 2012; Moualeu et al., 2015; Bowong and Alaoui, 2013; Dowdy et al., 2014; Dowdy et al., 2013a; Dowdy and Chaisson, 2009; Dowdy et al., 2008; Dowdy et al., 2006; Dowdy et al., 2013b; Dowdy et al., 2013c; Dye and Williams, 2008; Espindola et al., 2012; Fofana et al., 2017; Garcia et al., 1997; Gomes et al., 2007; Guzzetta et al., 2011; Bacaer et al., 2008; Basu et al., 2007; Basu et al., 2009; Bhunu and Garira, 2009), while five models evaluated the potential impact of interventions that were new, under development or hypothetical (Abu-Raddad et al., 2009; Knight et al., 2014; Lin

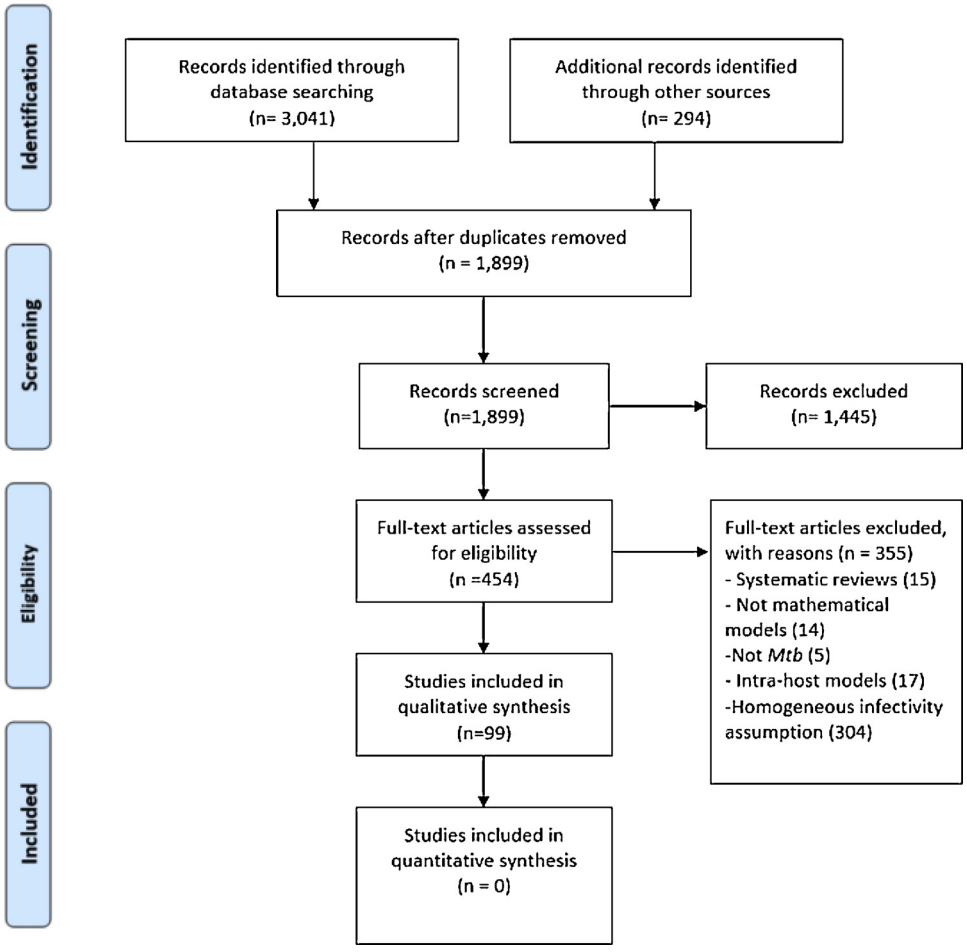


Fig. 1. Flow diagram of study selection process.

et al., 2012; Sun et al., 2013; Trauer et al., 2016b). Twenty-one models were built to capture TB epidemiology in a specific setting and make predictions by using local data (Arregui et al., 2018; Blower et al., 1995; Herrera et al., 2013; Houben et al., 2016; Legrand et al., 2008; Liao et al., 2012; Liao et al., 2013; Liu et al., 2011; Melnichenko and Romanyukha, 2009; Menzies et al., 2018; Moualeu et al., 2014; Moualeu-Ngangue et al., 2015; Mushayabasa and Bhunu, 2013; Nishiura et al., 2004; Oxlade et al., 2011; Pandey et al., 2017; Perelman et al., 2004; Sun et al., 2011; Wu et al., 2010; Zhang et al., 2015). Eight models were more theoretically focused, aiming to draw general conclusions about transmission dynamics, such as determining equilibrium points or reproductive numbers, and in some cases undertaking stability or sensitivity analyses around these quantities (Houben et al., 2016; Apriliani et al., 2016; Blower and Gerberding, 1998; Cohen and Murray, 2004; Liu et al., 2008; Liu and Sun, 2010; McBryde et al., 2017; Trauer et al., 2014). Four models assessed the impact of risk factors, such as smoking, age and diabetes mellitus (Garcia et al., 1997; Bhunu et al., 2011; Moualeu et al., 2012; Rodrigues et al., 2015), three evaluated the impact of immigration (Korthals Altes et al., 2018; Moualeu et al., 2018; Okuonghae and Aihie, 2010), and two evaluated the impact of *Mtb* re-infection (Moualeu et al., 2016; Sharomi et al., 2017). Six models additionally assessed aspects of the MDR-TB epidemic and its control (Kendall et al., 2015; Agusto et al., 2015; Ahmadin and Fatmawati, 2014; Bishai et al., 2010; Raimundo et al., 2014; Shrestha et al., 2014).

Models stratified the active TB compartment according to the level of infectiousness for a range of reasons, the commonest being to capture the impact of interventions directed at specific sub-groups of individuals with active TB or that had differential effectiveness depending on the level of infectiousness. Within this group, forty-three models stratified the active TB compartment in order to capture the impact of detection and treatment interventions (Tuite et al., 2017; Vynnycky et al., 2015; Okuonghae and Korobeinikov, 2007; Rayhan and Bakhtiar, 2017; Sachdeva et al., 2015; Salje et al., 2014; Salomon et al., 2006; Sharomi et al., 2008; Suen et al., 2014; Thomas et al., 2010; Huynh et al., 2015; Jung et al., 2002; Knight et al., 2015; Mellor et al., 2011; Bowong and Alaoui, 2013; Dowdy et al., 2014; Dowdy et al., 2013a; Dowdy and Chaisson, 2009; Dowdy et al., 2008; Dowdy et al., 2006; Dowdy et al., 2013b; Dowdy et al., 2013c; Fofana et al., 2017; Garcia et al., 1997; Basu et al., 2007; Basu et al., 2009; Bhunu and Garira, 2009; Abu-Raddad et al., 2009; Lin et al., 2012; Sun et al., 2013; Trauer et al., 2016b; Legrand et al., 2008; Nishiura et al., 2004; Perelman et al., 2004; Blower and Gerberding, 1998; Liu et al., 2008; Liu and Sun, 2010; Korthals Altes et al., 2018; Okuonghae and Aihie, 2010; Agusto et al., 2015; Huo and Zou, 2016; Kendall et al., 2017; Osgood et al., 2011). By contrast, forty-four models incorporated stratification in the process of estimating key parameters in TB epidemiology, such as the number of secondary infections per infectious individual and *Ro* (Trauer et al., 2016a; Okuonghae and Aihie, 2008; Okuonghae and Ikhimwin, 2016; Okuonghae and Omosigho, 2011; Hickson et al., 2011; Hickson et al., 2012; Hughes et al., 2006; Liao and Lin, 2012; Menzies et al., 2012; Moualeu et al., 2015; Gomes et al., 2007; Guzzetta et al., 2011; Bacaer et al., 2008; Blower et al., 1995; Herrera et al., 2013; Liao et al., 2012; Liao et al., 2013; Melnichenko and Romanyukha, 2009; Menzies et al., 2018; Moualeu et al., 2014; Moualeu-Ngangue et al., 2015; Mushayabasa and Bhunu, 2013; Oxlade et al., 2011; Pandey et al., 2017; Zhang et al., 2015; Apriliani et al., 2016; Cohen and Murray, 2004; McBryde et al., 2017; Trauer et al., 2014; Bhunu et al.,

2011; Moualeu et al., 2012; Moualeu et al., 2018; Moualeu et al., 2016; Sharomi et al., 2017; Ahmadin and Fatmawati, 2014; Raimundo et al., 2014; Aparicio and Castillo-Chavez, 2009; Kasaie and Dowdy, 2013; Li et al., 2011; Nyabadza and Kgosimore, 2012; Okuonghae, 2013). Eight models also stratified the active TB compartment to predict the future proportion of each type of TB simulated (Kendall et al., 2015; Espindola et al., 2012; Arregui et al., 2018; Sun et al., 2011; Wu et al., 2010; Rodrigues et al., 2015; Bishai et al., 2010; Shrestha et al., 2014). In a further six models, the rationale for stratifying the active TB compartment was not explained (Hill et al., 2012; Dye and Williams, 2008; Knight et al., 2014; Trauer et al., 2016b; Houben et al., 2016; Ackley et al., 2015).

3.2. Model structures

Models most commonly presented the natural history of TB using Susceptible, Exposed, Infectious, Recovered (SEIR) compartmental structures, although SEI, SEIS and SEIE structures were also used. (Some of these models employed more than one latency compartment; thus, "E" may represent multiple latency compartments.) In our review, the majority (64) of the models adopted a SEIR compartmental structure, while 12 used SEI structures and 11 employed SEIS.

The structures used by TB models to stratify the active TB compartment according to infectiousness level varied between models, with most models employing two active TB compartments. Fig. 2 presents a summary of the compartmental structures used and Table 1 maps the specific models to the compartmental structure employed. Henceforth, we refer to the specific model structures by the alphabetical structure names introduced in Fig. 2. It is important to note that several of these classifications include models that use the same compartmental structure, but use these structures to represent different factors (and so employ different parameter values). Stratifying the active TB compartment into two unconnected infectiousness levels (*Structure A*) was the commonest structure used, followed by a structure employing two infectious levels and with an additional transition process linking these two states (*Structure B*).

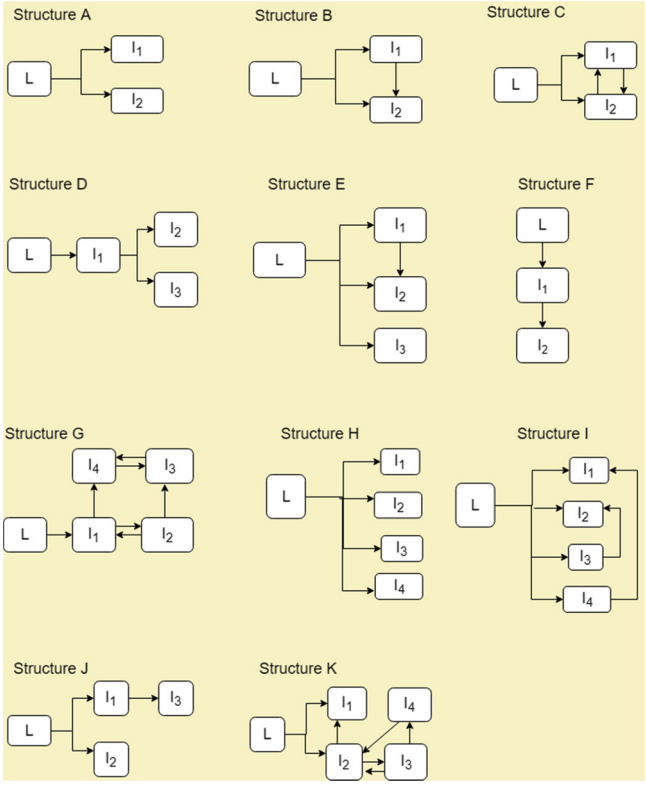


Fig. 2. Structures for the infectious compartments used by TB models that incorporate multiple levels of infectiousness. L: latent infection, I: active TB, with subscripts to indicate the multiple infectious compartments. Some of these models used more than one compartment for latency, thus "L" may represent multiple latency compartments here. The subscript numbers to the "I" compartments are arbitrary.

3.3. Factors for stratification

TB models stratified the active TB compartment by infectiousness in order to capture a range of clinical characteristics. These factors included: factors related to characteristics of the host, such as disease manifestation, co-morbidities and age; factors related to the organism, i.e. impaired fitness of *Mtb* due to drug resistance mutations; and factors related to the health care system. The factors used for stratification of the active TB compartment were sometimes combined in the included studies and are not mutually exclusive.

3.3.1. Factors related to disease characteristics

3.3.1.1. Pulmonary involvement. Six TB models explicitly classified the active TB compartment based on the clinical site of TB disease, as either representing cases with lung involvement (pulmonary TB), or with disease limited only to body organs other than the lungs (extrapulmonary TB) (Thomas et al., 2010; Hickson et al., 2012; Dye and Williams, 2008; Korthals Altes et al., 2018; Aparicio and Castillo-Chavez, 2009). These models universally assumed that only pulmonary TB cases were infectious without further stratification by smear-status. Five of these models employed *Structure A* (Thomas et al., 2010; Hickson et al., 2012; Mellor et al., 2011; Dye and Williams, 2008; Korthals Altes et al., 2018; Aparicio and Castillo-Chavez, 2009), while one model also incorporated conversion from non-infectious to infectious TB (*Structure B*) (Mellor et al., 2011).

3.3.1.2. Sputum smear-status. Sputum smear-status was a commonly used factor for stratifying active TB cases into varying levels of infectiousness. Twenty-two TB models stratified the active TB compartment by sputum smear-status (Tuite et al., 2017; Vynnycky et al., 2015; Menzies et al., 2012; Dowdy et al., 2014; Dowdy and Chaisson, 2009; Dowdy et al., 2008; Dowdy et al., 2006; Dowdy et al., 2013b; Dowdy et al., 2013c; Garcia et al., 1997; Gomes et al., 2007; Guzzetta et al., 2011; Abu-Raddad et al., 2009; Knight et al., 2014; Lin et al., 2012; Sun et al., 2013; Arregui et al., 2018; Houben et al., 2016; Menzies et al., 2018; Oxlade et al., 2011; Pandey et al., 2017; Ackley et al., 2015), of which eleven characterised active TB as either smear-positive or smear-negative TB (with smear-negative TB always considered less infectious). Among these models, seven employed *Structure A* (Dowdy et al., 2013b; Gomes et al., 2007; Knight et al., 2014; Lin et al., 2012; Oxlade et al., 2011; Pandey et al., 2017; Ackley et al., 2015), while the remaining used *Structure B* to incorporate a transition from smear-negative to smear-positive (Tuite et al., 2017; Menzies et al., 2012; Garcia et al., 1997; Houben et al., 2016; Menzies et al., 2018), thereby assuming that some smear-positive individuals must have initially been smear-negative at disease onset.

Employing *Structure E*, TB models used a three-tier stratification incorporating pulmonary involvement and smear-status to divide the active TB compartment into smear-positive TB, smear-negative TB and extrapulmonary TB (non-infectious) (Dowdy et al., 2014; Dowdy et al., 2013c; Abu-Raddad et al., 2009; Arregui et al., 2018). Five models included four active TB compartments, four of which used detection status to further cross-stratify smear-positive and TB smear-negative, while one model cross-stratified HIV status and smear-status (Vynnycky et al., 2015; Dowdy and Chaisson, 2009; Dowdy et al., 2008; Dowdy et al., 2006; Sun et al., 2013). Under each of these approaches, infectiousness was influenced by smear status, such that smear-positive individuals were considered more infectious than those smear-negative.

Twenty-one of the 22 models that stratified the active TB

Table 1
Infectious compartment structures used by compartmental TB single strain models.

Structure	Number of studies	Citation
A	27	(Okuonghae and Aihie, 2008; Okuonghae and Ikhimwin, 2016; Okuonghae and Korobeinikov, 2007; Okuonghae and Omosigho, 2011; Sharomi et al., 2008; Hill et al., 2012; Bowong and Alaoui, 2013; Dowdy et al., 2006; Dowdy et al., 2013b; Dowdy et al., 2013c; Dye and Williams, 2008; Gomes et al., 2007; Lin et al., 2012; Blower et al., 1995; Herrera et al., 2013; Liao et al., 2012; Liao et al., 2013; Oxlade et al., 2011; Pandey et al., 2017; Zhang et al., 2015; Apriliani et al., 2016; Korthals Altes et al., 2018; Okuonghae and Aihie, 2010; Sharomi et al., 2017; Aparicio and Castillo-Chavez, 2009; Okuonghae, 2013; Ackley et al., 2015)
B	11	(Hughes et al., 2006; Menzies et al., 2012; Garcia et al., 1997; Bacaer et al., 2008; Knight et al., 2014; Houben et al., 2016; Menzies et al., 2018; Moualeu et al., 2014; Bhunu et al., 2011; Moualeu et al., 2012; Moualeu et al., 2018)
C	1	(Huo and Zou, 2016)
D	2	(Dowdy et al., 2014; Dowdy and Chaisson, 2009)
E	2	(Abu-Raddad et al., 2009; Arregui et al., 2018)
F	2	(Dowdy et al., 2013a; Mushayabasa and Bhunu, 2013)
G	3	(Vynnycky et al., 2015; Melnichenko and Romanyukha, 2009; Perelman et al., 2004)
H	1	(Wu et al., 2010)
I	1	(Legrand et al., 2008)
J	6	(Hickson et al., 2011; Hickson et al., 2012; Moualeu et al., 2015; Moualeu-Ngangue et al., 2015; Liu et al., 2008; Moualeu et al., 2016)
K	1	(Osgood et al., 2011)
Others	5	(Salomon et al., 2006; Thomas et al., 2010; Dowdy et al., 2008; Sun et al., 2013; Moualeu et al., 2012)

Table 2
Proportions and relative infectiousness of active TB compartments among models that use two levels of infectiousness based on lungs involvement or smear status. Reference group.

Citation	Active TB compartment		Proportion progressing to high	Relative infectious of low	Conversion rate (low to high), per year
	High infectiousness	Low infectiousness			
(Ackley et al., 2015)	Infectious	Non-infectious	0.85	0	NA
(Aparicio and Castillo-Chavez, 2009)	Pulmonary	Extrapulmonary	0.7	0	NA
(Blower et al., 1995)	Infectious	Non-infectious	0.5-0.85	0	NA
(Dye and Williams, 2008)	Infectious	Non-infectious	0.6	0	NA
(Gomes et al., 2007)	Infectious	Non-infectious	0.66-0.87	0	NA
(Herrera et al., 2013)	Infectious	Non-infectious	0.7-0.85		NA
(Liao et al., 2012)	Infectious	Non-infectious	Normal distribution (0.78, 0.11)	0	NA
(Liao et al., 2013)	Infectious	Non-infectious	Normal distribution (0.78, 0.11)	0	NA
(Thomas et al., 2010)	Infectious	Non-infectious	HIV+ = 0.87 HIV- = 0.57	0	NA
(Hughes et al., 2006)	Infectious	Non-infectious	0.64	0	0.015
(Knight et al., 2014)	Infectious	Non-infectious	age < 15 = 0.1 age ≥15 = 0.5	0	0.015
(Dowdy et al., 2006)[a]	Smear-positive	Smear-negative	HIV- = 0.45 HIV+ = 0.35	0.22	NA
(Dowdy et al., 2013b)[#]	Smear-positive	Smear-negative	HIV- = 0.65 HIV+ = 0.5	0.15	NA
(Lin et al., 2012)[#]	Smear-positive	Smear-negative	HIV- adult = 0.7, HIV + adult = 0.48 HIV- children = 0.17 HIV + children = 0.1	0.22	NA
(Oxlade et al., 2011)	Smear-positive	Smear-negative	0.5	0.2	NA
(Pandey et al., 2017)	Smear-positive	Smear-negative	China, Korea, Philippines = 0.45 India = 0.63	0.2	NA
(Garcia et al., 1997)	Open cases	Non-open cases	0.5	0	to non-open[b]
(Houben et al., 2016)	Smear-positive	Smear-negative	0.85	0.22	0.015
(Menzies et al., 2012)	Smear-positive	Smear-negative	0.62	0.17-0.3	0.012-0.019

[a] Although these models include multiple stratifications which influenced progression proportions, infectiousness was only dependent on smear status.
[b] Infectious cases convert to non-infectious cases at rate of diagnosis and treatment.

compartment by smear-status considered both smear-positives and smear-negatives to be infectious, with smear-negatives having a lower level of infectiousness. The parameterisation for the relative infectiousness of smear-negative individuals compared to smear-positive was highly consistent across multiple modelling studies and almost universally fell in the range of 15–25%. Only one model that assessed the impact of changes in TB programs in low prevalence setting was an exception and assumed smear-negatives to be non-infectious (Garcia et al., 1997). Table 2 presents proportions and relative infectiousness of active TB compartments among models that incorporated two levels of infectiousness based on pulmonary involvement or smear status.

3.3.1.3. Type of infection. A model employing *Structure A* stratified active TB based on history of previous infection, as symptomatic primary infection or symptomatic re-infection, with the assumption that re-infected individuals were twice as infectious as primarily infected counterparts (Sharomi et al., 2017). An IBM further subdivided smear-positive and smear-negative TB into three groups based on the timing of disease onset from infection; namely primary TB, endogenous reactivation, and exogenous re-infection. However, under this approach the level of infectiousness was dependent only on smear-status, with smear-positive individuals four-times as infectious as smear-negatives (Guzzetta et al., 2011).

3.3.1.4. Stage of disease. A theoretical model of heterogeneous progression sub-divided the active TB compartment according to

progression time from infection to active TB, with fast progressors more infectious than their slowly progressing counterparts (Okuonghae, 2013). A model evaluating case finding strategies employed modified form of *Structure F* to stratify active TB into subclinical, pre-diagnostic and clinical phases (Dowdy et al., 2013a), with infectiousness increasing with progression between each of these three sequential stages. Similarly, an IBM simulated infectiousness using a linear function of time, with infectiousness increasing from zero to maximum infectiousness over the first nine months of the disease episode, and remaining unchanged thereafter (Kasaie and Dowdy, 2013).

3.3.2. Co-morbidities

Some models considered HIV co-infection and comorbid diabetes mellitus to affect the level of infectiousness of individuals with active TB. One model employed *Structure B* to stratify active TB based on HIV status, assuming that individuals co-infected with HIV were less infectious than HIV-negative (Bacaer et al., 2008). Using the same structure, another model sub-divided the active TB compartment based on diabetes status, assuming that individuals co-morbid with diabetic were more infectious than their non-diabetic counterparts (Moualeu et al., 2012). Smoking status was used in one model that assessed the impact of smoking on TB transmission dynamics, with smokers assumed to be more infectious (Bhunu et al., 2011).

3.3.3. Age

Twenty-three models incorporated age-specific infectiousness (Tuite et al., 2017; Vynnycky et al., 2015; Suen et al., 2014; Hughes et al., 2006; Huynh et al., 2015; Knight et al., 2015; Mellor et al., 2011; Menzies et al., 2012; Dye and Williams, 2008; Garcia et al., 1997; Guzzetta et al., 2011; Abu-Raddad et al., 2009; Knight et al., 2014; Arregui et al., 2018; Houben et al., 2016; Liao et al., 2012; Liao et al., 2013; Menzies et al., 2018; Wu et al., 2010; Rodrigues et al., 2015; Osgood et al., 2011; Aparicio and Castillo-Chavez, 2009; Nyabadza and Kgosimore, 2012). The great majority of these models assumed children to be markedly less infectious than adults, although there were considerable differences between models in the age categories chosen and parameterisation.

Twenty-one of these models incorporated other factors in addition to age, while the remaining two considered age as the only factor for infectious heterogeneity. Of the models that considered age as the only factor determining heterogeneous infectiousness, one stratified the active TB compartment into adults and children, assuming that only adults were infectious (Nyabadza and Kgosimore, 2012). The other model of TB epidemics in a Chinese city employed *Structure H* taking active TB cases of age 24 years as the reference group and assuming that children were 81% less infectious relative to this group, while younger adults were seven-times more infectious (Wu et al. (2010)).

3.3.4. Unspecified factors

Of all included models, eight classified the active TB compartment as either infectious or non-infectious, but the rationale for this classification was often not explained (Hill et al., 2012; Hughes et al., 2006; Knight et al., 2014; Blower et al., 1995; Herrera et al., 2013; Liao et al., 2012; Liao et al., 2013; Apriliani et al., 2016). Among these models, two (Hughes et al., 2006; Knight et al., 2014) employed *Structure B*, while six (Hill et al., 2012; Blower et al., 1995; Herrera et al., 2013; Liao et al., 2012; Liao et al., 2013; Apriliani et al., 2016) did not allow for progression between the active TB compartments (*Structure A*).

3.3.5. Drug resistance

Variation in infectiousness associated with drug-resistant *Mtb* strains was used to stratify active TB in 32 models (Trauer et al., 2016a; Rayhan and Bakhtiar, 2017; Sachdeva et al., 2015; Salje et al., 2014; Suen et al., 2014; Jung et al., 2002; Kendall et al., 2015; Liao and Lin, 2012; Espindola et al., 2012; Fofana et al., 2017; Basu et al., 2007; Basu et al., 2009; Bhunu and Garira, 2009; Trauer et al., 2016b; Liu et al., 2011; Nishiura et al., 2004; Sun et al., 2011; Blower and Gerberding, 1998; Cohen and Murray, 2004; Liu and Sun, 2010; McBryde et al., 2017; Trauer et al., 2014; Rodrigues et al., 2015; Agusto et al., 2015; Ahmadin and Fatmawati, 2014; Bishai et al., 2010; Raimundo et al., 2014; Shrestha et al., 2014; Kendall et al., 2017; Li et al., 2011; De Espíndola et al., 2011). In 31 of these models, cases with multidrug-resistant TB (MDR-TB) were assumed less infectious than cases with drug-susceptible TB (DS-TB), due to a fitness cost associated with the resistance-conferring mutation. Most of these models considered infectiousness of individuals with MDR-TB relative to individuals with DS-TB to be around 0.8.

Eighteen of the 32 TB models used two active TB compartments to represent DS-TB and MDR-TB without any additional stratification (Rayhan and Bakhtiar, 2017; Suen et al., 2014; Jung et al., 2002; Liao and Lin, 2012; Espindola et al., 2012; Fofana et al., 2017; Bhunu and Garira, 2009; Liu et al., 2011; Nishiura et al., 2004; Blower and Gerberding, 1998; Liu and Sun, 2010; Rodrigues et al., 2015; Agusto et al., 2015; Ahmadin and Fatmawati, 2014; Bishai et al., 2010; Raimundo et al., 2014; Shrestha et al., 2014; Li et al., 2011; De Espíndola et al., 2011). Out of the three models that stratified active TB into three compartments based on drug-susceptibility, with categories being DS-TB, MDR-TB and extensively drug-resistant TB (XDR-TB) and the assumption of a progressive gradient towards more resistant strains being less infectious (Agusto et al., 2015). The second such model incorporated detection status to stratify active TB compartment into three as undetected infectious TB, detected DS-TB and detected MDR-TB such that the undetected cases were the most infectious followed by detected MDR-TB. (Sun et al., 2011). The third model simulated active TB as either: on ineffective treatment (a treatment course without curative potential), on inadequate treatment (a treatment course with curative potential but taken for an insufficient duration) and not on treatment. This model assumed that individuals on ineffective or insufficient treatment were less infectious than individuals not on treatment (Fofana et al., 2017).

Six models integrated drug susceptibility with other factors, such as stage of disease, detection and treatment status, to stratify active TB into four compartments with varying levels of infectiousness (Trauer et al., 2016a; Basu et al., 2007; Basu et al., 2009; McBryde et al., 2017; Trauer et al., 2014; Kendall et al., 2017). For instance, a model of TB dynamics in South Africa stratified TB as either infectious or non-infectious and further sub-divided by drug susceptibility, assuming individuals with MDR-TB to be less infectious due to a resistance-associated fitness cost (Basu et al., 2007). A model evaluating the impact of diagnosis and treatment further stratified DS-TB and MDR-TB into early preclinical TB and symptomatic TB, to capture the variation in infectiousness associated with stage of disease in addition to a resistance-associated fitness cost (Kendall et al., 2017).

Further stratification of active TB into six compartments was undertaken in two models (Kendall et al., 2015; Cohen and Murray, 2004). A theoretical model of MDR-TB fitness characterised active TB compartment as detected or undetected, and furthered stratified as DS-TB, unfit MDR-TB or fit MDR-TB, assuming unfit MDR-TB to be the least infectious followed by fit MDR-TB and DS-TB (Cohen and Murray, 2004). The second model sub-divided both DS-TB and MDR-TB into early stage, active TB and TB on treatment (Kendall et al., 2015). This model assumed infectiousness variation by stage with early-stage TB the least infectious followed by TB on treatment. Further stratification of active TB into nine (Trauer et al., 2016b), 24 (Salje et al., 2014) and 28 (Sachdeva et al., 2015) compartments was used in models that integrated detection, treatment, HIV and smear-statuses with drug susceptibility status.

3.3.6. Health system-related factors

3.3.6.1. Detection and treatment. Nineteen TB models stratified the infectious compartment by diagnosis status (Okuonghae and Aihie,

2008; Okuonghae and Ikhimwin, 2016; Okuonghae and Korobeinikov, 2007; Okuonghae and Omosigho, 2011; Salomon et al., 2006; Sharomi et al., 2008; Hickson et al., 2011; Moualeu et al., 2015; Bowong and Alaoui, 2013; Legrand et al., 2008; Melnichenko and Romanyukha, 2009; Moualeu et al., 2014; Moualeu-Ngangue et al., 2015; Mushayabasa and Bhunu, 2013; Perelman et al., 2004; Liu et al., 2008; Moualeu et al., 2018; Okuonghae and Aihie, 2010; Moualeu et al., 2016). These models employed different assumptions regarding the greater infectiousness of undetected cases compared to those receiving treatment. Ten of these models sub-divided the active TB compartment into two compartments, detected and undetected, employing *Structure A* or *Structure B* (Okuonghae and Aihie, 2008; Okuonghae and Ikhimwin, 2016; Okuonghae and Korobeinikov, 2007; Okuonghae and Omosigho, 2011; Sharomi et al., 2008; Bowong and Alaoui, 2013; Moualeu et al., 2014; Mushayabasa and Bhunu, 2013; Moualeu et al., 2018; Okuonghae and Aihie, 2010). Further stratification into three compartments as provided by *Structure J* was employed in four models, three of which stratified active TB as diagnosed, lost to follow-up and undiagnosed infectious, with the assumption that undiagnosed individuals were most infectious followed by lost to follow-up (Moualeu et al., 2015; Moualeu-Ngangue et al., 2015; Moualeu et al., 2016). A theoretical model of global TB dynamics with the same compartmental structure stratified active TB as infectious treated, infectious untreated and non-infectious, with the assumption that untreated cases were seven times more infectious than patients under treatment (Liu et al., 2008).

A further four models implemented *Structure G* to divide active TB into four compartments based on detection status and other factors (Hickson et al., 2011; Legrand et al., 2008; Melnichenko and Romanyukha, 2009; Perelman et al., 2004). A model of HIV/TB co-infection included nine compartments based on detectability (a function of geographical location, local health services, and case-finding effort, as well as health seeking behaviour and intensity of symptoms), smear and treatment status, with the assumption that smear-positives were the most infectious and patients treated in high-quality treatment programs are non-infectious (Salomon et al., 2006).

3.3.6.2. Place of treatment. A model evaluating alternative TB treatment strategies in China used the place of treatment as a stratification factor and employed *Structure C* with the assumption that patients treated at home were more infectious than patients treated in hospitals (Huo and Zou, 2016). Similarly, a model employing *Structure A* assumed only hospitalised patients received treatment, thereby implying a higher level of infectiousness for non-hospitalised individuals with active TB (Zhang et al., 2015).

4. Discussion

Although the infectiousness of individuals with TB is known to be profoundly heterogeneous, with the epidemic significantly contributed to by a small group of highly infectious, the great majority of published TB models did not incorporate this phenomenon in any way. The infectiousness of individuals with active TB depends on their ability to release aerosolised droplets with sufficient bacilli concentrations (Turner et al., 2017). This process is particularly heterogeneous in the case of TB, due to associated risk factors such as the anatomical site of TB disease, demographic characteristics, behavioural characteristics and other co-morbidities (Melsew et al., 2018a). Moreover, instances of very extensive transmission of *Mtb* have been reported from case studies (Curtis et al., 1999; Valway et al., 1998), genotypic data analyses (Ypma et al., 2013) and contact investigation offspring data analyses (Melsew et al., 2019).

Although heterogeneous infectiousness is the reality for *Mtb* transmission, this does not necessarily imply that one must always stratify the active TB compartment irrespective of the modelling objectives. The principle of parsimony dictates the need to justify any additional complexity added to models through stratification. In some cases, the need for stratification is obvious, for example, if the objective of the model is to predict the relative magnitude of specific types of TB, such as pulmonary, extrapulmonary, drug-susceptible or multidrug-resistant. Similarly, if the objective is to evaluate the efficacy of public health interventions such as targeting children, HIV-positive persons, or patients with drug-resistant *Mtb* strains, modellers should consider the impact of heterogeneity. Similarly, models evaluating diagnostic interventions that have differential sensitivity based on factors associated with infectiousness, such as smear-status, should usually consider stratification. It is known that even traditional smear-based case finding interventions incorporate some targeting towards highly infectious patients because of the greater sensitivity of smear-based diagnostics for such patients (Cudahy and Shenoi, 2016). Active case finding interventions that employ very sensitive tools, and so may also detect less infectious TB patients, should consider the impact of heterogeneous infectiousness before claiming transmission reduction on the basis of the number of cases detected.

Most of the assumptions made by models that have incorporated heterogeneous infectiousness were broadly consistent with available epidemiological evidence. In our review, we found that TB models frequently stratified active TB based on sputum smear-status, which is consistent with our recent meta-analysis that found that contacts of smear-positive TB patients were more likely to be infected compared contacts of smear-negative patients (Melsew et al., 2018a). Positive sputum smear results are indicative of higher bacillary concentrations and disease that is in direct communication with the individuals' airway, such that these individuals are often highly infectious (Ait-Khaled et al., 2003). Several models also recognised that undetected active TB cases are typically more infectious than those receiving treatment. This is consistent with epidemiological studies that have showed that treatment rapidly reduces patients' infectiousness soon after commencement (Lohmann et al., 2012; Otero et al., 2016; Tornee et al., 2004; Xu et al., 2008). Using HIV co-infection as a factor to stratify the infectiousness of active TB cases is also broadly supported by meta-analysis results that found contacts exposed to HIV-positive TB patients had a 55% reduced risk of infection compared to those exposed to HIV-negative patients, although there were considerable differences between the individual studies (Cayla et al., 1996; Melsew et al., 2018b). There is also some evidence to support the higher infectiousness of smokers compared to non-smokers (Lohmann et al., 2012; Huang et al., 2014; Singh et al., 2005), although there is no strong epidemiological evidence that TB patients with co-morbid diabetes mellitus have increased infectiousness (Grandjean et al., 2015). Although evidence shows that drug-resistant strains are typically less transmissible than drug-susceptible strains, the overall fitness of the strain depends on both the transmission rate and duration of the infectious period. The poorer rates of MDR-TB diagnosis and so longer periods of infectiousness may offset the fitness cost, such that there is no clear consensus on the impact of drug resistance on TB transmission (Luciani et al., 2009; Cohen et al., 2003).

Models vary in regard to their assumptions concerning conversion from non-infectious states or states with limited infectiousness to more infectious forms of disease. Most heterogeneous models did not simulate this conversion, but rather assumed that individuals immediately became smear-positive or negative at the point of disease activation. Of those that did allow for this transition phenomenon, most models assumed smear-negatives could progress to smear-positivity with time. Active TB disease progression can be viewed as a dynamic continuum, with individuals progressing from a less infectious stage of disease to a more highly infectious stage as tissue damage increases (Pai et al., 2016). However, it is impossible to quantify this phenomenon as it is unethical to observe the natural history of untreated TB (Tiemersma et al., 2011). Nevertheless, models consistently assumed such small parameter values for the rates governing this transition as to be negligible, such that if this process is epidemiologically important then past

models are unlikely to be fully capturing it.

The impact of heterogeneous infectiousness in transmission dynamic models can be explored through the basic reproduction number, R_o, with marked reductions in R_0 potentially achievable by targeting control measures towards a small proportion of the population that are highly infectious (Trauer et al., 2018; Matthews et al., 2006). Increases in heterogeneous infectiousness may not necessarily lead to epidemics of a greater size for a given R_o although local outbreaks may act as driving sources of sustained disease burden (Andersson and Britton, 1998; Miller, 2007a; Miller, 2007b). As global TB control targets elimination, the phenomenon of heterogeneous infectiousness is likely to become more apparent, increasingly requiring TB models to incorporate heterogeneous infectiousness assumptions.

5. Conclusions

Given the evidence of extensive heterogeneity in the infectiousness of individuals with TB, there exists an argument for the stratification of active TB states by infectiousness in models. The main consideration for assuming heterogeneous infectiousness and thus stratifying active TB based on infectiousness is the objective of the model; in general, when the objective of the model is to evaluate effectiveness of targeted public health interventions or evaluation of diagnostic tools, modellers should consider incorporating heterogeneous infectiousness. Consensus has not been reached as to the optimal approach to stratifying TB models to capture the considerable heterogeneity in infectiousness between individuals, and it may be that there is no a single approach that will suit all epidemiological scenarios and research questions.

Availability of data and materials

A list of included studies has been made available. The study protocol can be accessed on PROSPERO, registration number: CRD42019111936.

Authors' contributions

YAM and JMT conceived the study, which was refined by ESM, ACC and RR. YAM developed the data extraction checklist, which was and approved by JMT, RR and AIA. YAM and AIA extracted the data. YAM drafted the manuscript, and all authors provided input into revisions and approved the final draft for submission.

Declaration of Competing Interests

The authors declare that they have no competing interests.

Acknowledgements

The authors are grateful to staff at the Ian Potter Library for their assistance in developing the search strategy. We are also grateful to the Monash Graduate Scholarship program for providing a scholarship to YAM for his PhD study.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.epidem.2019.100374.

References

Abu-Raddad, L.J., et al., 2009. Epidemiological benefits of more-effective tuberculosis vaccines, drugs, and diagnostics. Proc. Natl. Acad. Sci. 106 (33), 13980–13985.
Ackley, S.F., et al., 2015. Modeling historical tuberculosis epidemics among Canadian First Nations: effects of malnutrition and genetic variation. PeerJ 3, e1237.
Agusto, F.B., et al., 2015. Mathematical model of MDR-TB and XDR-TB with isolation and lost to follow-up. Abstr. Appl. Anal. 2015.
Ahmadin and Fatmawati, 2014. Mathematical modeling of drug resistance in tuberculosis transmission and optimal control treatment. Appl. Math. Sci. 8 (92), 4547–4559.
Ait-Khaled, N., Enarson, D.A., Initiative, S.T., 2003. Tuberculosis: a Manual for Medical Students. World Health Organization, Geneva.
Andersson, H., Britton, T., 1998. Heterogeneity in epidemic models and its effect on the spread of infection. J. Appl. Probab. 35 (3), 651–661.
Aparicio, J.P., Castillo-Chavez, C., 2009. Mathematical modelling of tuberculosis epidemics. Math. Biosci. Eng.: MBE 6 (2), 209–237.
Apriliani, V., Jaharuddin, Sianturi, P., 2016. Mathematical model of tuberculosis spread within two groups of infected population. Appl. Math. Sci. 10 (41–44), 2131–2140.
Arregui, S., et al., 2018. Data-driven model for the assessment of Mycobacterium tuberculosis transmission in evolving demographic structures. Proc. Natl. Acad. Sci. U.S.A. 115 (14), E3238–E3245.
Bacaer, N., et al., 2008. Modeling the joint epidemics of TB and HIV in a South African township. J. Math. Biol. 57 (4), 557–593.
Basu, S., et al., 2007. Prevention of nosocomial transmission of extensively drug-resistant tuberculosis in rural South African district hospitals: an epidemiological modelling study. Lancet 370 (9597), 1500–1507.
Basu, S., et al., 2009. Primary and secondary tuberculosis preventive treatment in HIV clinics: simulating alternative strategies. Int. J. Tuberc. Lung Dis. 13 (5), 652–658.
Bhunu, C., Garira, W., 2009. A two strain tuberculosis transmission model with therapy and quarantine. Math. Model. Anal. 14 (3), 291–312.
Bhunu, C., Mushayabasa, S., Tchuenche, J., 2011. A theoretical assessment of the effects of smoking on the transmission dynamics of tuberculosis. Bull. Math. Biol. 73 (6), 1333–1357.
Bishai, J.D., Bishai, W.R., Bishai, D.M., 2010. Heightened vulnerability to MDR-TB epidemics after controlling drug-susceptible TB. PLoS One 5 (9), e12843.
Blower, S.M., Gerberding, J.L., 1998. Understanding, predicting and controlling the emergence of drug-resistant tuberculosis: a theoretical framework. J. Mol. Med. 76 (9), 624–636.
Blower, S.M., et al., 1995. The intrinsic transmission dynamics of tuberculosis epidemics. Nat. Med. 1 (8), 815–821.
Bowong, S., Alaoui, A.A., 2013. Optimal intervention strategies for tuberculosis. Commun. Nonlinear Sci. Numer. Simul. 18 (6), 1441–1453.
Cayla, J.A., et al., 1996. The influence of intravenous drug use and HIV infection in the transmission of tuberculosis. AIDS 10 (1), 95–100.
Cohen, T., Murray, M., 2004. Modeling epidemics of multidrug-resistant M. tuberculosis of heterogeneous fitness. Nat. Med. 10 (10), 1117–1121.
Cohen, T., White, P., 2016. Transmission-dynamic models of infectious diseases. Infect. Dis. Epidemiol. 223.
Cohen, T., Sommers, B., Murray, M., 2003. The effect of drug resistance on the fitness of Mycobacterium tuberculosis. Lancet Infect. Dis. 3 (1), 13–21.
Cudahy, P., Shenoi, S.V., 2016. Diagnostics for pulmonary tuberculosis. Postgrad. Med. J. 92 (1086), 187–193.
Curtis, A.B., et al., 1999. Extensive transmission of Mycobacterium tuberculosis from a child. N. Engl. J. Med. 341 (20), 1491–1495.
De Espíndola, A.L., et al., 2011. An agent-based computational model of the spread of tuberculosis. J. Stat. Mech. Theory Exp. 2011 (5).
Dowdy, D.W., Chaisson, R.E., 2009. The persistence of tuberculosis in the age of DOTS: reassessing the effect of case detection. Bull. World Health Organ. 87, 296–304.
Dowdy, D.W., et al., 2006. The potential impact of enhanced diagnostic techniques for tuberculosis driven by HIV: a mathematical model. AIDS 20 (5), 751–762.
Dowdy, D.W., et al., 2008. Impact of enhanced tuberculosis diagnosis in South Africa: a mathematical model of expanded culture and drug susceptibility testing. Proc. Natl. Acad. Sci.
Dowdy, D.W., Basu, S., Andrews, J.R., 2013a. Is passive diagnosis enough? The impact of subclinical disease on diagnostic strategies for tuberculosis. Am. J. Respir. Crit. Care Med. 187 (5), 543–551.
Dowdy, D.W., et al., 2013b. Population-level impact of same-day microscopy and Xpert MTB/RIF for tuberculosis diagnosis in Africa. PLoS One 8 (8), e70485.
Dowdy, D.W., et al., 2013c. Population-level impact of active tuberculosis case finding in an Asian megacity. PLoS One 8 (10), e77517.
Dowdy, D.W., et al., 2014. A user-friendly, open-source tool to project impact and cost of diagnostic tests for tuberculosis. Elife 3, e02565.
Dye, C., Williams, B.G., 2008. Eliminating human tuberculosis in the twenty-first century. J. R. Soc. Interface 5 (23), 653–662.
Espindola, A.L., et al., 2012. Exploration of the parameter space in an agent-based model of tuberculosis spread: emergence of drug resistance in developing vs developed countries. Int. J. Mod. Phys. C 23 (6), 1250046.
Fofana, M.O., et al., 2017. A multistrain mathematical model to investigate the role of pyrazinamide in the emergence of extensively drug-resistant tuberculosis. Antimicrob. Agents Chemother. 61 (3), e00498–16.
Garcia, A., Maccario, J., Richardson, S., 1997. Modelling the annual risk of tuberculosis infection. Int. J. Epidemiol. 26 (1), 190–203.
Gomes, P.D., Leal-Toledo, R.C., Cunha, C., 2007. Dynamics of tuberculosis under dots strategy, in Biomat 2006. World Sci. 161–180.
Grandjean, L., et al., 2015. Transmission of multidrug-resistant and drug-susceptible

tuberculosis within households: a prospective cohort study. PLoS Med. 12 (6) (no pagination) (e1001843).
Guzzetta, G., et al., 2011. Modeling socio-demography to capture tuberculosis transmission dynamics in a low burden setting. J. Theor. Biol. 289, 197–205.
Herrera, M., et al., 2013. Modeling the spread of tuberculosis in semiclosed communities. Comput. Math. Methods Med. 2013.
Hickson, R., Mercer, G., Lokuge, K., 2011. Sensitivity Analysis of a Model for Tuberculosis. In 19th International Congress on Modelling and Simulation.
Hickson, R.I., Mercer, G.N., Lokuge, K.M., 2012. A metapopulation model of tuberculosis transmission with a case study from high to low burden areas. PLoS One 7 (4), e34411.
Hill, A., Becerra, J., Castro, K., 2012. Modelling tuberculosis trends in the USA. Epidemiol. Infect. 140 (10), 1862–1872.
Houben, R., et al., 2016. TIME Impact–a new user-friendly tuberculosis (TB) model to inform TB policy decisions. BMC Med. 14 (1), 56.
Huang, C.C., et al., 2014. Cigarette smoking among tuberculosis patients increases risk of transmission to child contacts. Int. J. Tuberc. Lung Dis. 18 (11), 1285–1291.
Hughes, G.R., Currie, C.S., Corbett, E.L., 2006. Modeling tuberculosis in areas of high HIV prevalence. Simulation Conference, 2006. WSC 06. Proceedings of the Winter. IEEE.
Huo, H.-F., Zou, M.-X., 2016. Modelling effects of treatment at home on tuberculosis transmission dynamics. Appl. Math. Model. 40 (21–22), 9474–9484.
Huynh, G.H., et al., 2015. Tuberculosis control strategies to reach the 2035 global targets in China: the role of changing demographics and reactivation disease. BMC Med. 13 (1), 88.
Jung, E., Lenhart, S., Feng, Z., 2002. Optimal control of treatments in a two-strain tuberculosis model. Discret. Contin. Dyn. Syst. - Ser. B 2 (4), 473–482.
Kasaie, P., Dowdy, D.W., Kelton, W.D., 2013. An agent-based simulation of a tuberculosis epidemic: understanding the timing of transmission. Proceedings of the 2013 Winter Simulation Conference: Simulation: Making Decisions in a Complex World. IEEE Press.
Kendall, E.A., Fofana, M.O., Dowdy, D.W., 2015. Burden of transmitted multidrug resistance in epidemics of tuberculosis: a transmission modelling analysis. Lancet Respir. Med. 3 (12), 963–972.
Kendall, E.A., et al., 2017. MDR-TB treatment as prevention: the projected population-level impact of expanded treatment for multidrug-resistant tuberculosis. PLoS One 12 (3), e0172748.
Knight, G.M., et al., 2014. Impact and cost-effectiveness of new tuberculosis vaccines in low-and middle-income countries. Proc. Natl. Acad. Sci. 111 (43), 15520–15525.
Knight, G.M., et al., 2015. Tuberculosis prevention in South Africa. PLoS One 10 (4), e0122514.
Korthals Altes, H., et al., 2018. Latent tuberculosis infection in foreign-born communities: import vs. transmission in The Netherlands derived through mathematical modelling [Erratum appears in PLoS One. 2018 May 24;13(5):e0198376; PMID: 29795693]. PLoS ONE [Electronic Resource] 13 (2), e0192282.
Legrand, J., et al., 2008. Modeling the impact of tuberculosis control strategies in highly endemic overcrowded prisons. PLoS One 3 (5), e2100.
Li, X.-Z., et al., 2011. A tuberculosis (TB) model with undetected compartment: an application to China. J. Biol. Syst. 19 (2), 205–236.
Liao, C.-M., Lin, Y.-J., 2012. Assessing the transmission risk of multidrug-resistant Mycobacterium tuberculosis epidemics in regions of Taiwan. Int. J. Infect. Dis. 16 (10), e739–e747.
Liao, C.M., et al., 2012. A probabilistic transmission and population dynamic model to assess tuberculosis infection risk. Risk Anal.: Int. J. 32 (8), 1420–1432.
Liao, C.-M., Lin, Y.-J., Cheng, Y.-H., 2013. Modeling the impact of control measures on tuberculosis infection in senior care facilities. Build. Environ. 59, 66–75.
Liberati, A., et al., 2009. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. PLoS Med. 6 (7), e1000100.
Lin, H.-H., et al., 2012. The impact of new tuberculosis diagnostics on transmission: why context matters. Bull. World Health Organ. 90, 739–747.
Liu, Y., Sun, Z., 2010. A new model for MDR-TB infection with undetected TB cases. Control and Decision Conference (CCDC), 2010 Chinese. IEEE.
Liu, L., Zhou, Y., Wu, J., 2008. Global dynamics in a TB model incorporating case detection and two treatment stages. Rocky Mt. J. Math. 1541–1559.
Liu, Y., et al., 2011. Modeling transmission of tuberculosis with MDR and undetected cases. Discrete Dyn. Nat. Soc. 2011.
Lohmann, E.M., et al., 2012. Grading of a positive sputum smear and the risk of Mycobacterium tuberculosis transmission. Int. J. Tuberc. Lung Dis. 16 (11), 1477–1484.
Luciani, F., et al., 2009. The epidemiological fitness cost of drug resistance in Mycobacterium tuberculosis. Proc. Natl. Acad. Sci. 106 (34), 14711–14715.
Matthews, L., et al., 2006. Heterogeneous shedding of Escherichia coli O157 in cattle and its implications for control. Proc. Natl. Acad. Sci. 103 (3), 547–552.
McBryde, E.S., et al., 2017. The risk of global epidemic replacement with drug-resistant Mycobacterium tuberculosis strains. Int. J. Infect. Dis. 56, 14–20.
McLean, A.R., 2013. Infectious disease modeling. Infectious Diseases. Springer, pp. 99–115.
Mellor, G.R., Currie, C.S., Corbett, E.L., 2011. Incorporating household structure into a discrete-event simulation model of tuberculosis and HIV. ACM Trans. Model. Comput. Simul. (TOMACS) 21 (4), 26.
Melnichenko, A., Romanyukha, A.A., 2009. A model of tuberculosis epidemiology: data analysis and estimation of parameters. Math. Model. Comput. Simul. 1 (4), 428–444.
Melsew, Y., et al., 2018a. Risk factors for infectiousness of patients with tuberculosis: a systematic review and meta-analysis. Epidemiol. Infect. 146 (3), 345–353.
Melsew, Y., et al., 2018b. Risk factors for infectiousness of patients with tuberculosis: a systematic review and meta-analysis. Epidemiol. Infect. 1–9.
Melsew, Y.A., et al., 2019. The role of super-spreading events in Mycobacterium tuberculosis transmission: evidence from contact tracing. BMC Infect. Dis. 19 (1), 244.
Menzies, N.A., et al., 2012. Population health impact and cost-effectiveness of tuberculosis diagnosis with Xpert MTB/RIF: a dynamic simulation and economic evaluation. PLoS Med. 9 (11), e1001347.
Menzies, N.A., et al., 2018. Prospects for tuberculosis elimination in the United States: results of a transmission dynamic model. Am. J. Epidemiol.
Miller, J.C., 2007a. Epidemic size and probability in populations with heterogeneous infectivity and susceptibility. Phys. Rev. E 76 (1), 010101.
Miller, J.C., 2007b. Predicting the Size and Probability of Epidemics in a Population With Heterogeneous Infectiousness and Susceptibility. arXiv preprint q-bio/0702007. .
Moualeu, D., et al., 2012. Analysis of the impact of diabetes on the dynamical transmission of tuberculosis. Math. Model. Nat. Phenom. 7 (3), 117–146.
Moualeu, D.P., Bowong, S., Kurths, J., 2014. Parameter estimation of a tuberculosis model in a patchy environment: case of Cameroon. BIOMAT 2013: International Symposium on Mathematical and Computational Biology. World Scientific.
Moualeu, D.P., et al., 2015. Optimal control for a tuberculosis model with undetected cases in Cameroon. Commun. Nonlinear Sci. Numer. Simul. 20 (3), 986–1003.
Moualeu, D., et al., 2016. Analysis of a tuberculosis model with undetected and lost-sight cases. Commun. Nonlinear Sci. Numer. Simul. 41, 48–63.
Moualeu, D.P., et al., 2018. A patchy model for the transmission dynamics of tuberculosis in sub-Saharan Africa. Int. J. Dyn. Control. 6 (1), 122–139.
Moualeu-Ngangue, D.P., et al., 2015. Parameter identification in a tuberculosis model for Cameroon. PLoS One 10 (4).
Mushayabasa, S., Bhunu, C.P., 2013. Modeling the impact of early therapy for latent tuberculosis patients and its optimal control analysis. J. Biol. Phys. 39 (4), 723–747.
Nishiura, H., Patanarapelert, K., Tang, I.M., 2004. Predicting the future trend of drug-resistant tuberculosis in Thailand: assessing the impact of control strategies. Southeast Asian J. Trop. Med. Public Health 35, 649–656.
Nyabadza, F., Kgosimore, M., 2012. Modeling the dynamics of tuberculosis transmission in children and adults. J. Math. Stat. 8 (2), 229–240.
Okuonghae, D., 2013. A mathematical model of tuberculosis transmission with heterogeneity in disease susceptibility and progression under a treatment regime for infectious cases. Appl. Math. Model. 37 (10–11), 6786–6808.
Okuonghae, D., Aihie, V., 2008. Case detection and direct observation therapy strategy (DOTS) in Nigeria: its effect on TB dynamics. J. Biol. Syst. 16 (1), 1–31.
Okuonghae, D., Aihie, V.U., 2010. Optimal control measures for tuberculosis mathematical models including immigration and isolation of infective. J. Biol. Syst. 18 (1), 17–54.
Okuonghae, D., Ikhimwin, B.O., 2016. Dynamics of a mathematical model for tuberculosis with variability in susceptibility and disease progressions due to difference in awareness level. Front. Microbiol. 6, 1530.
Okuonghae, D., Korobeinikov, A., 2007. Dynamics of tuberculosis: the effect of direct observation therapy strategy (DOTS) in Nigeria. Math. Model. Nat. Phenom. 2 (1), 113–128.
Okuonghae, D., Omosigho, S.E., 2011. Analysis of a mathematical model for tuberculosis: What could be done to increase case detection. J. Theor. Biol. 269 (1), 31–45.
Osgood, N.D., et al., 2011. Estimating the relative impact of early-life infection exposure on later-life tuberculosis outcomes in a Canadian sample. Res. Hum. Dev. 8 (1), 26–47.
Otero, L., et al., 2016. A prospective longitudinal study of tuberculosis among household contacts of smear-positive tuberculosis cases in Lima, Peru. BMC Infect. Dis. 16 (1) (no pagination) (259).
Oxlade, O., et al., 2011. Developing a tuberculosis transmission model that accounts for changes in population health. Med. Decis. Mak. 31 (1), 53–68.
Pai, M., et al., 2016. Tuberculosis. Nat. Rev. Dis. Primers 2, 16076.
Pandey, S., et al., 2017. Estimating tuberculosis incidence from primary survey data: a mathematical modeling approach. Int. J. Tuberc. Lung Dis. 21 (4), 366–374.
Perelman, M., et al., 2004. Tuberculosis epidemiology in Russia: the mathematical model and data analysis. Russian Journal of Numerical Analysis and Mathematical Modelling rnam 19 (4), 305–314.
Raimundo, S.M., Yang, H.M., Venturino, E., 2014. Theoretical assessment of the relative incidences of sensitive and resistant tuberculosis epidemic in presence of drug treatment. Math. Biosci. Eng. 11 (4), 971–993.
Rayhan, S., Bakhtiar, T., 2017. Two-strain tuberculosis transmission model under Three control strategies. IOP Conference Series: Earth and Environmental Science. IOP Publishing.
Rodrigues, C.G., Espíndola, A.L., Penna, T., 2015. An agent-based computational model for tuberculosis spreading on age-structured populations. Phys. A Stat. Mech. Its Appl. 428, 52–59.
Sachdeva, K.S., et al., 2015. The potential impact of up-front drug sensitivity testing on India's epidemic of multi-drug resistant tuberculosis. PLoS One 10 (7), e0131438.
Salje, H., et al., 2014. The importance of implementation strategy in scaling up Xpert MTB/RIF for diagnosis of tuberculosis in the Indian health-care system: a transmission model. PLoS Med. 11 (7), e1001674.
Salomon, J.A., et al., 2006. Prospects for advancing tuberculosis control efforts through novel therapies. PLoS Med. 3 (8), e273.
Sharomi, O., et al., 2008. Mathematical analysis of the transmission dynamics of HIV/TB coinfection in the presence of treatment. Math. Biosci. Eng. 5 (1), 145.
Sharomi, O.Y., et al., 2017. Exogenous re-infection does not always cause backward bifurcation in TB transmission dynamics. Appl. Math. Comput. 298, 322–335.
Shrestha, S., et al., 2014. Drivers and trajectories of resistance to new first-line drug regimens for tuberculosis. Open Forum Infectious Diseases. Oxford University Press.
Singh, M., et al., 2005. Prevalence and risk factors for transmission of infection among children in household contact with adults having pulmonary tuberculosis. Arch. Dis. Child. 90 (6), 624–628.

Suen, S., Bendavid, E., Goldhaber-Fiebert, J.D., 2014. Disease control implications of India's changing multi-drug resistant tuberculosis epidemic. PLoS One 9 (3), e89822.
Sun, Z., et al., 2011. Modeling transmission of tuberculosis with MDR and undetected cases. Discrete Dyn. Nat. Soc. 2011.
Sun, A.Y., et al., 2013. Modeling the impact of alternative strategies for rapid molecular diagnosis of tuberculosis in Southeast Asia. Am. J. Epidemiol. 178 (12), s1740–1749.
Thomas, E.G., et al., 2010. Modelling the spread of tuberculosis, including drug resistance and HIV: a case study in Papua New Guinea's Western Province. The ANZIAM J. 52 (1), 26–45.
Tiemersma, E.W., et al., 2011. Natural history of tuberculosis: duration and fatality of untreated pulmonary tuberculosis in HIV negative patients: a systematic review. PLoS One 6 (4), e17601.
Tornee, S., et al., 2004. Risk factors for tuberculosis infection among household contacts in Bangkok, Thailand. Southeast Asian J. Trop. Med. Public Health 35 (2), 375–383.
Trauer, J.M., Denholm, J.T., McBryde, E.S., 2014. Construction of a mathematical model for tuberculosis transmission in highly endemic regions of the Asia-Pacific. J. Theor. Biol. 358, 74–84.
Trauer, J.M., et al., 2016a. Scenario analysis for programmatic tuberculosis control in Western Province, Papua New Guinea. Am. J. Epidemiol. 183 (12), 1138–1148.
Trauer, J.M., et al., 2016b. Modelling the effect of short-course multidrug-resistant tuberculosis treatment in Karakalpakstan, Uzbekistan. BMC Med. 14 (1), 187.
Trauer, J.M., et al., 2018. The importance of heterogeneity to the epidemiology of Tuberculosis. Clin. Infect. Dis.
Tuite, A.R., et al., 2017. Stochastic agent-based modeling of tuberculosis in Canadian Indigenous communities. BMC Public Health 17 (1), 73.
Turner, R.D., et al., 2017. Tuberculosis infectiousness and host susceptibility. J. Infect. Dis. 216 (suppl_6), S636–S643.
Valway, S.E., et al., 1998. An outbreak involving extensive transmission of a virulent strain of Mycobacterium tuberculosis. N. Engl. J. Med. 338 (10), 633–639.
Vynnycky, E., et al., 2015. Tuberculosis control in South African gold mines: mathematical modeling of a trial of community-wide isoniazid preventive therapy. Am. J. Epidemiol. 181 (8), 619–632.
Waaler, H., Geser, A., Andersen, S., 1962. The use of mathematical models in the study of the epidemiology of tuberculosis. Am. J. Public Health Nations Health 52, 1002–1013.
WHO, 2014. The End TB Strategy: Global Strategy and Targets for Tuberculosis Prevention, Care and Control After 2015. Geneva, Switzerland.
WHO, 2018. Global Tuberculosis Report 2018. Geneva.
Wu, P., et al., 2010. The transmission dynamics of tuberculosis in a recently developed Chinese city. PLoS One 5 (5), e10468.
Xu, L., et al., 2008. Dose-response relationship between treatment delay of smear-positive tuberculosis patients and intra-household transmission: a cross-sectional study. Trans. R. Soc. Trop. Med. Hyg. 102 (8), 797–804.
Ypma, R.J., et al., 2013. A sign of superspreading in tuberculosis: highly skewed distribution of genotypic cluster sizes. Epidemiology 24 (3), 395–400.
Zhang, J., Li, Y., Zhang, X., 2015. Mathematical modeling of tuberculosis data of China. J. Theor. Biol. 365, 159–163.

Appendix A. PRISMA 2009 checklist for systematic review of Heterogeneous infectiousness in mathematical models of tuberculosis

	#	Checklist item	Reported on page #
TITLE			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	1
ABSTRACT			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number. The abstract is structured into **Background, methods, results and conclusions** according to BMC medicine guideline	2
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	3-4
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	4
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	4
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	5
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	5
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	5 Appendix B
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	5

Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	5-6
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	6
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	NA*
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	NA*
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I^2) for each meta-analysis.	NA*

From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(7): e1000097. doi:10.1371/journal.pmed1000097

NA* This study being a methodological review intends to present diverse approaches used in transmission dynamic models of TB to capture infectiousness heterogeneity. Hence are not suitable for epidemiological pooling .Therefore, risk of bias assessment, meta analysis, meta regression, sensitivity or subgroup analyses, effect estimates and confidence intervals, a forest plots were not done.

Appendix B: Electronic search strategy

Database	**Search strategy**
MEDLINE	TB OR Tuberculosis OR Mycobacterium tuberculosis AND Dynamic* OR Model* OR Simulat*
EMBASE	TB OR Tuberculosis OR Mycobacterium tuberculosis AND Dynamic* OR Model* OR Simulat*
Biosis	TB OR Tuberculosis OR Mycobacterium tuberculosis AND Dynamic* OR Model* OR Simulat*
Global Health	TB OR Tuberculosis AND Dynamic* OR Model* OR Simulat*
Scopus	TB OR Tuberculosis AND Dynamic* OR Model* OR Simulat*

Chapter 6

6. Capturing heterogeneous infectiousness in dynamic transmission models of tuberculosis: a compartmental modelling approach

6.1. Chapter overview

The systematic review of previous TB modelling studies in Chapter 5 indicated the need for a data-driven model of TB that can capture heterogeneous infectiousness, which was demonstrated in the previous chapters. This Chapter presents a new transmission dynamic model of TB incorporating heterogeneous infectiousness assumption with three levels of TB patients' infectiousness, namely: non-spreaders, low-spreaders and super-spreaders. This Chapter uses data on the number of secondary infections produced in Chapter 3 to estimate model parameters that pertain to heterogeneous infectiousness using a Bayesian inference method. By building a data-driven model with heterogeneous infectiousness assumption, this Chapter evaluated the effectiveness of an active case finding intervention targeted towards super-spreaders and compares it to an untargeted intervention. It also highlights how intensifying targeted active TB case finding towards super-spreaders is more effective as compared to mass or population-wide interventions. By addressing thesis question 5, this Chapter reported that targeting highly infectious individuals with TB is considerably more effective than untargeted intervention.

6.2. Abstract

Infectiousness heterogeneity among individuals with tuberculosis (TB) is substantial and is likely to have a significant impact on the long-term dynamics of TB and the effectiveness of interventions. However, there is a gap in capturing heterogeneous infectiousness and evaluating its impact on the effectiveness of interventions.

Informed by observed distribution of secondary infections, we constructed a deterministic model of TB transmission using ordinary differential equations. The model incorporated assumption of heterogeneous infectiousness with three levels of infectivity, namely non-spreaders, low-spreaders and super-spreaders. We evaluated the effectiveness of dynamic transmission untargeted and targeted implementation of an intervention intended to represent active case finding with a point-of-care diagnostic tool. The simulated intervention detected 20% of all TB patients who would otherwise have been missed by the health system during their disease episode and was compared across four epidemiological scenarios.

Our model suggested that targeting the active case finding intervention towards super-spreaders was more effective than untargeted intervention in all setting scenarios, with more effectiveness in settings with low case detection and high transmission intensity. For instance, a targeted intervention achieved a 42.2% reduction in TB incidence, while the untargeted intervention achieved only a 20.7% reduction over 20 years, given the same number of people treated. The increase in the proportion of the intervention targeted towards super-spreaders had a similar effect on incidence reduction as the increase in the overall intervention. Although the most marked impact on equilibrium TB incidence came from the rate of late reactivation, the proportion of super-spreaders and their relative infectiousness had shown substantial impact.

Targeting active case-finding interventions to highly infectious cases likely to be particularly beneficial in settings where case detection is poor. Heterogeneity-related parameters had an equivalent effect to several other parameters that have been established as being very important to TB transmission dynamics.

Keywords: heterogeneous infectiousness; M. tuberculosis transmission; mathematical model;

6.3. Introduction

Tuberculosis (TB) is the world's leading cause of death from a single infectious agent, in 2019, ranking above HIV/AIDS with an estimated 10.0 million cases and 1.2 million deaths worldwide [1]. The World Health Organization developed the new End TB Strategy for post-2015 TB elimination activities, with the ambition of a 95% reduction in TB deaths and a 90% reduction in TB incidence by 2035 by comparison to 2015 rates [2]. However, the natural history of TB remains poorly understood, and consequently, the uncertain potential impact of control interventions limits confidence about the possibility of its elimination. The heterogeneous transmission of *Mycobacterium tuberculosis (Mtb)* within populations is well-established, but its epidemiological impact is poorly understood. Moreover, up to a third of TB cases are not diagnosed, so finding and treating infectious cases is key to achieving TB control, and finding and treating highly infectious people is the key to TB control in high transmission settings [3, 4].

We define infectiousness heterogeneity as the variability in the capacity of infectious patients to produce secondary infections, which may be attributable to characteristics of the host, the agent and the environment [5-9]. Heterogeneity of infectious individuals in spreading *Mtb* infection is well-recognised, with a small group of highly infectious individuals producing a large proportion of secondary infections, while many others produce very few or none [10]. Previous TB genomic epidemiology has identified TB super-spreading events by quantifying

heterogeneity in the infectiousness of TB patients through fitting a standard statistical distribution (the negative binomial distribution, NBD) to the distribution of secondary cases produced by each infectious patient (the "offspring distribution") [11]. More recently, data from TB contacts in a low-transmission setting have allowed estimation of the proportion of all TB patients that can be categorised as super-spreaders, finding it to be approximately 10% [9]. In such studies, super-spreaders were typically defined as patients who produced a number of secondary "infections/cases" greater than the 99th centile of a standard Poisson offspring distribution, with distribution mean equal to the average number of secondary infections per index [9, 12]. This heterogeneity is likely to have implications for both the burden of disease and the effectiveness of control interventions. Thus, we propose that it is necessary to capture this heterogeneity when modelling TB transmission dynamics.

Some past compartmental models of TB transmission dynamics have attempted to capture heterogeneity in patients' infectiousness by stratifying the active TB compartment into different levels of infectivity. Amongst the most typical approaches is stratification as either infectious (usually representing pulmonary TB) or non-infectious (extrapulmonary TB) [13-17]. This approach implies that all pulmonary TB patients are equally infectious, while all extrapulmonary patients are entirely non-infectious, and so does not capture the heterogeneity among pulmonary patients. Other TB models have considered sputum smear status as a factor in stratifying patients' levels of infectiousness, considering both smear-negative and smear-positive pulmonary TB to be infectious, with the relative infectiousness of smear-negative patients compared to smear-positive typically set between 15 and 25% [18-33]. While stratifying pulmonary TB patients based on smear status captures an additional clinical attribute that is important in determining infectiousness, even these models still do not capture the full picture of TB patients' infectiousness variation, since several behavioural and demographic factors other than smear status can affect the level of infectiousness [5].

An individual-based model simulated patterns of meetings and *Mtb* transmission between three different types of contact to determine the effects of variation in infectiousness and susceptibility on transmission location. This study suggested that the majority of disease resulted from infection by a small proportion of people with TB or super-spreaders [34]. This study defined super-spreading defined based on the number of secondary cases instead of secondary infections, however, as TB disease activation may take long time and also depend on mainly on the characteristics of contact person [35], this definition may not show true heterogeneity in *Mtb* transmission.

In this study, we present a deterministic compartmental *Mtb* transmission model that incorporates three levels of TB patients' infectivity, namely non-spreaders, low-spreaders and super-spreaders. Using this framework, we analysed how infectiousness heterogeneity affects disease dynamics and evaluated the effectiveness of a hypothetical active case finding intervention and the impact of targeting super-spreaders.

6.4. Methods

Literature review

Before constructing our model, we systematically reviewed how previous TB models have approached heterogeneous infectiousness [36]. In the review, we found that TB models frequently stratified the active TB compartment according to one or more patient-related factors. We constructed a flexible model with three levels of infectiousness, without restricting to a given factor, and the ability to transition between these levels using empirical measures of heterogeneous infectiousness to parametrise the model.

Empiric data

Data from the VTP which are stored by the Victorian Department of Health and Human Services (DHHS) were used for the following analysis. Index patients were classified as

confirmed cases of TB notified from 1 January 2005 to 31 December 2015 in residents of Victoria. The data set includes contact tracing information and results of testing for *Mtb* infection, with cases of subsequent active TB disease linked to these contact episodes now extending to March 2017 (see [37] for earlier publication of linkage process). We constructed empirical offspring distributions from the detailed contact tracing data set of the VTP.

Ethical approval was obtained from Monash University, Human Research Ethics Committee (Project Number: 7776) and permission was given by the VTP and DHHS.

Model structure

Using ordinary differential equations (ODE), we constructed a deterministic model of *Mtb* transmission in a hypothetical high TB burden setting. To represent heterogeneous infectiousness and super-spreading in *Mtb* transmission, we stratified the active TB compartment into three sub-classes with different levels of infectiousness: non-spreaders (I_0), low-spreaders (I_1) and super-spreaders (I_2). We did not use any specific factor to discriminate I_0, I_1 and I_2 a priori. Instead, we use the offspring data (number of secondary infections per index) and a Bayesian approach to find the most realistic parameterisation to implement heterogeneous infectiousness in our model.

Fig 1 presents the model structure. The model simulates a closed population, such that births (π) accrue into the fully susceptible compartment (S) to replace all deaths. Natural death occurs from all compartments at a constant rate (μ). Susceptible individuals are dynamically infected, with the force of infection (λ) contributed by super-spreaders and low-spreaders. Following infection, individuals enter a rapid-sojourn early latent compartment (L_A), from which they may progress to one of the three active TB compartments (I_0, I_1 or I_2) at a total rate of ε or enter the low-risk late latent stage (L_B) at rate κ. From the late latent state they may progress more slowly to disease (total rate ν), also entering the active TB compartments

(I_0, I_1, I_2). Re-infection may occur during late latency with a reduced force of infection λ_r ($\lambda_r = r \times \lambda$) and, if re-infection occurs, these individuals similarly enter the early latent compartment, progressing to the disease at the same rates as for newly infected persons. Individuals with active disease may either die due to TB (rate μ_i) or background mortality (μ); or spontaneously recover and return to the late latent state (rate γ); or be detected and treated by the health system (rate δ) and return to the susceptible compartment (S).

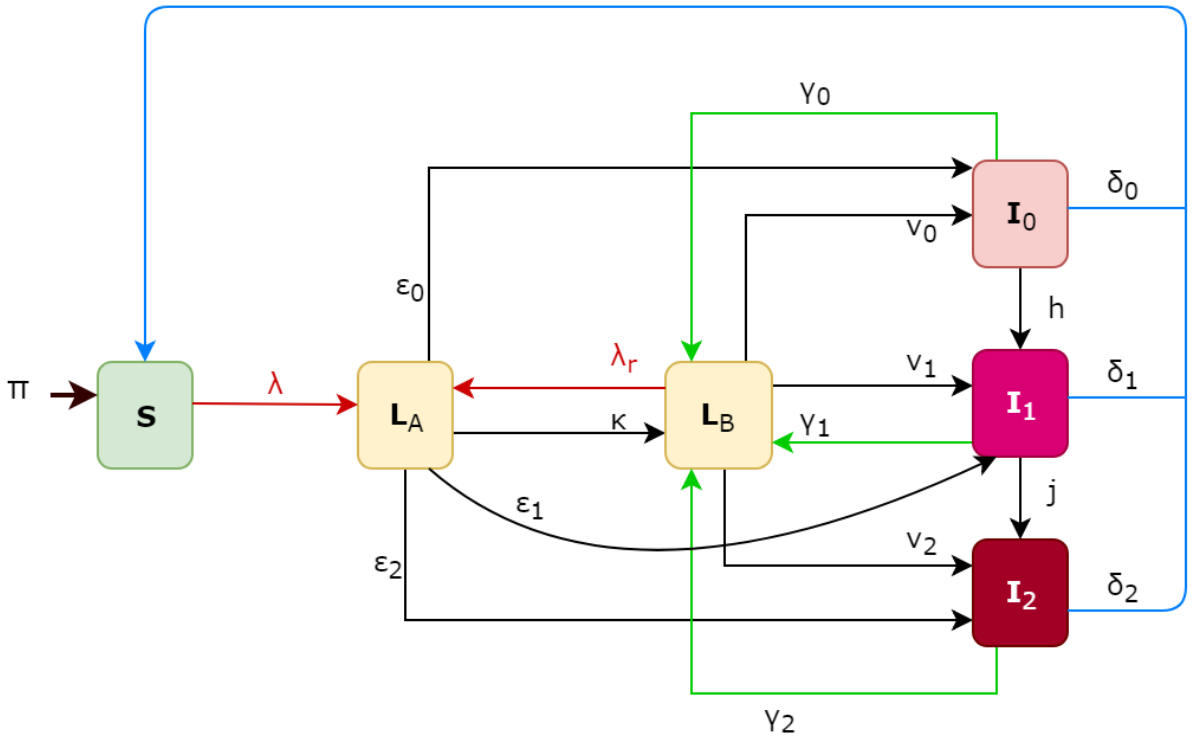


Fig 1: Transmission dynamic model of TB with heterogeneous infectiousness. Boxes represent compartments in the model and arrows represent flows. Not shown here but included in the model are natural mortality from each compartment and TB-related mortality from each active TB compartment (I_0, I_1 and I_2). Red arrows represent transmission flows; black arrows represent disease progression processes; green arrows represent self-recovery, blue arrows represent detection and treatment cure.

Estimation of the infectiousness parameters using empiric evidence

The four model parameters that capture heterogeneous infectiousness in the model were the proportions of non-spreaders, low-spreaders and super-spreaders, and the relative infectiousness of super-spreaders compared to low-spreaders. In order to estimate these parameters, we first determined the statistical distribution of the number of secondary infections per index associated with the stochastic equivalent (continuous-time Markov) model of our ODE-based model. We obtained a mixture of three classic statistical distributions: two geometric distributions for low-spreaders and super-spreaders and one Dirac delta distribution for non-spreaders. A detailed demonstration of this result is provided in the ***Supplementary Material***.

A Hamiltonian Monte Carlo (HMC) algorithm was used to generate 20,000 samples from the posterior distributions of the parameters. The 20,000 samples used for inference were obtained from 40,000 iterations of the HMC, burning the first 20,000 draws. Burn-in size and convergence were assessed by inspecting the trace plots of the estimated parameters. The reported 95% credible intervals were obtained by computing the 2.5^{th} and 97.5^{th} percentiles of the parameters' posterior distributions. The parameter estimates from the Bayesian process is provided in **Supplementary fig.1**.

The resulting statistical distribution was then compared in a Bayesian context with empiric data obtained from our previous analysis of TB contact tracing data in the Australian state of Victoria [9]. A comparison between the distribution of the number of secondary infections per index simulated by our calibrated model with the one observed empirically is presented in Fig 2.

The parameter values other than those pertaining to infectiousness introduced above were estimated from a review of relevant evidence, with the relevant sources presented in **Table 1**.

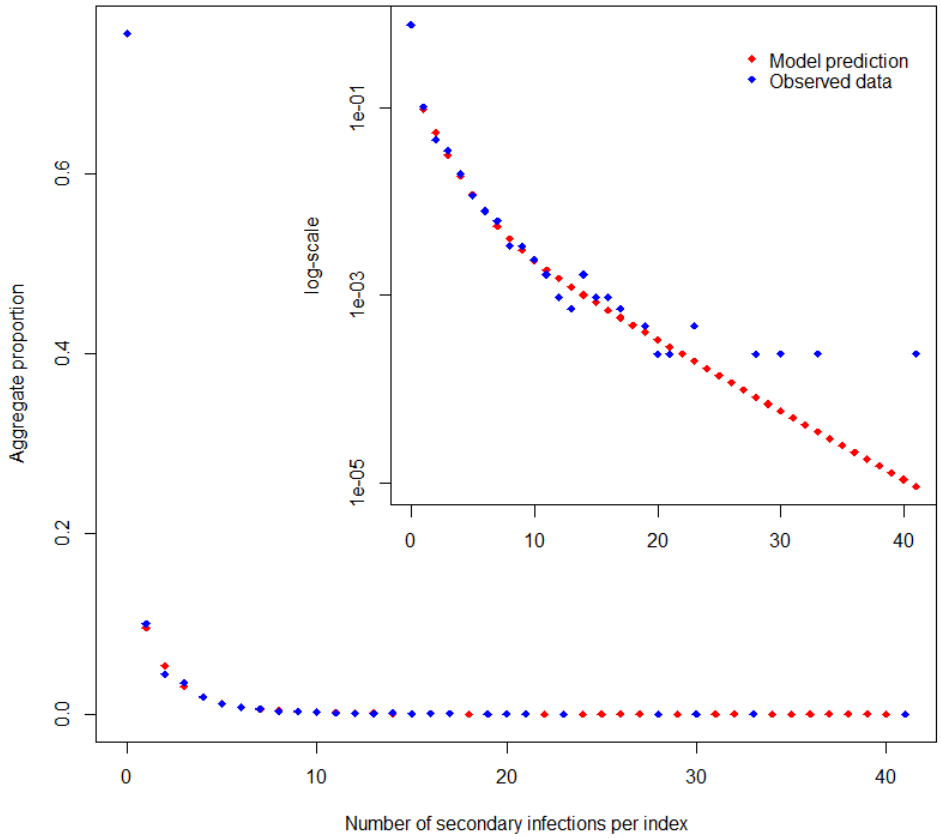


Fig 2. Comparison of the effective modelled secondary infection distribution with empiric data from Victoria, Australia (red points represent the modelled distribution while blue points are observed empiric data).

Model calibration

For the baseline equilibrium state, the ODE model was calibrated to a high TB burden settings with an incidence rate of 200 per 100,000 population per year by adjusting the transmission rate parameter for low-spreaders.

The model was programmed in R 3.4.0 (R Project for Statistical Computing), and we used a package called 'deSolve', solver for initial value problems of ODE [38, 39]. The Bayesian parameter estimation was done using the package 'rstan' (v.2.18.2) [40] and the source code for the model is publicly available from https://github.com/Yayehirad/TB_heterogeneity

Table 1: Model parameter values

Corresponding model parameter	Symbol	Base model value	Uncertainty range	Source
Natural mortality rate, per year	μ	0.0154	0.0133 - 0.0182	[41]
Rate of early progression, per year	ε	0.401	0.307 - 0.548	[42]
Rate of progression to late latency after infection, per year	κ	3.599	3.452- 3.693	[42]
Rate of late progression, per year	ν	0.002	0.0009 - 0.004	[42]
Average duration of infectiousness, years	D	3	2-4	[43]
TB-related mortality among Super-spreaders, proportion	μ_{I2}	0.534	0.327 - 0.78	[43, 44]
TB-case fatality among low-spreaders, proportion	μ_{I1}	0.077	0.077 - 0.096	
TB-case fatality among non-spreaders, proportion	μ_{I0}	0.077	0.077 - 0.096	
Conversion from non-infectious to low-spreaders, proportion	h	0.014	0 - 0.06	[32, 45]
Conversion from low-spreaders to super-spreaders, proportion	j	0.014	0 - 0.06	Assumed
Relative reduction in force of re-infection compared to new infection	r	0.21	0.15 - 0.4	[46]
Parameters calibrated from data				
Proportion of infectious among all TB		0.42	0.39 - 0.46	Estimated from data [9]
Proportion of in super-spreaders among infectious		0.18	0.1 - 0.29	Estimated from data [9]
Transmission rate per low-spreader per-year	β1	12	10 - 15	Calibrated
Relative infectiousness of super-spreaders compared to low-spreaders	m	4.67	3.69 - 5.88	Estimated from data [9]
Intervention parameters				
Baseline case detection rate, proportion	CDR	0.65	0.5 - 0.75	Assumed
Intervention (proportion of missed cases detected)	q	0.2	0 - 0.4	Assumed
Proportion of intervention targeted to super-spreaders	d	0.5	0 - 1.0	Assumed

Sensitivity analysis

We first performed a one-way sensitivity analysis on each parameter to understand the sensitivity of equilibrium incidence to parameter ranges in the absence of interventions. We then undertook multidimensional sensitivity analyses to assess the correlation between equilibrium incidence and each parameter by taking 1000 samples from each parameter ranges with Latin hypercube sampling (LHS) technique [47, 48].

Intervention simulations

We simulated an intervention intended to represent active case finding with a point-of-care diagnostic tool, accelerating successful diagnosis and treatment of individuals as early as possible before they start transmitting the infection. This could be conceptually considered as implemented by a highly sensitive point-of-care diagnostic technology following community awareness campaigns that target individuals with noticeable symptoms in order to increase their rate of presentation [49]. To target the intervention, clinical symptoms and behavioural characteristics can be used to identify those individuals who are more likely to be super-spreaders.

We implemented an intervention capable of finding q proportion of TB cases that would otherwise remain undetected by the passive case detection process:

i.e. $CDR_{intervention} = q \times (1 - CDR_{baseline}) + CDR_{baseline}$, where, CDR is the case detection rate, being the proportion of all cases detected during their disease episode and q is the proportion of cases detected under the intervention scenario that would otherwise be missed by the baseline CDR. We set the maximum attainable level of CDR in all intervention scenarios including targeting super-spreaders at 90%. In targeting the intervention towards super-spreaders, the maximum possible CDR targeted to super-spreaders under the intervention occurs when the CDR for non-spreaders and low-spreaders is kept at the baseline. With this restriction, we defined a single parameter (d) to represent the extent of intervention targeting relative to the maximum amount of targeting possible, which ranges from zero to one, with zero representing untargeted intervention and one representing 100% targeting. That is, the CDR targeted to super-spreaders under the intervention scenario was $CDR_{intervention}$ plus d multiplied by the difference between the maximum possible CDR and $CDR_{intervention}$. This implementation ensures equivalent effort or the number of people

treated for untargeted and targeted interventions; details of these calculations are given in the ***Supplementary Material***.

We then evaluated the effectiveness of targeting super-spreaders ($d > 0$) compared to untargeted intervention ($d = 0$) in four independent scenarios. These four scenarios were settings with high transmission and high case detection, high transmission and low case detection, low transmission and high case detection, and low transmission and low case detection. These four scenarios were simulated by calibrating the transmission parameter for low-spreaders and CDR. In addition, we evaluated the impact of levels of targeting (d) for different coverage levels of the intervention (q) in these four setting scenarios, while other model parameters were sampled from the same plausible ranges as in the baseline sensitivity analysis using LHS.

6.5. Results

Baseline sensitivity analyses

The sensitivity analyses show that the most marked impact on equilibrium incidence arose from the rate of late TB progression followed by the CDR **(Fig 3).** The proportion of super-spreaders among persons with infectious TB and the relative infectiousness of super-spreaders compared to low-spreaders also had a substantial impact. **Supplementary Fig 2** presents a multidimensional sensitivity analysis, using the LHS method to sample 1000 parameter sets, which was consistent with the results of the one-way sensitivity analysis.

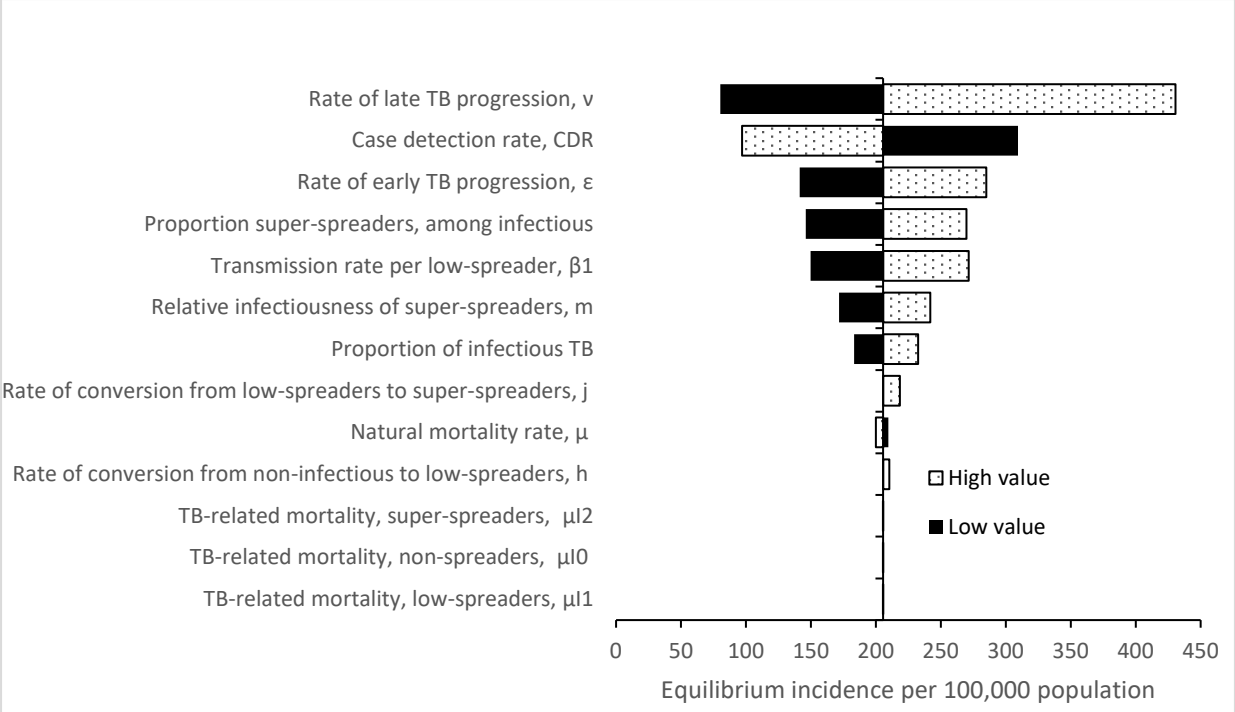


Fig 3: One-way sensitivity analyses of equilibrium TB incidence to variation between extremes of plausible parameter values. Maximum and minimum parameter values are given in Table 1. Values on the x-axis represent the equilibrium TB incidence, and the vertical line indicates the equilibrium incidence obtained with the baseline parameter set.

Impact of targeted interventions

Targeting the active case finding intervention towards super-spreaders was more effective than mass intervention in all setting scenarios. However, its effectiveness was particularly marked in settings with low case detection, regardless of transmission rates. For example, in the first scenario of both high transmission and high CDR, a 20% untargeted active case finding intervention led to a 22.8% reduction in incidence over 20 years, while with an 80% targeting an equal active case finding intervention resulted in a 35.9% reduction in TB incidence over the same period. However, in the scenario of high transmission but low CDR, the targeted intervention is much more effective (achieving a 42.2% reduction) than the untargeted intervention, which only reduced incidence by 22.8% over 20 years, considering a

20% active case finding and an 80% level of targeting. Interventions in settings of low transmission and high case detection rates had a relatively minor impact on the burden (***Supplementary Fig. 3***).

We evaluated the impact of intensifying active case finding (q) from zero to 40% detection and varying the proportion of this intervention targeting the super-spreaders (d). As shown in **Fig 4,** both these quantities are essential – i.e. the increase in CDR and the extent of targeting provided the increase in CDR is more than negligible. The most significant impact of targeting was in the setting of high transmission and low CDR at all coverage levels of the intervention. Targeting super-spreaders was more effective at all levels of relative CDR increases in settings with high transmission and low case detection (Panel B), while in settings with high transmission and high case detection, targeting super-spreaders has a significant effect if the relative increase in CDR is low (Panel A).

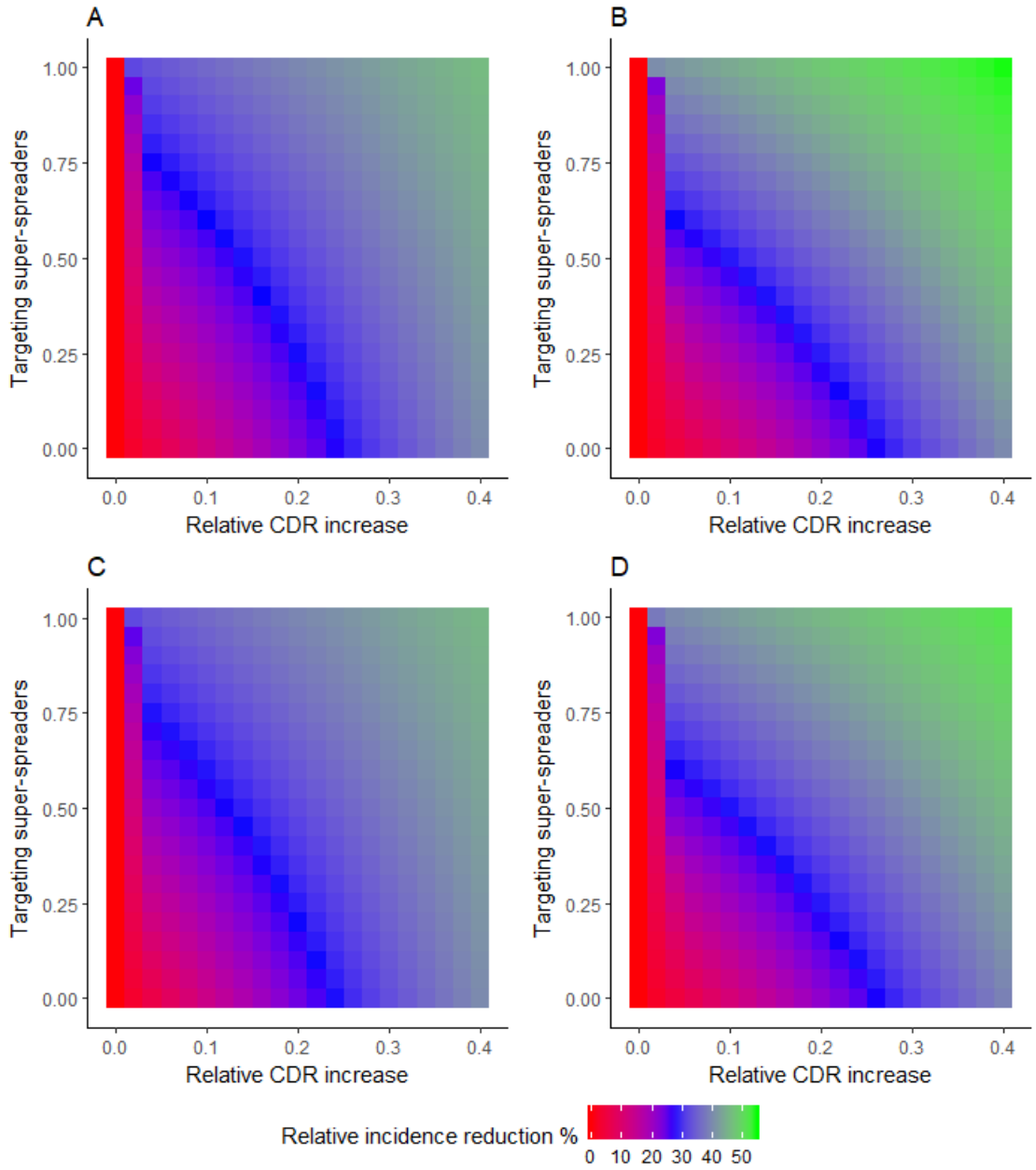


Fig 4: Comparison of extent of active case finding and the proportion of targeting to super-spreaders on the 20-year projected relative reduction in TB incidence under the four scenarios. A) High transmission and high case detection setting; B) high transmission and low case detection; C) low transmission and high case detection; D) low transmission and low case detection.

Fig 5 presents the results of the one-way sensitivity analysis performed to observe the impact of high and low extremes of parameter values on the intervention's 20-year projected relative

incidence reduction. The sensitivity analyses show that the most marked impact on intervention effectiveness arose from the rate of late progression, followed by the rate of early progression and the proportion of super-spreaders. Notably, while higher rates of late progression were correlated with lower effectiveness of the active case finding intervention, higher rates of early progression resulted in greater reductions in the incidence.

Supplementary fig 4 presents the results of the multidimensional sensitivity analyses assessing the impact of parameter ranges on intervention effectiveness (relative reduction in annual incidence with 20-years projections). This multidimensional analysis supports the results of the one-way sensitivity analysis: that the proportion of super-spreaders had a considerable impact on intervention effectiveness, and was the most important parameter after the disease progression parameters.

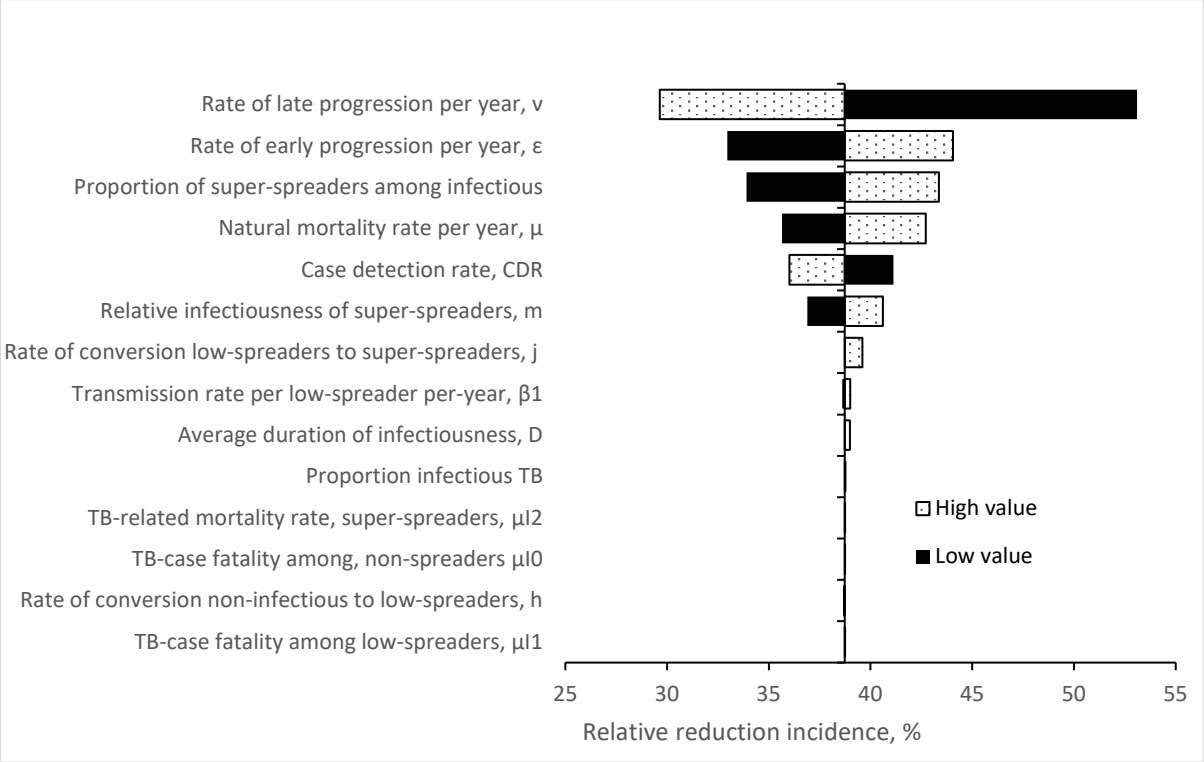


Fig 5: One-way sensitivity analyses of 20-year projected relative reduction in incidence to variation between extremes of plausible parameter values. The maximum and minimum parameter values are given in Table 1. Values on the x-axis represent the relative reduction in

incidence following the implementation of a 20% active case finding intervention with 80% targeting. The vertical line indicates the relative reduction in incidence obtained with the baseline parameter set.

6.6. Discussion

Our model suggested that targeted active case finding interventions directed toward people likely to be super-spreaders could have a substantial impact on the TB burden, particularly in settings with high transmission but low case detection rates. This model also showed that parameters related to infectiousness heterogeneity such as the proportion of super-spreaders and their relative infectiousness compared to other low-spreaders were as crucial as other previously recognised epidemiological parameters, such as the latency progression parameters, in determining the burden of TB. The current model is also able to incorporate a range of commonly used approaches to capturing heterogeneous infectiousness of active TB cases. The choice of three infectiousness levels for active TB and the values assigned to each compartment was based on empirical data estimates, strengthening the validity of our findings.

Active case finding has been described as “turning off the tap” in TB control intervention strategies since it represents a method of identifying individuals with TB and promptly initiating treatment to avoid further onward transmissions [50]. Previous TB modelling has suggested that active case finding with highly sensitive diagnostic tools can reduce delay to treatment and so significantly reduce TB transmission [51]. The use of point-of-care diagnostic technologies with high sensitivity to detect persons with TB soon after they begin to transmit, such as GeneXpert Omni [49], could enable the implementation of active case finding interventions similar to that considered in this study. In fact, any TB active case finding intervention necessarily incorporates some level of targeting towards the highly infectious individuals, because highly infectious patients’ disease characteristics (such as

higher sputum-bacilli concentration and lung-cavitation) make them more easily detectable with existing microbiological and radiological diagnostic tools. This means that implementing more sensitive diagnostics may paradoxically decrease targeting of super-spreaders – whereas traditional smear-based interventions might ensure that resources are targeted to the most infectious. Although it is difficult to determine which patients are super-spreaders a priori, previous studies have characterised highly infectious TB patients and identified some of the clinical, demographic and behavioural predictors that could be used in community awareness and detection campaigns [5, 9]. In agreement with previous modelling [52], our model suggested that active case finding interventions are particularly valuable in high transmission and low case detection settings to limit onward transmission. In the implementation of the intervention, although we did not perform cost-effectiveness analyses, our study shows that finding super-spreaders (less than 10% of all TB cases) is equally as effective epidemiologically as finding 40% of general TB population. In addition, the cost of identifying and treating super-spreaders that only comprises less than 10% of total TB cases can be compared with the cost of treating all active TB cases in a population. Our findings concerning the effectiveness of interventions can provide broad directions to future TB modellers and policymakers, although predictions from our model are not intended to provide location-specific estimates.

Our sensitivity analyses showed that a higher late reactivation rate had a negative impact on the effectiveness of active case finding interventions, while higher rates of early progression increased the impact on the 20-year projected reduction of incidence following the intervention. This suggests that active case finding interventions are more effective if the TB epidemic is more dependent on early progression and intense recent transmission than late TB reactivation since the intervention would have a more significant effect on rapidly reducing transmission. A recent study showed that variation in the latency progression

parameters had important impacts on model predictions around the effectiveness of preventive therapy interventions [53], while our analysis complements this finding by showing that these parameter variations also have indispensable effects on predictions regarding the effectiveness of case-finding interventions.

In our analyses, the rate of conversion from low-spreader to super-spreader had little significance, and impact of the rate of conversion from non-infectious to low-spreaders was negligible. It is known that active TB cases' ability to spread *Mtb* may increase as they progress clinically to more severe disease, e.g. from smear-negative to smear-positive or from non-cavitary to cavitary-TB [54]. In previous TB models, spontaneous conversion from less infectious to more infectious states such as smear-negative to smear-positive was included with rates around 1.5% per year [18, 25, 32, 45, 55-57]. These parameters do not affect model outputs substantially, such that they can be omitted for model simplicity if the actual values are consistent with what has previously been assumed. Nevertheless, the exact values of these quantities remain uncertain, and future research to refine these quantities may modify this conclusion.

The current model structure enhances flexibility around the assumption of infectiousness heterogeneity that allowed for reflection of a broad range of factors that can alter active TB cases infectivity level [5, 58]. Thus, the model is able to incorporate many previous compartmental modelling structural approaches, including those that stratified active TB cases' infectivity into two levels, as non-infectious and infectious [13-17], or model structures that incorporate three levels of infectivity, as non-infectious, smear-negative infectious and smear-positive infectious [30, 31, 59]. However, in the application of this model to a particular epidemiological setting, data to estimate the proportion of super-spreaders is essential, since this information has a particularly marked impact on both baseline disease burden and intervention effectiveness.

We used empirical data to parameterise our model, but a noteworthy limitation is that these results may not be generalizable to other settings. However, we are not aware of any past work that has used empiric data to inform model parameters on the proportions of infectious TB and super-spreaders at all. As with any model-based analyses, our study has limitations that arise from its assumptions. In addition to those introduced above, our model is not intended to represent a specific setting, and as such does not incorporate stratification by age, HIV or multidrug-resistant TB. The other limitation is that we used a very simplified implementation of an active case finding intervention with a theoretical point-of-care diagnostic tool, without considering all the complexities in TB diagnosis and treatment implementation.

6.7. Conclusions

The approaches we used to inform model parameters related to infectiousness heterogeneity from the observed distribution of secondary infections can be useful for future modelling studies of TB, in particular, and other infectious diseases, in general. The principles of implementation could be used in future TB modelling studies that represent specific epidemiological settings, especially when TB contact investigation data are available to estimate the proportion of super-spreaders. In the usage of advanced point-of-care technologies, targeting active case-finding interventions to super-spreaders is likely to be especially beneficial in low CDR settings.

Acknowledgements

The authors would like to thank Monash University for providing the PhD scholarship to YAM.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Availability of data and materials

The R-code used to analyse this model is publicly available.

Authors' contributions

YAM, RR and JMT conceived the study. YAM developed structure and ODE of the model, which was reviewed by JMT and RR. ESM, ACA, JMT conceived the strategy of the intervention in the model. YAM coded the model in R that RR and JMT reviewed and evaluated. YAM drafted the manuscript, and all authors provided input into revisions and approved the final draft for submission.

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Competing interests

The authors declare that they have no competing interests.

Supplementary material

Supplementary Material. doc

6.8. References

1. WHO, *Global Tuberculosis Report 2019*. 2019: Geneva.
2. WHO, *The End TB Strategy: Global strategy and targets for tuberculosis prevention, care and control after 2015*. 2014: Geneva, Switzerland

3. Herbert, N., et al., *World TB Day 2014: finding the missing 3 million.* The Lancet, 2014. **383**(9922): p. 1016-1018.
4. Zumla, A., et al., *The WHO 2014 global tuberculosis report—further to go.* The Lancet Global Health, 2015. **3**(1): p. e10-e12.
5. Melsew, Y., et al., *Risk factors for infectiousness of patients with tuberculosis: a systematic review and meta-analysis.* Epidemiology & Infection, 2018: p. 1-9.
6. Turner, R.D., et al., *Tuberculosis infectiousness and host susceptibility.* The Journal of infectious diseases, 2017. **216**(suppl_6): p. S636-S643.
7. Issarow, C.M., N. Mulder, and R. Wood, *Environmental and social factors impacting on epidemic and endemic tuberculosis: a modelling analysis.* Royal Society open science, 2018. **5**(1): p. 170726.
8. Trauer, J.M., et al., *The Importance of Heterogeneity to the Epidemiology of Tuberculosis.* Clinical Infectious Diseases, 2018. **69**(1): p. 159-66.
9. Melsew, Y.A., et al., *The role of super-spreading events in Mycobacterium tuberculosis transmission: evidence from contact tracing.* BMC Infectious Diseases, 2019. **19**(1): p. 244.
10. Lloyd-Smith, J.O., S.J. Schreiber, and W.M. Getz. *Moving beyond averages: Individual-level variation in*. in *Mathematical Studies on Human Disease Dynamics: Emerging Paradigms and Challenges: AMS-IMS-SIAM Joint Summer Research Conference on Modeling the Dynamics of Human Diseases: Emerging Paradigms and Challenges, July 17-21, 2005, Snowbird, Utah*. 2006. American Mathematical Soc.
11. Ypma, R.J., et al., *A sign of superspreading in tuberculosis: highly skewed distribution of genotypic cluster sizes.* Epidemiology, 2013. **24**(3): p. 395-400.
12. Lloyd-Smith, J.O., et al., *Superspreading and the effect of individual variation on disease emergence.* Nature, 2005. **438**(7066): p. 355-359.
13. Aparicio, J.P. and C. Castillo-Chavez, *Mathematical modelling of tuberculosis epidemics.* Mathematical Biosciences & Engineering: MBE, 2009. **6**(2): p. 209-37.
14. Hickson, R., G. Mercer, and K. Lokuge. *Sensitivity analysis of a model for tuberculosis*. in *19th international congress on modelling and simulation*. 2011.
15. Hickson, R.I., G.N. Mercer, and K.M. Lokuge, *A metapopulation model of tuberculosis transmission with a case study from high to low burden areas.* PloS one, 2012. **7**(4): p. e34411.
16. Korthals Altes, H., et al., *Latent tuberculosis infection in foreign-born communities: Import vs. transmission in The Netherlands derived through mathematical modelling.[Erratum appears in PLoS One. 2018 May 24;13(5):e0198376; PMID: 29795693]*. PLoS ONE [Electronic Resource], 2018. **13**(2): p. e0192282.
17. Thomas, E.G., et al., *Modelling the spread of tuberculosis, including drug resistance and HIV: a case study in Papua New Guinea's Western Province.* The ANZIAM Journal, 2010. **52**(1): p. 26-45.
18. Dowdy, D.W., et al., *Impact of enhanced tuberculosis diagnosis in South Africa: a mathematical model of expanded culture and drug susceptibility testing.* Proceedings of the National Academy of Sciences, 2008.
19. Dowdy, D.W., et al., *The potential impact of enhanced diagnostic techniques for tuberculosis driven by HIV: a mathematical model.* AIDS, 2006. **20**(5): p. 751-62.
20. Dowdy, D.W., et al., *Population-level impact of same-day microscopy and Xpert MTB/RIF for tuberculosis diagnosis in Africa.* PLoS One, 2013. **8**(8): p. e70485.
21. Lin, H.-H., et al., *The impact of new tuberculosis diagnostics on transmission: why context matters.* Bulletin of the World Health Organization, 2012. **90**: p. 739-747.
22. Oxlade, O., et al., *Developing a tuberculosis transmission model that accounts for changes in population health.* Medical Decision Making, 2011. **31**(1): p. 53-68.

23. Pandey, S., et al., *Estimating tuberculosis incidence from primary survey data: a mathematical modeling approach.* The International Journal of Tuberculosis and Lung Disease, 2017. **21**(4): p. 366-374.
24. Sun, A.Y., et al., *Modeling the impact of alternative strategies for rapid molecular diagnosis of tuberculosis in Southeast Asia.* American journal of epidemiology, 2013. **178**(12): p. s1740-1749.
25. Houben, R., et al., *TIME Impact–a new user-friendly tuberculosis (TB) model to inform TB policy decisions.* BMC medicine, 2016. **14**(1): p. 56.
26. Menzies, N.A., et al., *Prospects for Tuberculosis Elimination in the United States: Results of a Transmission Dynamic Model.* American journal of epidemiology, 2018.
27. Menzies, N.A., et al., *Population health impact and cost-effectiveness of tuberculosis diagnosis with Xpert MTB/RIF: a dynamic simulation and economic evaluation.* PLoS medicine, 2012. **9**(11): p. e1001347.
28. Garcia, A., J. Maccario, and S. Richardson, *Modelling the annual risk of tuberculosis infection.* International Journal of Epidemiology, 1997. **26**(1): p. 190-203.
29. Dowdy, D.W., et al., *A user-friendly, open-source tool to project impact and cost of diagnostic tests for tuberculosis.* Elife, 2014. **3**: p. e02565.
30. Arregui, S., et al., *Data-driven model for the assessment of Mycobacterium tuberculosis transmission in evolving demographic structures.* Proceedings of the National Academy of Sciences of the United States of America, 2018. **115**(14): p. E3238-E3245.
31. Dowdy, D.W., et al., *Population-level impact of active tuberculosis case finding in an Asian megacity.* PLoS One, 2013. **8**(10): p. e77517.
32. Salomon, J.A., et al., *Prospects for advancing tuberculosis control efforts through novel therapies.* PLoS Medicine, 2006. **3**(8): p. e273.
33. Vynnycky, E., et al., *Tuberculosis control in South African gold mines: mathematical modeling of a trial of community-wide isoniazid preventive therapy.* American journal of epidemiology, 2015. **181**(8): p. 619-632.
34. McCreesh, N. and R.G. White, *An explanation for the low proportion of tuberculosis that results from transmission between household and known social contacts.* Scientific reports, 2018. **8**(1): p. 5382.
35. Melsew, Y.A., et al., *Profiles of tuberculosis disease activation among contacts of patients with tuberculosis.* European Respiratory Journal, 2019. **54**(4): p. 1900353.
36. Melsew, Y.A., et al., *Heterogeneous infectiousness in mathematical models of tuberculosis: A systematic review.* Epidemics, 2019: p. 100374.
37. Moyo, N., E. Tay, and J. Denholm, *Evaluation of tuberculin skin testing in tuberculosis contacts in Victoria, Australia, 2005–2013.* Public health action, 2015. **5**(3): p. 188-193.
38. RCoreTeam, *R: A Language and Environment for Statistical Computing*. 2014, R Foundation for Statistical Computing: Vienna, Austria.
39. Soetaert, K., T. Petzoldt, and R.W. Setzer, *Solving differential equations in R: package deSolve.* Journal of Statistical Software, 2010. **33**.
40. Team, S.D., *RStan: the R interface to Stan, Version 2.18.2.* 2018.
41. WHO, *ETHIOPIA:FactsheetsofHealthStatistics 2016*. 2016.
42. Ragonnet, R., et al., *Optimally capturing latency dynamics in models of tuberculosis transmission.* Epidemics, 2017.
43. Tiemersma, E.W., et al., *Natural history of tuberculosis: duration and fatality of untreated pulmonary tuberculosis in HIV negative patients: a systematic review.* PloS one, 2011. **6**(4): p. e17601.

44. Trauer, J.M., J.T. Denholm, and E.S. McBryde, *Construction of a mathematical model for tuberculosis transmission in highly endemic regions of the Asia-Pacific.* Journal of theoretical biology, 2014. **358**: p. 74-84.
45. Hughes, G.R., C.S. Currie, and E.L. Corbett. *Modeling tuberculosis in areas of high HIV prevalence*. in *Simulation Conference, 2006. WSC 06. Proceedings of the Winter*. 2006. IEEE.
46. Andrews, J.R., et al., *Risk of progression to active tuberculosis following reinfection with Mycobacterium tuberculosis.* Clinical infectious diseases, 2012. **54**(6): p. 784-791.
47. Drake, J.M. and P. Rohani *Sensitivity analysis of deterministic models through Latin hypercube sampling: A model for the spread of Ebola virus disease*. 2015.
48. Gomero, B., *Latin Hypercube Sampling and Partial Rank Correlation Coefficient Analysis Applied to an Optimal Control Problem*. 2012, University of Tennessee.
49. FIND, C.a. *World's most portable molecular diagnostics system unveiled at AACC.* 2015 [cited 2019 May 14]; Available from: https://www.prnewswire.com/news-releases/worlds-most-portable-molecular-diagnostics-system-unveiled-at-aacc-300119213.html.
50. Yuen, C.M., et al., *Turning off the tap: stopping tuberculosis transmission through active case-finding and prompt effective treatment.* The Lancet, 2015. **386**(10010): p. 2334-2343.
51. Uys, P.W., et al., *Potential of rapid diagnosis for controlling drug-susceptible and drug-resistant tuberculosis in communities where Mycobacterium tuberculosis infections are highly prevalent.* Journal of clinical microbiology, 2009. **47**(5): p. 1484-1490.
52. Dowdy, D.W., S. Basu, and J.R. Andrews, *Is passive diagnosis enough? The impact of subclinical disease on diagnostic strategies for tuberculosis.* American journal of respiratory and critical care medicine, 2013. **187**(5): p. 543-551.
53. Sumner, T. and R.G. White, *Modelling the impact of tuberculosis preventive therapy: the importance of disease progression assumptions.* bioRxiv, 2019: p. 666669.
54. Pai, M., et al., *Tuberculosis.* Nature Reviews Disease Primers, 2016. **2**: p. 16076.
55. Knight, G.M., et al., *Impact and cost-effectiveness of new tuberculosis vaccines in low-and middle-income countries.* Proceedings of the National Academy of Sciences, 2014. **111**(43): p. 15520-15525.
56. Legrand, J., et al., *Modeling the impact of tuberculosis control strategies in highly endemic overcrowded prisons.* PLoS One, 2008. **3**(5): p. e2100.
57. Mellor, G.R., C.S. Currie, and E.L. Corbett, *Incorporating household structure into a discrete-event simulation model of tuberculosis and HIV.* ACM Transactions on Modeling and Computer Simulation (TOMACS), 2011. **21**(4): p. 26.
58. Grassly, N.C. and C. Fraser, *Mathematical models of infectious disease transmission.* Nature Reviews Microbiology, 2008. **6**(6).
59. Abu-Raddad, L.J., et al., *Epidemiological benefits of more-effective tuberculosis vaccines, drugs, and diagnostics.* Proceedings of the National Academy of Sciences, 2009. **106**(33): p. 13980-13985.

6.9. Supplementary material

Model Equations

$$\frac{dS}{dt} = \pi . N + \ \delta_0 . I_0 + \ \delta_1 . I_1 + \delta_2 . I_2 \ - (\lambda + \ \mu) . S$$

$$\frac{dL_A}{dt} = \lambda . S + \lambda_r . L_B - (\varepsilon_0 + \varepsilon_1 + \varepsilon_2 + \ \mu + \ \kappa) . L_A$$

$$\frac{dL_B}{dt} = \ \kappa . L_a \ + \ \gamma_0 I_0 + \gamma_1 I_1 + \gamma_2 I_2 \ - (\nu_0 + \nu_1 + \nu_2 + \ \lambda_r + \mu) . L_B$$

$$\frac{dI_0}{dt} = \ \varepsilon_0 . L_A + \ \nu_0 . L_B - \ (\delta_0 + \mu_{i0} + \mu + \gamma_0 + h) . I_0$$

$$\frac{dI_1}{dt} = \ \varepsilon_1 . L_A + \nu_1 . L_B + h . I_0 - \ (\delta_1 + \mu_{i1} + \mu + \gamma_1 + j) . I_1$$

$$\frac{dI_2}{dt} = \ \varepsilon_2 . L_A + + \nu_2 . L_B + j . I_1 - \ (\delta_2 + \mu_{i2} + \mu + \gamma_2) . I_2$$

$$\beta_2 = m \times \beta_1 \ , \ \ \lambda = \frac{\beta_1 I_1 + \beta_2 I_2}{N}, \ \lambda_r = r \times \lambda$$

Where, N is the total population, β_2 is the transmission rate per super-spreader per year, β_1 is the transmission rate per low-spreader. The rate of treatment commencement is determined from the proportion of all cases detected during their disease episode. This proportion is reported from programmatic performance as the "case detection rate" (CDR). We calculated the rate at which patients are detected and start treatment during their infectious period (δ) from the CDR. The proportion of cases assumed to be detected and treated from an active TB compartment is found from $CDR = \frac{\delta}{\delta + \gamma + \mu_i + \mu}$. Solving for δ, $\delta = \left(\frac{\mathrm{CDR}(\gamma + \mu_i + \mu)}{(1 - \mathrm{CDR})}\right)$

The full definition of parameters is provided in Table 1

Estimation of the parameters defining infectiousness heterogeneity

In this section, we demonstrate that the distribution of the number of secondary infections associated with each of our infectious compartments and assuming equivalent periods of infectiousness follows a geometric distribution. The time (T) spent in the infectious compartment follows an exponential distribution parameterised with the rate of leaving the infectious compartment (γ): $T \sim Exp(\gamma)$.

Let us denote X the number of secondary infections. As transmission events occur at a constant rate β, the number of secondary infections generated during a fixed duration $t \geq 0$ follows a Poisson distribution with parameter βt. That is, $(X|T = t) \sim Poisson(\beta t)$ and therefore, $\forall\, v \in \{0,1,2,3 \ldots\ldots\}$,

$$P(X = v|\mathrm{T} = \mathrm{t}) = (\beta t)^v \frac{e^{-\beta t}}{v!} .$$

Using the law of total probability for continuous random variables, the probability of observing a given number of offspring (v) can be obtained by:

$P(X = v) = \int_0^\infty P(X = v|\mathrm{T} = \mathrm{t}).f_T(t).dt$, where, $f_T(t)$ is the probability density function of T (exponentially distributed).

By substituting $P(X = v|\mathrm{T} = \mathrm{t})$ and $f_T(t)$, we obtain:

$$P(X = v) = \int_0^\infty (\beta t)^v \frac{e^{-\beta t}}{v!}.\gamma e^{-\gamma t}.dt$$

$$P(X = v) = \frac{\beta^v}{v!}\gamma \int_0^\infty t^v e^{-(\gamma+\beta).t}.dt$$

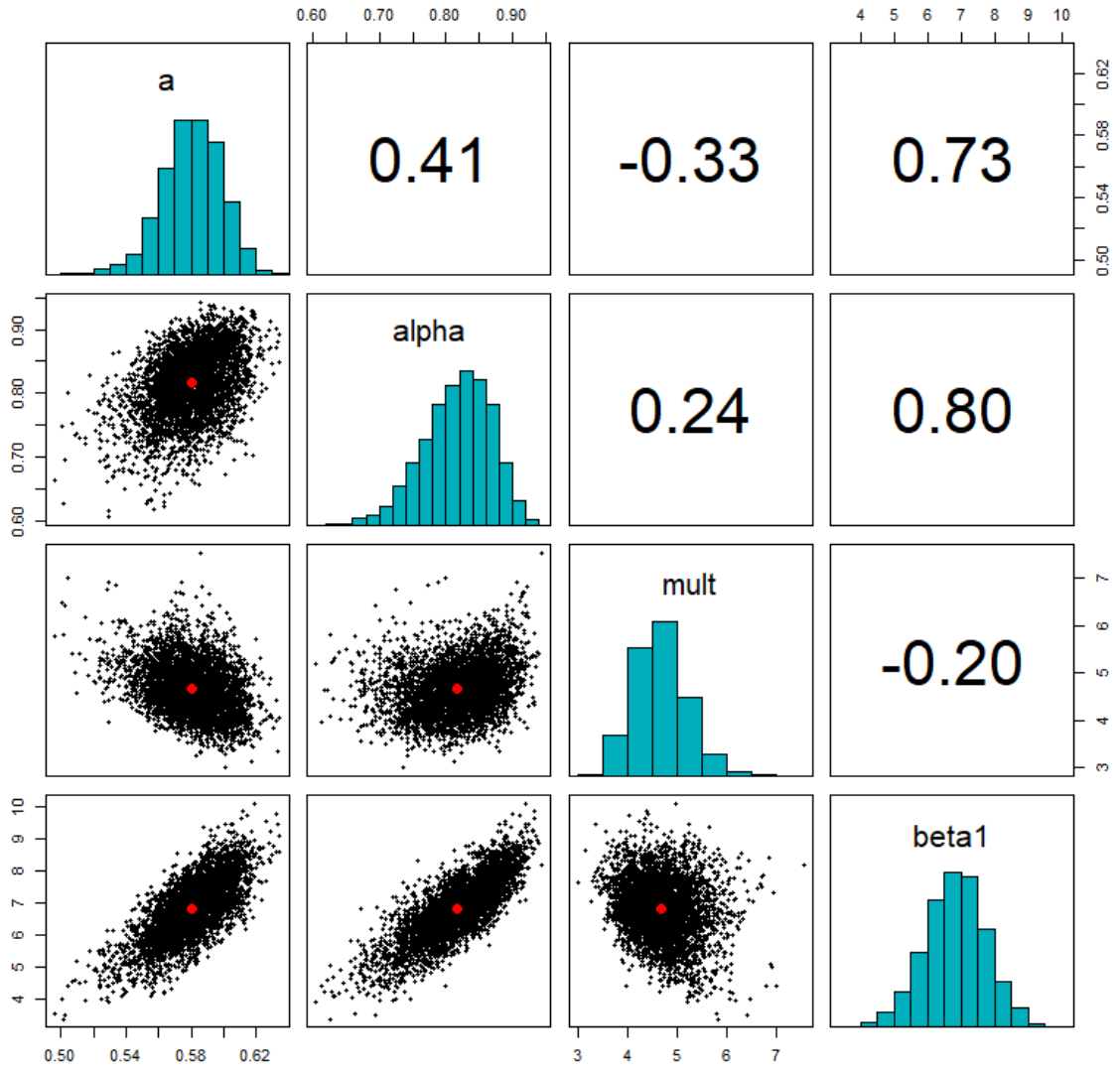
$$P(X = v) = \frac{\beta^v}{v!}\gamma.(\gamma + \beta)^{-(v+1)}\, v!$$

$$P(X = v) = \gamma.\frac{\beta^v}{(\gamma + \beta)^{(v+1)}}$$

$$P(X = v) = \frac{\gamma}{\gamma + \beta}.\left(\frac{\beta}{\gamma + \beta}\right)^v, \qquad let\ p = \frac{\beta}{\gamma + \beta}$$

$$P(X = v) = (1 - p).p^v$$

Thus, the number of secondary infections associated with an infectious compartment follows a geometric distribution with success probability $\frac{\beta}{\gamma+\beta}$.



Supplementary fig. 1 Pairwise combination of heterogeneity parameters estimates with Bayesian technique from Victorian TB contact tracing data (the numbers in the upper panels are Pearson's correlation coefficients, the black dots in the lower panels represent samples, and the red dot is just to show the centre). a is the proportion of non-spreaders; $alpha$ is the proportion of low-spreaders among infectious; $mult$ is the relative infectiousness of super-spreaders relative to low-spreaders; $beta1$ is transmission rate per low-spreader ($beta1$ estimate was not used in our ODE model since it was calibrated separately to a high-incidence setting).

Equations for effective reproductive number R_0

We followed the method used by the previous model to estimate R_o [2], with the overall R_o of the model calculated as the sum R_o of low-spreaders ($R_o^{I_1}$) and R_o of super-spreaders ($R_o^{I_2}$) which is computed as follows:

$R_o^{I_1} = \beta_1 \times D_1 \times p_1$, where β_1 is the effective contact rate of a low-spreader, D_1 is the time spent in the low-spreaders compartment, and p_1 is the probability of an infected person to become low-spreader.

-

- $D_1 = \frac{1}{(j+\mu_{i1}+\mu+\gamma_1)}$
- $p_1 = [\frac{\varepsilon_1}{\varepsilon_0+\varepsilon_2+\varepsilon_1+\kappa+\mu} + \frac{\kappa}{\varepsilon_0+\varepsilon_2+\varepsilon_1+\kappa+\mu} \times \frac{\nu_1}{\nu_0+\nu_1+\nu_2+\mu} + \left(\frac{\varepsilon_0}{\varepsilon_0+\varepsilon_2+\varepsilon_1+\kappa+\mu} + \frac{\kappa}{\varepsilon_0+\varepsilon_2+\varepsilon_1+\kappa+\mu} \times \frac{\nu_0}{\nu_0+\nu_1+\nu_2+\mu}\right) \times \frac{h}{h+\mu_{i0}+\mu+\gamma_0}]$, thus

- $R_o^{I_1} = \beta_1 \times \frac{1}{(j+\mu_{i1}+\mu+\gamma_1)} \times (\frac{\varepsilon_1}{\varepsilon_0+\varepsilon_2+\varepsilon_1+\kappa+\mu} + \frac{\kappa}{\varepsilon_0+\varepsilon_2+\varepsilon_1+\kappa+\mu} \times \frac{\nu_1}{\nu_0+\nu_1+\nu_2+\mu} + \left(\frac{\varepsilon_0}{\varepsilon_0+\varepsilon_2+\varepsilon_1+\kappa+\mu} + \frac{\kappa}{\varepsilon_0+\varepsilon_2+\varepsilon_1+\kappa+\mu} \times \frac{\nu_0}{\nu_0+\nu_1+\nu_2+\mu}\right) \times \frac{h}{h+\mu_{i0}+\mu+\gamma_0})$

Similarly, $R_o^{I_2} = \beta_2 \times D_2 \times p_2$, where β_2 is the effective contact rate of a super-spreader, D_2 is the time spent in the super-spreaders compartment, and p_2 is the probability of an infected person to become super-spreader.

- $D_2 = \frac{1}{(\mu_{i2}+\mu+\gamma_2)}$

- $p_2=\frac{\varepsilon_2}{\varepsilon_0+\varepsilon_2+\varepsilon_1+\kappa+\mu}+\frac{\kappa}{\varepsilon_0+\varepsilon_2+\varepsilon_1+\kappa+\mu}\times\frac{\nu_2}{\nu_0+\nu_1+\nu_2+\mu}+\frac{\varepsilon_0}{\varepsilon_0+\varepsilon_2+\varepsilon_1+\kappa+\mu}\times\frac{h}{h+\mu_{i0}+\mu+\gamma_0}\times\frac{j}{j+\mu_{i1}+\mu+\gamma_1}+\frac{\varepsilon_0}{\varepsilon_0+\varepsilon_2+\varepsilon_1+\kappa+\mu}\times\frac{h}{h+\mu_{i0}+\mu+\gamma_0}\times\frac{j}{j+\mu_{i1}+\mu+\gamma_1}+\frac{\kappa}{\varepsilon_0+\varepsilon_2+\varepsilon_1+\kappa+\mu}\times\frac{\nu_0}{\nu_0+\nu_1+\nu_2+\mu}\times\frac{h}{h+\mu_{i0}+\mu+\gamma_0}\times\frac{j}{j+\mu_{i1}+\mu+\gamma_1}+\frac{\kappa}{\varepsilon_0+\varepsilon_2+\varepsilon_1+\kappa+\mu}\times\frac{\nu_1}{\nu_0+\nu_1+\nu_2+\mu}\times\frac{h}{h+\mu_{i0}+\mu+\gamma_0}\times\frac{j}{j+\mu_{i1}+\mu+\gamma_1}$, thus

- $R_o^{I_2}=\beta_2\times\frac{1}{(\mu_{i2}+\mu+\gamma_2)}\times(\frac{\varepsilon_2}{\varepsilon_0+\varepsilon_2+\varepsilon_1+\kappa+\mu}+\frac{\kappa}{\varepsilon_0+\varepsilon_2+\varepsilon_1+\kappa+\mu}\times\frac{\nu_2}{\nu_0+\nu_1+\nu_2+\mu}+\frac{\varepsilon_0}{\varepsilon_0+\varepsilon_2+\varepsilon_1+\kappa+\mu}\times\frac{h}{h+\mu_{i0}+\mu+\gamma_0}\times\frac{j}{j+\mu_{i1}+\mu+\gamma_1}+\frac{\varepsilon_0}{\varepsilon_0+\varepsilon_2+\varepsilon_1+\kappa+\mu}\times\frac{h}{h+\mu_{i0}+\mu+\gamma_0}\times\frac{j}{j+\mu_{i1}+\mu+\gamma_1}+\frac{\kappa}{\varepsilon_0+\varepsilon_2+\varepsilon_1+\kappa+\mu}\times\frac{\nu_0}{\nu_0+\nu_1+\nu_2+\mu}\times\frac{h}{h+\mu_{i0}+\mu+\gamma_0}\times\frac{j}{j+\mu_{i1}+\mu+\gamma_1}+\frac{\kappa}{\varepsilon_0+\varepsilon_2+\varepsilon_1+\kappa+\mu}\times\frac{\nu_1}{\nu_0+\nu_1+\nu_2+\mu}\times\frac{h}{h+\mu_{i0}+\mu+\gamma_0}\times\frac{j}{j+\mu_{i1}+\mu+\gamma_1})$

- If we assume the conversion rates h and j are zero the formula for Ro can be simplified as follows:

- $R_o^{I_1}=\beta_1\times\frac{1}{(j+\mu_{i1}+\mu+\gamma_1)}\times(\frac{\varepsilon_1}{\varepsilon_0+\varepsilon_2+\varepsilon_1+\kappa+\mu}+\frac{\kappa}{\varepsilon_0+\varepsilon_2+\varepsilon_1+\kappa+\mu}\times\frac{\nu_1}{\nu_0+\nu_1+\nu_2+\mu})$, and

- $R_o^{I_2}=\beta_2\times\frac{1}{(\mu_{i2}+\mu+\gamma_2)}\times(\frac{\varepsilon_2}{\varepsilon_0+\varepsilon_2+\varepsilon_1+\kappa+\mu}+\frac{\kappa}{\varepsilon_0+\varepsilon_2+\varepsilon_1+\kappa+\mu}\times\frac{\nu_2}{\nu_0+\nu_1+\nu_2+\mu})$, thus
- $Ro = R_o^{I_1}+R_o^{I_2}$

Intervention implementation equations

$CDR_{intervention}=CDR_{baseline}+q*(1-CDR_{baseline})$,

where $q\ is\ the\ proportion\ of\ missed\ cases\ detected\ by\ the\ intervention, q\epsilon(0,1)$.

We set the maximum possible attainable level of CDR was assumed to 90%, i.e. $CDR_{max}=$ 0.9 in all intervention scenarios.

In order to ensure that the overall coverage of the case detection intervention was the same regardless of targeting, we first defined the overall case detection rate for the intervention ($CDR_{intervention}$) as:

$CDR_{intervention} = a * CDR_{0i} + b * CDR_{1i} + c * CDR_{2i}$, where $a = \frac{I_0}{I_0+I_1+I_2}$, $b = \frac{I_1}{I_0+I_1+I_2}$ and $c = \frac{I_2}{I_0+I_1+I_2}$ are the proportional prevalence of I_0, I_1 and I_2 respectively. Under the intervention scenario, we set the detection rate for I_0 and I_1 to be equal, i.e. $CDR_{0i} = CDR_{1i}$. Thus, CDR_{0i} can be derived from $CDR_{intervention}$ and CDR_{2i} through the formula:

$CDR_{oi} = \frac{CDR_{intervention} - c*CDR_{2i}}{a+b}$, where

CDR_{oi} *is the detection under intervention scenario targeted at* I_0 ,

CDR_{1i} *is the detection under intervention scenario targeted at* I_1,

CDR_{2i} *is targeted at* I_2 and

$CDR_{baseline} \leq CDR_{0i}, CDR_{baseline} \leq CDR_{2i}, 0 \leq CDR \leq 1$. The maximum level of targeting for the intervention (i.e. the maximum value of CDR_{2i} or $CDR2_{imax}$) will occur when CDR_{0i} is unchanged from baseline (i.e. $CDR_{0i} = CDR_{baseline}$) and can be calculated as:

$CDR_{2imax} = \frac{(CDR_{intervention} - (a+b)*CDR_{baseline})}{c}$, however, this value should not be greater than CDR_{max}.

Thus, $CDR_{2imax} = \min(\frac{(CDR_{intervention} - (a+b)*CDR_{baseline})}{c}, CDR_{max})$. We, therefore, define a single parameter (d) representing the extent of intervention targeting relative to the maximum amount of targeting possible, which ranges from zero to one. Therefore the intervention case detection rate for super-spreaders (CDR_{2i}) is given by:

$CDR_{2i} = CDR_{intervention} + (CDR_{2imax} - CDR_{intervention}) * d,$

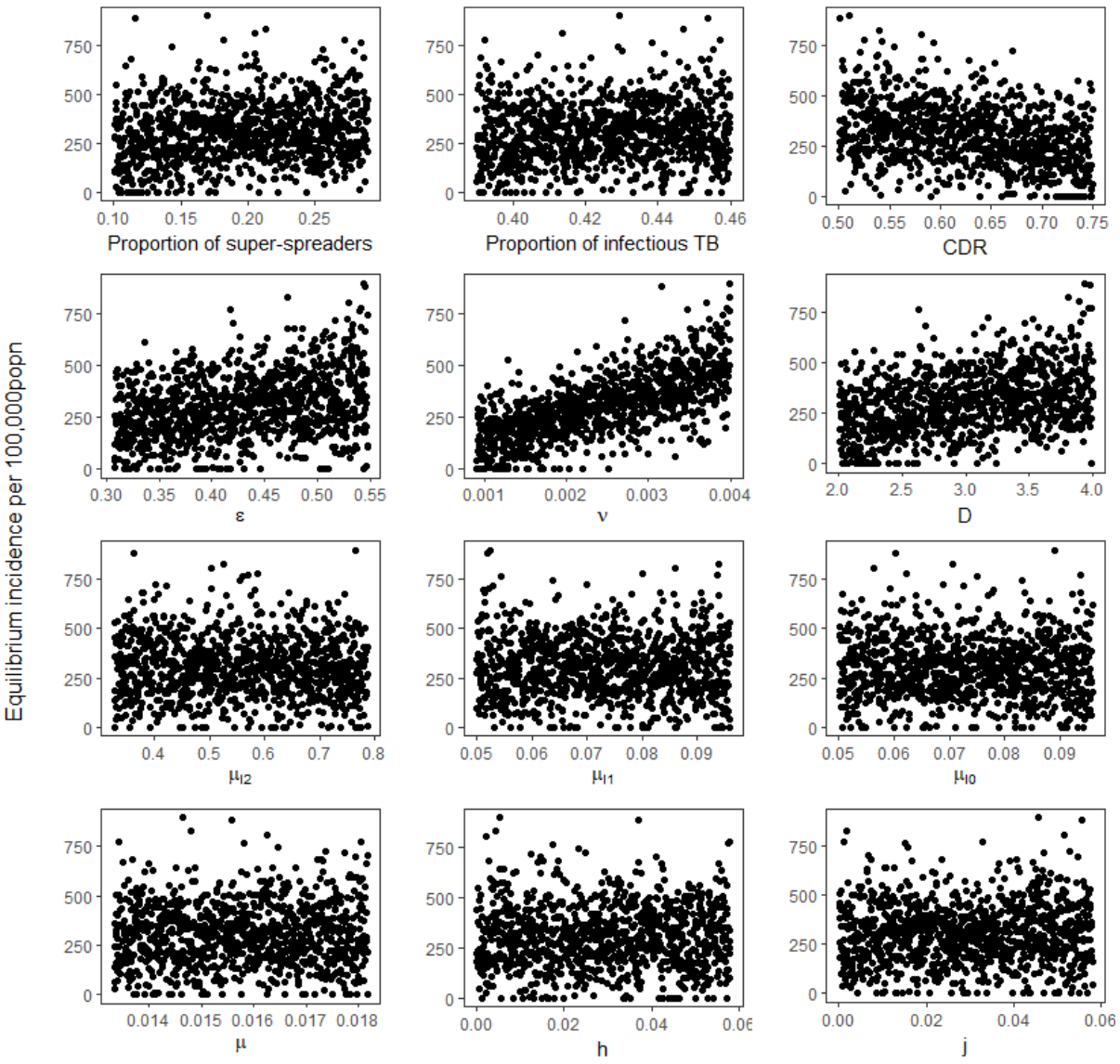
where d *is the proportion of the intervention targeted to* $I_2, d\epsilon(0,1)$. While the

intervention case detection rate for non-spreaders (CDR_{oi}) and low-spreaders (CDR_{1i}) is given as follows:

$$CDR_{oi} = \frac{CDR_{intervention} - c * CDR_{2i}}{a + b}$$

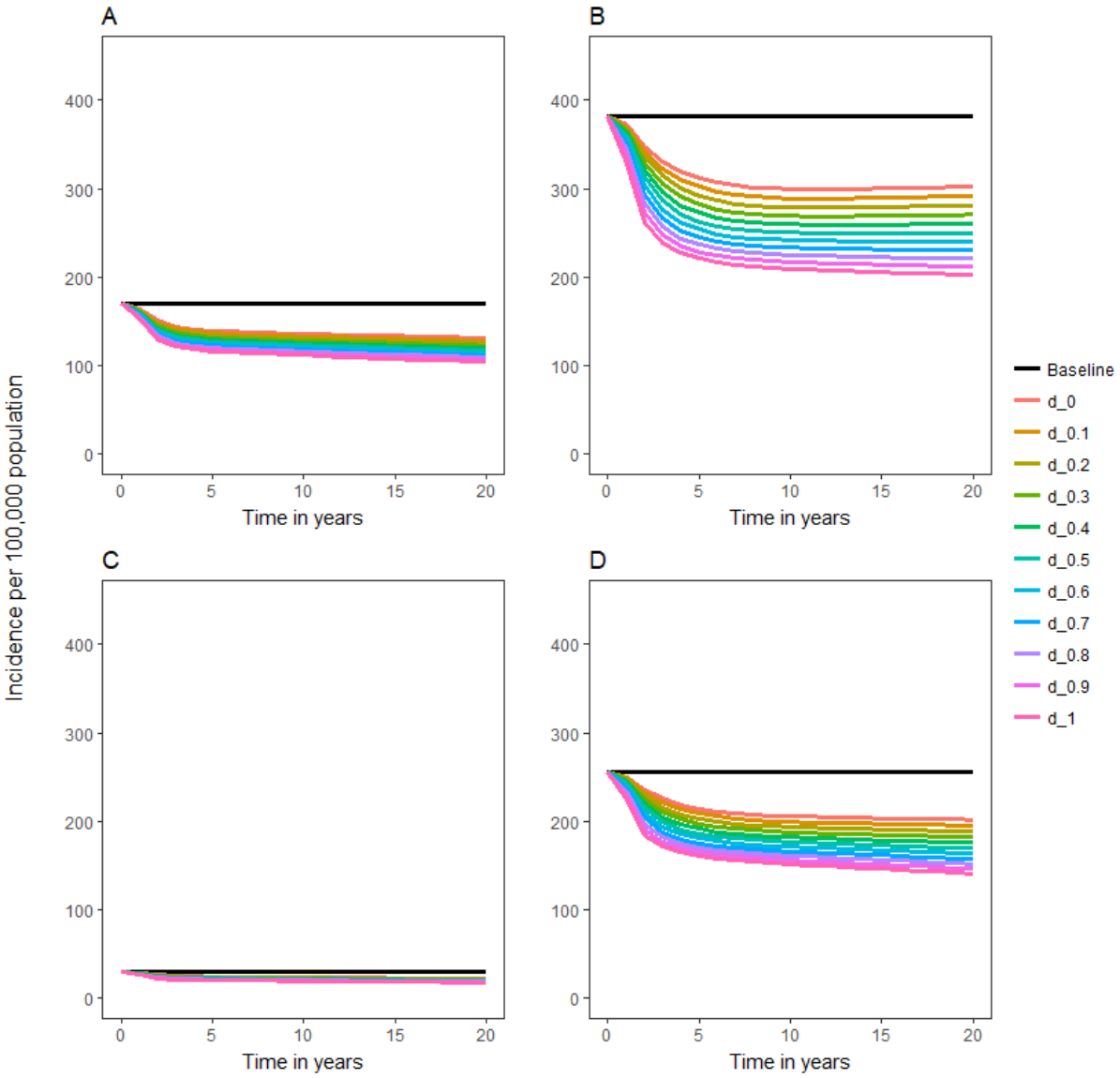
$$CDR_{1i} = CDR_{0i}$$

Multidimensional sensitivity analyses



Supplementary fig. 2: Multidimensional sensitivity analysis for the equilibrium incidence. A Latin hypercube method was used to create 1000 parameter sets, keeping the transmission rate (β) constant. The values obtained for the equilibrium incidence over the ranges of parameters are represented. Parameter values and definitions are given in Table 1.

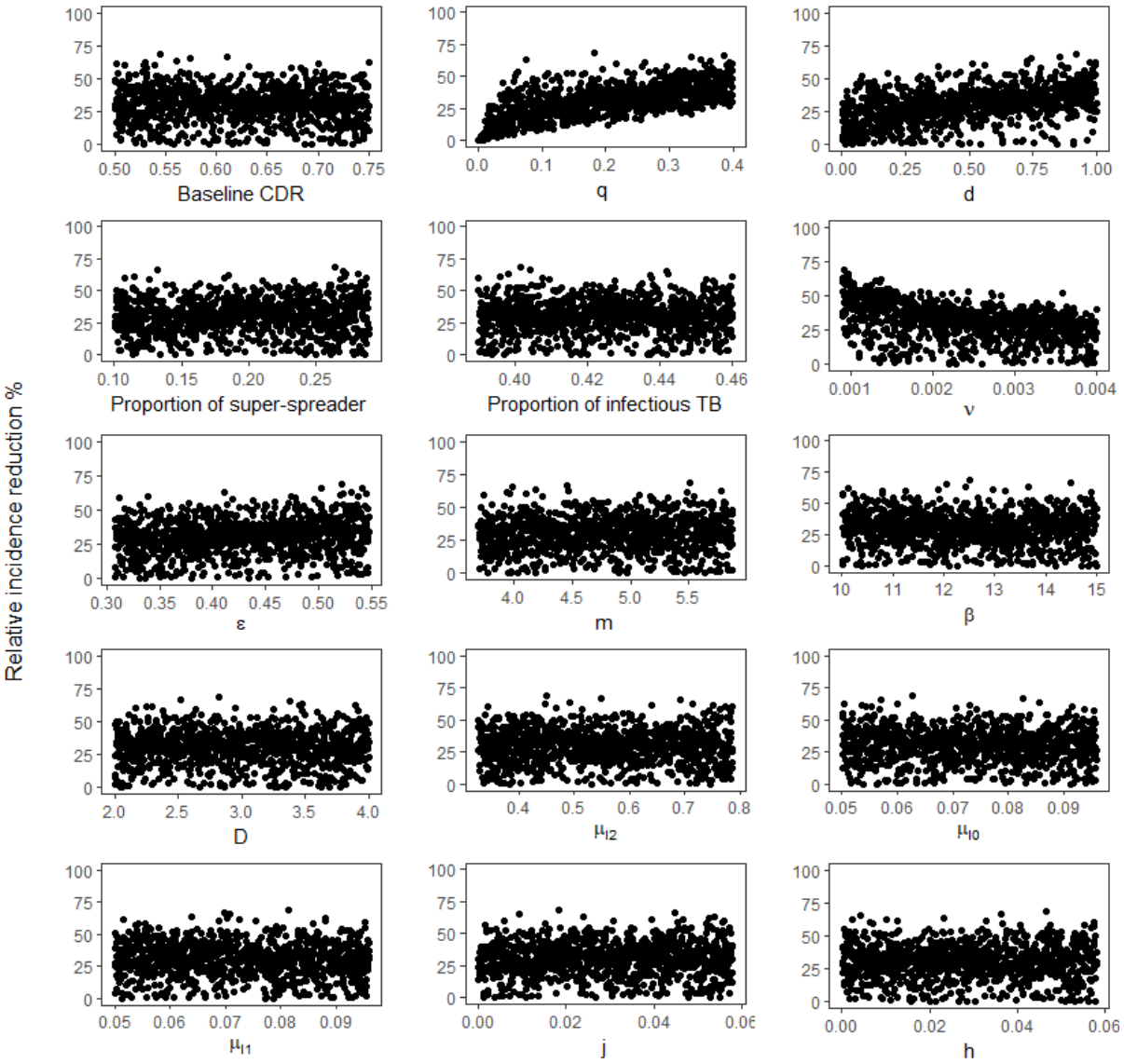
Intervention scenario results



Supplementary fig. 3: Comparison of the effect of mass and targeted interventions on the 20-year projected reduction in TB incidence under four scenarios. The intervention simulates the diagnosis of 20% of undetected cases, with the proportion of this intervention targeted towards super-spreaders ranging from zero to one, i.e. q=0.2 and d∈ [0 - 1]. Panels represent the setting scenarios: A) high transmission and high case detection setting; B) high transmission and low case detection; C) low transmission and high case detection; D) low

transmission and low case detection. In the legend, "Baseline" means the baseline equilibrium incidence, 'd_0' represents incidence after initiation of a population-wide mass case-finding intervention and the rest shows incidence after initiation of the intervention with given level of targeting, for example, *'d_0.1'* indicates 10% targeting, while 'd_0.5' indicates 50% targeting towards super-spreaders.

Intervention effectiveness sensitivity analyses



Supplementary fig 4: Multidimensional sensitivity analysis of the effect of interventions on targeting TB super-spreaders. A Latin hypercube method was used to create 1000 parameter sets. The values obtained for the relative reduction in incidence in TB within 20 years over the ranges of parameters are represented. Parameter values and definitions are given in Table 1.

References used in the supplementary material

1. Melsew, Y.A., et al., *The role of super-spreading events in Mycobacterium tuberculosis transmission: evidence from contact tracing.* BMC Infectious Diseases, 2019. **19**(1): p. 244.
2. Trauer, J.M., J.T. Denholm, and E.S. McBryde, *Construction of a mathematical model for tuberculosis transmission in highly endemic regions of the Asia-Pacific.* Journal of theoretical biology, 2014. **358**: p. 74-84.
3. Ragonnet, R., et al., *Optimally capturing latency dynamics in models of tuberculosis transmission.* Epidemics, 2017.
4. Menzies, N.A., et al., *Progression from latent infection to active disease in dynamic tuberculosis transmission models: a systematic review of the validity of modelling assumptions.* The Lancet Infectious Diseases, 2018.

Chapter 7

7. Conclusions and future directions

7.1. Conclusions and future directions

In this thesis, I estimated the impact of infectiousness heterogeneity on *Mtb* transmission and proposed a data-driven TB model that captures the phenomenon of super-spreading and could be used to predict the effectiveness of targeted interventions. The thesis objectives were achieved by integrating published evidence, TB contact investigation data and dynamic transmission models of TB. This thesis presents evidence that will assist the global effort to end TB by evaluating the effectiveness of targeted interventions towards the most significant sources of infection, with the objective of reducing TB incidence. The evidence produced in this thesis regarding the degree to which patients with TB vary in their ability to transmit infection, and the identified characteristics of highly infectious patients, could be used to support the implementation of the WHO's End TB Strategy, which includes systematic screening of contacts and high-risk groups [1, 2]. More importantly, the evaluation of targeted active case finding interventions aligns with the recent release of point-of-care tools such as GeneXpert Omni, now undergoing a multi-country randomised control trial to evaluate its impact in comparison with current tests [3-5]. The evaluation of the effectiveness of the targeted implementation of a novel point-of-care tool will also complement the *End TB Strategy's "Intensified research and innovation"* pillar, which includes a focus on optimisation of interventions.

Although heterogeneous infectiousness is the reality of *Mtb* transmission, there are practical challenges in determining its magnitude and incorporating it into TB transmission models. The availability of detailed population-based contact investigation data and linkage of these data to subsequent TB cases is typically deficient in high-burden settings. To quantify heterogeneous infectiousness and optimally decide on targeted interventions, we need to know *who infects whom* and to identify characteristics that can be used to predict patients' infectiousness. To address this challenge, I used prospectively collected data from a well-

resourced but low-burden setting. These data were used to inform the transmission model and to evaluate the effectiveness of targeted active case finding interventions. From the modelling results presented, the utility of targeting super-spreaders appears impressive, although the practicalities of accurately identifying active cases who will become super-spreaders present substantial challenges. Using systematic reviews and meta-analyses of published TB contact investigation studies, I summarised the characteristics of TB patients that are associated with greater infectiousness, which can be used to identify highly infectious patients or super-spreaders.

7.2. Key findings

Presenting the findings of a systematic review and meta-analyses of contact tracing studies, Chapter 2 assessed the risk factors associated with the capacity of index TB patients to transmit the infection to their contacts. Sputum smear positivity, the presence of lung cavitation and HIV seronegativity were the factors most consistently associated with greater infectiousness of index cases. This Chapter's findings are broadly consistent with the premise that more severe lung parenchymal involvement is the major factor for infectiousness. Clinical factors were studied and reported in a variety of ways, but findings were generally consistent with the message that severe or prolonged symptoms led to a greater spread of infection. This Chapter also demonstrated that demographic and behavioural characteristics such as older age, female sex, race, smoking and heavy alcohol consumption are frequently associated with infectiousness. However, the associations were not consistent across studies, which may be because such associations are often setting-specific. This highlights that infectiousness is a result of a multitude of interrelated factors and not merely a consequence of the clinical manifestation of their disease (***Appendix C***).

The magnitude of the variation among patients with TB in their capacity to produce secondary infections has not been previously quantified from contact investigation data,

unlike for other respiratory infections such as SARS [6, 7]. However, TB is well known to display heterogeneous pathogenesis, with extrapulmonary TB typically producing no secondary infections, while pulmonary TB shows highly variable levels of infectiousness. Chapter 3 filled this gap in our understanding by quantifying the infectiousness heterogeneity in *Mtb* transmission using population-based contact investigation data. This Chapter reported that infectiousness heterogeneity was much more critical to *Mtb* transmission than it was for many other infectious agents, with super-spreading events constituting the vast majority (three-quarters) of all transmission events. The finding from this Chapter of extensive variability between index TB patients is relevant to three epidemiological quantities: the number of contacts identified, the number of contacts infected and the number of cases of secondary active TB subsequently occurring. These estimates provide complementary information on the multitude of factors that can affect infectiousness. The focus in this Chapter was on quantifying infectiousness heterogeneity with respect to the number of secondary infections instead of secondary active TB disease. To assess heterogeneity in disease progression, I also investigated the profiles and patterns of contacts' progression to active TB disease after exposure to pulmonary TB patients in Chapter 4.

Chapter 4 reported that higher rates of progression to active TB disease were found among males, children and TST-positive contacts. The great majority of TB disease episodes among children were early progressions (occurring within six months of exposure), although high rates of late reactivation were observed as children reached about 15 years of age. Significant risk factors for TB disease activation included younger age, close contact and high TB burden country of birth. BCG vaccination history appeared to be protective for early progression but a risk factor for late reactivation, with the vast majority of adult late reactivations occurring in BCG vaccinated contacts. Two of these findings are particularly interesting: the first is that

after a window period, high rates of late TB reactivation were seen among infected children as they reached adolescence. This suggests that children should be strongly considered for preventive interventions even after their initial high-risk period has elapsed. The second is that BCG appears to have an important effect on TB epidemiology in preventing early progression, which is most relevant to children but increases rates of late reactivation in adults. As adult TB episodes are more critical to *Mtb* transmission, this effect requires further investigation with more comprehensive data on contacts' receipt of preventive therapy. This Chapter also finds the index patient characteristic of having recurrent TB to be a risk factor for contact progression to TB; this may be due to the higher infectiousness of these patients, which may be attributable to a longer duration of infectiousness or more frequently smear-positive disease.

The evidence concerning heterogeneous infectiousness presented in Chapters 2, 3 and 4 could be used as inputs for TB transmission models to test and evaluate the impact of heterogeneity parameters at the population level. Such models could also be used to predict the effectiveness of interventions with assumptions that reflect a higher degree of realism regarding heterogeneous infectiousness. To investigate and summarise past progress in this aspect in TB modelling studies, Chapter 5 described a systematic review of past TB modelling studies and described their methods of capturing heterogeneous infectiousness.

Although patients with TB have profoundly heterogeneous infectiousness and super-spreading contributes significantly to transmission events, my review of published TB models in Chapter 5 reported that the majority of published studies did not attempt to capture this phenomenon adequately. In TB models that incorporated multiple levels of infectiousness, the most common approach to stratification of the active TB compartment was to incorporate two levels of infectivity, namely non-infectious and infectious TB, with various clinical characteristics proposed as the rationale for stratifying the active TB compartments. Most of

the factors used for stratifying active TB in these models, in particular, sputum smear status and pulmonary involvement, were found to be supported by epidemiological studies. However, as described in previous chapters, heterogeneous infectiousness results from a great many factors and may not be accurately captured through stratification according to a single factor. Thus, there is a need for more comprehensive, data-driven TB models, which can capture heterogeneous infectiousness and evaluate the effectiveness of targeted interventions.

In Chapter 6, I constructed a model able to capture heterogeneous infectiousness of active TB cases by stratifying them as non-spreaders, low-spreaders and super-spreaders, without conceptually limiting this to a single clinical factor responsible for these differences. This TB model incorporated empirical parameters related to heterogeneous infectiousness that were estimated using the offspring distribution produced for the number of secondary infections per-index in Chapter 3. This model demonstrated that the parameters related to infectiousness heterogeneity were among the most critical quantities in determining the burden of TB. With this model, I evaluate the effectiveness of an active case finding intervention targeted to the most infectious patients compared to the untargeted implementation of the same intervention. The model suggested that targeted implementation of the intervention was more effective than an untargeted approach, particularly in settings of high transmission but low case detection. These model results provided valuable insights into the optimal implementation of active case finding interventions and the use of future TB point-of-care tools.

7.3. Strengths and Limitations

The strengths of this thesis arise from the approach I use to integrate evidence from published studies, extensive population-based contact investigation data and mathematical modelling to achieve the objectives listed above and to establish an evidence base around heterogeneous infectiousness in *Mtb* transmission. The data used in Chapters 3 and 4 constitute a large dataset of TB contact investigation for patients identified from 2005 to 2015 across the whole

of the state of Victoria. The dataset was well organised with contacts traced by public health nurses and identified by their index patients' and their unique numbers. The linkage for subsequent contact TB disease is extended to 2017, which provides a sufficient time period for observing TB disease progression. The model presented in Chapter 6 was novel in its use of these contact investigation data to inform its implementation of heterogeneous infectiousness.

This thesis has a number of limitations that are relevant to its interpretation. The first limitation is its use of data from a low TB incidence setting to understand *Mtb* transmission, which occurs to a much greater extent in high-burden settings. It is therefore essential to acknowledge that there may be different transmission patterns and levels of heterogeneity in high-burden settings. Heterogeneity may be context-specific – as the number of secondary cases is not just a consequence of disease characteristics, but also depends on health system access and other non-biological characteristics of the local context (***Appendix C***). However, I was unable to identify contact investigation data of a sufficiently large size with subsequent linkage to TB cases from high-burden settings, which are mostly resource-limited, to undertake population-based contact investigation. Moreover, the fact that the setting from which our empiric data were obtained was of low transmission substantially increases the probability that infection can be correctly attributed to only the index case, rather than other unlinked sources.

The second limitation was the lack of a universally agreed standard cut-point for defining TB super-spreaders. I adopted the definitions used for analysing SARS data, that the 99^{th} centile and above of the offspring distribution categorised as super-spreading; cut-point that is increasingly gaining acceptance across various diseases [7].

The third limitation was the use of hypothetical settings and a hypothetical point-of-care tool in the evaluation of targeted interventions, although the assumed tool is currently undergoing a multi-national trial in which it is being evaluated against currently available diagnostic tools. However, the generality of the model presented in this thesis should facilitate its adaptation to represent a specific active case-finding intervention once parameters for that setting are estimated.

The fourth limitation of this thesis is that heterogeneity in *Mtb* transmission was only investigated from the perspective of the index patients, and so ignored heterogeneity in the susceptibility of contacts. However, the susceptibility level of contacts most importantly affects their risk of developing active TB after infection rather than their risk of being infected [8]. This was the rationale underpinning my decision to quantify heterogeneity mainly based on the number of secondary infections produced, whereas for heterogeneity in TB disease progression I used contact characteristics only to estimate the hazard of active TB after exposure.

7.4. Implications and future directions

This thesis produces and quantifies estimates for the heterogeneity in infectiousness of TB patients that adds to the scientific understanding of *Mtb* transmission. I have presented evidence that shows that patients with TB have profound heterogeneity in their infectiousness with an extreme case of super-spreading. These findings provide information for TB clinicians and public health systems in designing approaches to limit transmission from highly infectious patients and prevent disease among exposed contacts. Besides, the distinct patterns of TB activation observed among children, with TB reactivation frequently occurring after a window period, should help to guide policy for preventive treatment and post-exposure follow-up. Stakeholders of TB control programs should now better appreciate that the size and impact of heterogeneity are increasing as the global TB control community looks

towards ending the epidemic. Specifically, since just a very few patients may cause severe localised outbreaks, targeted implementations of interventions are likely to be a particularly effective approach.

The methodological novelty of this thesis provides a springboard for future researchers aiming to incorporate the reality of infectiousness heterogeneity in different settings. In the process of quantifying infectiousness heterogeneity, I produced an empiric distribution of the number of secondary infections per index and fitted a negative binomial distribution to these data. I used a Bayesian Hamiltonian Monte Carlo algorithm to estimate model parameters related to infectiousness heterogeneity from this distribution of secondary infections. In addition, the approach of intervention simulation can be used in future modelling studies evaluating similar targeted interventions.

Although this study provides estimates of the magnitude and impact of TB infectiousness heterogeneity, important research questions remain to be answered. The need to characterise super-spreaders in specific settings in order to optimise targeted interventions is probably the most critical example. The patterns of TB reactivation in children as they reach adolescence and the effect of BCG on adult TB reactivation also warrant further investigation, as these groups are highly infectious and so contribute substantially to transmission. Future TB modelling studies should also investigate the effectiveness of interventions, such as preventive treatment, using more realistic assumptions concerning infectiousness heterogeneity and setting-specific data. In transmission models of TB, incorporating assumptions of heterogeneous infectiousness increases realism, although it also increases model complexity. Thus, the necessity of adding extra complexity must be justified while considering modelling objectives, such as evaluating targeted public health and diagnostic interventions, and predicting the relative magnitude of specific types of TB.

From my doctoral studies, I conclude that the role of infectiousness heterogeneity in *Mtb* transmission is substantial and transmission is substantially driven by super-spreaders. Hence, interventions targeted towards super-spreaders and their contacts are likely to be considerably more effective than untargeted implementations.

7.5. References

1. WHO, *The End TB Strategy: Global strategy and targets for tuberculosis prevention, care and control after 2015.* 2014: Geneva, Switzerland
2. Goosby, E., et al., *The Lancet Commission on tuberculosis: building a tuberculosis-free world.* The Lancet, 2018. **391**(10126): p. 1132-1133.
3. Institute, I.H. *Evaluation of Cepheid GeneXpert Omni combined with the Xpert MTB/RIF Ultra in Primary Health Care in Tanzania*. 2016 [cited 2019 24 August]; Available from: http://ihi.or.tz/project/cepheid-genexpert-omni-combined-with-the-xpert-mtbrif-ultra-for-detection-of-tuberculosis-and-rifampicin-resistance-in-adults-with-presumptive-pulmonary-tuberculosis-at-primary-level-diagnostic-cent/.
4. Ejalu, D.L., *Cost Effectiveness analysis of GeneXpert Omni compared to GeneXpert MTB/Rif for point of care diagnosis of Tuberculosis among presumptive cases in Pingire health center III, Serere district.* 2018, Makerere University.
5. Institute, B. *Rapid and accurate diagnosis of TB: New trial in PNG*. 2019 [cited 2019 24 August]; Available from: https://www.burnet.edu.au/projects/284_rapid_and_accurate_diagnosis_of_tb_new_trial_in_png.
6. Kucharski, A. and C. Althaus, *The role of superspreading in Middle East respiratory syndrome coronavirus (MERS-CoV) transmission.* Euro surveillance, 2015. **20**(25): p. pii= 21167.
7. Lloyd-Smith, J.O., et al., *Superspreading and the effect of individual variation on disease emergence.* Nature, 2005. **438**(7066): p. 355-359.
8. Lawn, S.D. and A.I. Zumla, *Tuberculosis.* Lancet, 2011. **378**: p. 57-52.

Appendices

Appendix A: Ethics Approval

Monash University Human Research Ethics Committee

Approval Certificate

This is to certify that the project below was considered by the Monash University Human Research Ethics Committee. The Committee was satisfied that the proposal meets the requirements of the *National Statement on Ethical Conduct in Human Research* and has granted approval.

Project Number: 7776
Project Title: Understanding the Causes of Heterogeneity in Tuberculosis Transmission
Chief Investigator: Dr James Trauer
Expiry Date: 22/02/2022

Terms of approval - failure to comply with the terms below is in breach of your approval and the *Australian Code for the Responsible Conduct of Research*.

1. The Chief Investigator is responsible for ensuring that permission letters are obtained, if relevant, before any data collection can occur at the specified organisation.
2. Approval is only valid whilst your hold a position at Monash University.
3. It is responsibility of the Chief Investigator to ensure that all investigators are aware of the terms of approval and to ensure the project is conducted as approved by MUHREC.
4. You should notify MUHREC immediately of any serious or unexpected adverse effects on participants or unforeseen events affecting the ethical acceptability of the project.
5. The Explanatory Statement must be on Monash letterhead and the Monash University complaints clause must include your project number.
6. Amendments to approved projects including changes to personnel must not commence without written approval from MHUREC.
7. Annual Report - continued approval of this project is dependent on the submission of an Annual Report.
8. Final Report - should be provided at the conclusion of the project. MUHREC should be notified if the project is discontinued before the expected completion date.
9. Monitoring - project may be subject to an audit or any other form of monitoring by MUHREC at any time.
10. Retention and storage of data - The Chief Investigator is responsible for the storage and retention of the original data pertaining to the project for a minimum period of five years.

Thank you for your assistance.

Professor Nip Thomson

Chair, MUHREC

CC: Assoc Professor Manoj Gambhir, Professor Allen Cheng

List of approved documents:

Document Type	File Name	Date	Version
Supporting Documentation	ethics_proposal_yaye_2017_2	10/02/2017	2

Appendix B: Publication from Chapter 4

Profiles of tuberculosis disease activation among contacts of patients with tuberculosis

To the Editor:

The risk of a person progressing to tuberculosis (TB) disease after infection with *Mycobacterium tuberculosis* remains poorly understood, with some contacts developing TB in the early period following exposure, while others take many years to progress or never do so [1, 2]. We described profiles and patterns of contacts' progression to TB disease following exposure by linking a large, prospectively collected contact investigation dataset from Victoria, Australia to data on subsequent cases of active TB disease, after obtaining ethical approval from the Monash University, Human Research Ethics Committee. Unlike many past studies, this approach offers the opportunity to disaggregate by various characteristics of both index patient and exposed contact.

The main outcomes of interest were early progression (development of active TB within 6 months of exposure) and late reactivation (development of TB thereafter). Our definition for early progression is supported by previous findings that the risk of TB is highest in the first 3–9 months after infection [3, 4]. We used a survival analysis approach to describe patterns of TB disease reactivation and used logistic and Cox regression to quantify risks associated with early and late reactivation respectively, incorporating variables pertaining to both contact and index patients.

17 740 contacts of pulmonary TB patients were included in the analysis, of which 82.0% were aged 15 years and above. Over more than 7 years of median follow-up, 224 (1.3%) contacts were diagnosed with TB disease, resulting in an overall incidence rate of 189 cases per 100 000 person-years. Most cases (62.5%) accrued during the first 6 months following exposure and so were defined as early progression episodes. We found that male contacts had a slightly greater rate of TB than their female counterparts (log-rank $p=0.03$), while children aged <5 years at exposure had the greatest rate of active TB development, followed by children aged 5–14 years (log-rank $p<0.001$). Tuberculin skin test (TST) positivity (⩾10 mm) was associated with a greater rate of TB (4.6%, 148/3213) than for TST-negative contacts (0.2%, 29/12 233), despite TST positives being more likely to receive preventive therapy. Younger age of contact, close contact, high-burden country of birth, absence of BCG vaccination and recurrent TB in the index patient were associated with early progression, while close contact, high-burden country of birth, BCG vaccination history and BCG status not stated were associated with contacts' higher hazard of late reactivation (table 1).

Every TB case observed among children aged <5 years occurred in the first 18 months of follow-up, of which 44/51 (86.3%) were early progressions. Although a similar proportion (49/60, 81.6%) of early progressions was found among child TB contacts aged 5–14 years, there were seven TB activation episodes that occurred after more than 2 years of follow-up in this age group. These episodes predominantly occurred as these children aged into adolescence and reached approximately 15 years of age. This pattern has also been observed in historical works, which have reported that when a first exposure occurred in children (aged 1–14 years), TB typically did not develop until after the age of 15 years [3]. A prospective study from Hong Kong provided similar evidence that infected children were at greater risk of TB beyond the age of 15 years after an initial low-risk window period [4]. The predilection of TB to affect young

@ERSpublications
The risk of TB reactivation among infected children increases as they reach the age of adolescence. BCG vaccination history seems to increase the risk of TB disease reactivation among adults exposed to *Mycobacterium tuberculosis*. http://bit.ly/2YReXej

Cite this article as: Melsew YA, Cheng AC, McBryde ES, *et al*. Profiles of tuberculosis disease activation among contacts of patients with tuberculosis. *Eur Respir J* 2019; 0: 1900353 [https://doi.org/10.1183/13993003.00353-2019].

TABLE 1 Regression analyses for early progression and late reactivation of tuberculosis (TB) among contacts of TB patients in Victoria, Australia 2005–2015

	TB disease		Crude OR/HR (95% CI)	Adjusted OR/HR (95% CI)	p-value
	No	Yes			
Factors associated with early progression#					
Contact characteristics					
Sex					
Female	9132	62	Ref	Ref	
Male	8349	78	1.38 (0.98–1.93)	1.23 (0.88–1.73)	0.227
Not stated	119	0	NA	NA	
Age					
0–4 years	1153	44	11.78 (7.75–17.84)	7.79 (4.80–12.62)	<0.001*
5–14 years	1951	49	7.75 (5.17–11.61)	5.23 (3.39–8.07)	<0.001*
15 years and above	14496	47	Ref	Ref	
Type of contact					
Close contacts	7871	123	8.94 (5.54–15.40)	5.73 (3.39–9.69)	<0.001*
Other contacts	9729	17	Ref	Ref	
Country of birth					
Low burden	11824	100	Ref	Ref	
High burden	5090	40	0.73 (0.52–1.02)	1.76 (1.11–2.81)	0.017*
Not stated	686	0	NA	NA	
Tuberculin skin test					
Negative	12222	11	Ref	Not included	
Positive	3112	101	36.06 (20.24–71.31)*		
Not performed	2266	28	13.73 (7.02–28.83)*		
BCG					
No	4995	69	Ref		
Yes	7505	37	0.36 (0.24–0.53)	0.47 (0.29–0.77)	0.002*
Not stated/unknown	5100	34	0.48 (0.32–0.73)	0.79 (0.50–1.23)	0.288
Index characteristics					
Sex					
Male	9439	75	Ref		
Female	8161	65	1.00 (0.72–1.40)	1.14 (0.80–1.62)	0.485
Age					
1–14 years	521	10	3.76 (1.76–7.31)	0.20 (0.03–1.45)	0.110
15–24 years	7249	37	Ref		
25–44 years	5181	73	2.76 (1.87–4.15)	1.04 (0.69–1.55)	0.860
45–64 years	1713	17	1.94 (1.06–3.40)	1.57 (0.91–2.71)	0.101
65 years and above	2936	3	0.20 (0.05–0.55)	0.60 (0.32–1.13)	0.111
Country of birth					
High burden	8897	77	Ref	Ref	
Low burden	8703	77	0.84 (0.60–1.17)	1.15 (0.81–1.63)	0.446
Method of case finding					
Clinical presentation	16628	132	Ref	Ref	
Other methods	972	8	1.04 (0.46–1.99)	1.71 (0.91–3.21)	0.096
Site of TB					
Pulmonary	14655	122	Ref	Ref	
Pulmonary plus other site	2945	18	0.73 (0.43–1.17)	1.20 (0.76–1.90)	0.435
Type of TB					
New	16584	130	Ref	Ref	
Recurrent	774	10	1.65 (0.81–2.99)	2.68 (1.42–5.06)	0.002*
Unknown	242	0	NA	NA	
Chest radiography finding					
Abnormal, no cavity	10172	79	Ref	Ref	
Abnormal, cavity	5959	55	1.19 (0.84–1.68)	1.20 (0.82–1.75)	0.347
Abnormal, no details	265	1	0.49 (0.07–3.51)	0.30 (0.04–2.21)	0.236
Normal	339	2	0.76 (0.19–3.10)	0.72 (0.17–3.04)	0.655
Not stated/done	865	3	0.45 (0.14–1.42)	0.42 (0.13–1.36)	0.146
Sputum smear					
Yes	2212	15	Ref	Ref	
No	3609	31	0.85 (0.49–1.47)	0.88 (0.47–1.62)	0.677
Not done/recorded	11779	94	1.08 (0.72–1.62)	1.29 (0.83–2.01)	0.264

Continued

TABLE 1 Continued

	TB disease		Crude OR/HR (95% CI)	Adjusted OR/HR (95% CI)	p-value
	No	Yes			
Factors associated with late reactivation[¶]					
Contact characteristics					
Sex					
Female	7585	31	Ref	Ref	
Male	6733	35	1.26 (0.77–2.06)	1.20 (0.74–1.96)	0.460
Not stated	112	0	NA	NA	
Type of contact					
Close contacts	5785	50	4.43 (2.49–7.86)	3.25 (1.81–5.82)	<0.001*
Other contacts	8645	16	Ref		
Country of birth					
Low burden	9246	23	Ref	Ref	
High burden	4602	43	3.73 (2.23–6.24)	2.17 (1.26–3.75)	0.005*
Not stated	582	0	NA	NA	
Tuberculin skin test					
Negative	9629	16	Ref	Not included	
Positive	2729	35	6.94 (3.78–2.74)*		
Not performed	2072	15	5.06 (2.45–0.43)*		
BCG					
No	3374	3	Ref	Ref	
Yes	6794	46	7.89 (2.43–25.65)	4.54 (1.35–15.28)	0.014*
Not stated	4262	17	4.48 (1.30–15.42)	3.52 (1.01–12.26)	0.048*
Index characteristics					
Sex					
Male	7632	24	Ref	Ref	
Female	6798	42	2.08 (1.14–3.82)	0.55 (0.29–1.03)	0.06
Country of birth					
High burden	7672	36	Ref	Ref	
Low burden	6758	30	1.39 (0.76–2.54)	1.43 (0.78–2.60)	0.24
Method of case finding					
Clinical presentation	13533	64	Ref	Ref	
Other methods	897	2	0.50 (0.11–2.27)	0.42 (0.09–1.89)	0.26
Type of TB					
New case	13553	64	Ref	Ref	
Recurrent	704	2	0.72 (0.14–3.69)	0.82 (0.16–4.11)	0.81
Unknown	173	0	NA	NA	
Chest radiography finding					
Abnormal, no cavity	8025	40	Ref	Ref	
Abnormal, cavity	5285	25	0.92 (0.56–1.52)	0.62 (0.16–4.11)	0.17
Abnormal, no details	185	0	NA	NA	
Normal	291	0	NA	NA	
Not done/stated	644	1	0.27 (0.04–1.99)	0.40 (0.04–3.89)	0.43

*: statistically significant. [#]: logistic regression (all contacts, n=17740); analysis results are presented as odds ratios with 95% confidence intervals. [¶]: Cox regression (contacts age ⩾15 years, n=14496); analysis results are presented as hazard ratios with 95% confidence intervals.

adults has been recognised since Hippocrates, although there are multiple possible explanations for the considerable burden of TB in adolescents, particularly assortative mixing by age leading into the years in which active TB is most infectious [5–7]. Given our very low transmission setting and the fact that all of these children were born in low-burden settings, our findings provide considerable evidence for the explanation that infected children enter a higher risk phase as they progress towards adulthood, which has major implications for clinical care and public health responses. The absence of any cases occurring in adolescence among those aged 0–4 years at the time of exposure likely reflects insufficient follow-up duration, as only the oldest such contacts entering during the earliest period of the study would have reached adolescence, and could also reflect more consistent use of preventive therapy in this group.

Consistent with previous studies, we found that BCG vaccination was highly protective against early progression in contacts aged 0–14 years at exposure, which is a similar effect size to those reported by a previous meta-analysis [8]. However, unexpectedly, risk of late reactivation was greater among BCG vaccinated and BCG not stated adult contacts than BCG unvaccinated counterparts. Previous studies have

consistently reported that BCG is effective in preventing TB disease among children and that its effectiveness is reduced when all ages are considered [9–11]. We note similar findings from the largest ever trial estimating the efficacy of BCG, which reported nonsignificantly higher rates of TB among every age group of vaccinated adults than in the unvaccinated over 15 years of follow-up [12]. One possible interpretation of this evidence is that BCG is effective in reducing TB-related child morbidity and mortality, but may increase late reactivation rates in adults. These phenomena may lead to perverse effects on the overall epidemic because the adult forms of TB are typically most infectious, and they are critical when considering BCG revaccination programmes and interpreting recent trials of novel vaccines whose outcomes consider immunological responses to *M. tuberculosis* rather than incidence of TB disease [13]. Although we found no evidence of an association between BCG vaccination and absence of preventive therapy, this or other unmeasured factors could confound these findings. Furthermore, adults with unknown BCG status were also associated with higher rates of late reactivation, such that ensuring comprehensive collection of BCG status will be a future focus of our research.

Although the development of active TB in contacts after infection is largely dependent on contacts' endogenous factors, we also found associations with index patients' characteristics. Specifically, index patients with recurrent TB appeared more infectious, with their contacts at a considerably greater odds of early progression than for new TB cases. This may be explained by patients with recurrent TB having longer infectious periods or more frequently being smear-positive TB than those with a first episode, although this effect was not seen for late reactivation.

Important limitations of our study include that transmission was inferred based on epidemiological links due to the absence of genotyping data and we did not have universal information on the provision of preventive therapy that could have significantly influenced our findings. However, because preventive therapy is recommended universally for all contacts of patients with active TB in our setting, we would generally expect this to reduce the absolute TB risk across the cohort, but be less likely to systematically bias the trends reported in this manuscript.

In conclusion, child contacts were at highest risk of early progression, but high rates of late reactivation among infected children as they reached adolescence imply that children should be strongly considered for preventive interventions even after their initial high-risk period has elapsed. BCG seems to have an important effect on TB epidemiology in preventing early activation but increasing rates of late reactivation in adults. This effect requires further investigation with more comprehensive data on contacts' receipt of preventive therapy.

Yayehirad A. Melsew[1,2], Allen C. Cheng[1], Emma S. McBryde[3,4], Justin T. Denholm[5,6], EeLaine Tay[7], Romain Ragonnet[1] and James M. Trauer[1,5]

[1]Dept of Epidemiology and Preventive Medicine, School of Public Health and Preventive Medicine, Monash University, Melbourne, Australia. [2]Dept of Epidemiology and Biostatistics, Institute of Public Health, University of Gondar, Gondar, Ethiopia. [3]Australian Institute of Tropical Health and Medicine, James Cook University, Townsville, Australia. [4]Dept of Medicine at Royal Melbourne Hospital, University of Melbourne, Melbourne, Australia. [5]The Victorian Tuberculosis Program at the Peter Doherty Institute, Melbourne, Australia. [6]Dept of Microbiology and Immunology, University of Melbourne, Melbourne, Australia. [7]Health Protection branch, Dept of Health and Human Services, Melbourne, Australia.

Correspondence: Yayehirad A. Melsew, Monash University, Epidemiology and Preventive Medicine, 99 Commercial Rd, Melbourne, Victoria, 3004, Australia. E-mail: 078yayu@gmail.com or yayehirad.melsew@monash.edu

Received: 20 Feb 2019 | Accepted after revision: 25 May 2019

Conflict of interest: Y.A. Melsew has nothing to disclose. A.C. Cheng has nothing to disclose. E.S. McBryde has nothing to disclose. J.T. Denholm has nothing to disclose. E. Tay has nothing to disclose. R. Ragonnet has nothing to disclose. J. M. Trauer has nothing to disclose.

References

1 Turner RD, Chiu C, Churchyard GJ, *et al.* Tuberculosis infectiousness and host susceptibility. *J Infect Dis* 2017; 216: Suppl. 6, S636–S643.
2 Reichler MR, Khan A, Sterling TR, *et al.* Risk and timing of tuberculosis among close contacts of persons with infectious tuberculosis. *J Infect Dis* 2018; 40: 1–9.
3 Zeidberg LD, Gass R, Dillon A, *et al.* The Williamson County Tuberculosis Study. A twenty-four-year epidemiologic study. *Am Rev Respir Dis* 1963; 87: 1–88.
4 Leung CC, Yew WW, Au KF, *et al.* A strong tuberculin reaction in primary school children predicts tuberculosis in adolescence. *Pediatr Infect Dis J* 2012; 31: 150–153.
5 Snow KJ, Sismanidis C, Denholm J, *et al.* The incidence of tuberculosis among adolescents and young adults: a global estimate. *Eur Respir J* 2018; 51: 1702352.
6 Daniel T, Iversen P. Hippocrates and tuberculosis. *Int J Tuberc Lung Dis* 2015; 19: 373–374.

7 Palmer CE, Shaw L, Comstock G. Community trials of BGG vaccination. *Am Rev Tuberc Pulm Dis* 1958; 77: 877–907.
8 Mangtani P, Abubakar I, Ariti C, *et al.* Protection by BCG vaccine against tuberculosis: a systematic review of randomized controlled trials. *Clin Infect Dis* 2014; 58: 470–480.
9 Rodrigues LC, Diwan VK, Wheeler JG. Protective effect of BCG against tuberculous meningitis and miliary tuberculosis: a meta-analysis. *Int J Epidemiol* 1993; 22: 1154–1158.
10 Roy A, Eisenhut M, Harris R, *et al.* Effect of BCG vaccination against *Mycobacterium tuberculosis* infection in children: systematic review and meta-analysis. *BMJ* 2014; 349: g4643.
11 Soysal A, Millington KA, Bakir M, *et al.* Effect of BCG vaccination on risk of *Mycobacterium tuberculosis* infection in children with household tuberculosis contact: a prospective community-based study. *The Lancet* 2005; 366: 1443–1451.
12 Radhakrishna S, Frieden TR, Subramani R. Influence of sex, age and nontuberculous infection at intake on the efficacy of BCG: re-analysis of 15-year data from a double-blind randomized control trial in south India. *Indian J Med Res* 2006; 123: 119–124.
13 Nemes E, Geldenhuys H, Rozot V, *et al.* Prevention of *M. tuberculosis* infection with H4: IC31 vaccine or BCG revaccination. *N Engl J Med* 2018; 379: 138–149.

Author Queries

Journal: ERJ

Manuscript: ERJ-00353-2019

No Queries

Appendix C: Additional publication relevant to the thesis

Clinical Infectious Diseases

VIEWPOINTS

The Importance of Heterogeneity to the Epidemiology of Tuberculosis

James M. Trauer,[1] Peter J. Dodd,[2] M. Gabriela M. Gomes,[3,4] Gabriela B. Gomez,[5] Rein M. G. J. Houben,[6,7] Emma S. McBryde,[8] Yayehirad A. Melsew,[1] Nicolas A. Menzies,[9] Nimalan Arinaminpathy,[10] Sourya Shrestha,[11] and David W. Dowdy[11]

[1]School of Public Health and Preventive Medicine, Monash University, Melbourne, Victoria, Australia; [2]Health Economic and Decision Science, University of Sheffield, United Kingdom; [3]Liverpool School of Tropical Medicine, United Kingdom; [4]CIBIO-InBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, Universidade do Porto, Portugal; [5]Department of Global Health and Development, [6]Tuberculosis Centre, and [7]Infectious Disease Epidemiology Department, London School of Hygiene and Tropical Medicine, United Kingdom; [8]Australian Institute of Tropical Health and Medicine, James Cook University, Townsville, Queensland; [9]Department of Global Health and Population, Harvard T. H. Chan School of Public Health, Boston, Massachusetts; [10]Department of Infectious Disease Epidemiology, School of Public Health, Imperial College London, United Kingdom; and [11]Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland

Although less well-recognized than for other infectious diseases, heterogeneity is a defining feature of tuberculosis (TB) epidemiology. To advance toward TB elimination, this heterogeneity must be better understood and addressed. Drivers of heterogeneity in TB epidemiology act at the level of the infectious host, organism, susceptible host, environment, and distal determinants. These effects may be amplified by social mixing patterns, while the variable latent period between infection and disease may mask heterogeneity in transmission. Reliance on notified cases may lead to misidentification of the most affected groups, as case detection is often poorest where prevalence is highest. Assuming that average rates apply across diverse groups and ignoring the effects of cohort selection may result in misunderstanding of the epidemic and the anticipated effects of control measures. Given this substantial heterogeneity, interventions targeting high-risk groups based on location, social determinants, or comorbidities could improve efficiency, but raise ethical and equity considerations.

Keywords. tuberculosis; heterogeneity; epidemiology; case detection; interventions.

Although estimates of the global burden of tuberculosis (TB) suggest gradual decline, this aggregate profile masks a patchy, heterogeneous epidemic that predominantly afflicts society's most marginalized groups. Meanwhile, the causative organism is now the world's leading infectious killer and dramatic reductions in burden will be necessary if the bold new End TB targets are to be realised [1]. Heterogeneity in disease distribution increases as the burden of an infectious disease declines and becomes more unevenly distributed across space or social networks [2]—a phenomenon that is well recognized in the case of diseases such as malaria [3]. There are many reasons to suspect that TB epidemics are highly heterogeneous, such as the prominence of highly localized or household transmission, the wide geographical variation in disease burden within and between countries, and the many individual-level factors strongly associated with risk of disease. Here we describe key drivers of heterogeneity in TB burden, discuss the challenges in quantifying this heterogeneity, and consider implications for transmission dynamics and the design of interventions.

Received 1 August 2018; editorial decision 24 October 2018; accepted 31 October 2018; published online November 1, 2018.

Correspondence: J. M. Trauer, School of Public Health and Preventive Medicine, Monash University, 553 St Kilda Road, Melbourne, Victoria 3004, Australia (james.trauer@monash.edu).

Clinical Infectious Diseases® 2019;69(1):159–66

DOI: 10.1093/cid/ciy938

DRIVERS OF HETEROGENEITY

Risk of infectious disease is dependent on characteristics of the infectious host, the organism, the susceptible host, and the environment (Figure 1; Table 1). The complex interplay between the pathogen and the host's immune system and the propensity for *Mycobacterium tuberculosis* (*Mtb*) to enter a latent state following infection mean that many exposed individuals will never progress to active TB disease. Therefore, individual characteristics that predispose to susceptibility to infection, progression to disease after infection, and infectiousness during disease episodes all contribute to heterogeneity, although the risk factors associated with each differ considerably. For example, risk of exposure is driven by sociodemographic factors (eg, crowding, contact patterns), susceptibility to infection once exposed is influenced by processes that impair local immune responses (eg, smoking), progression to disease may reflect systemic immune status (eg, human immunodeficiency virus [HIV], nutrition), and likelihood of onward transmission may be altered by cough symptomatology and disease duration (eg, through access to care).

The Infectious Host

Medical and demographic factors also strongly influence the extent to which each affected person propagates *Mtb* infection. Smear-positive adults and particularly those with cavitary pulmonary TB transmit infection more extensively [13], while many others, such as those with only extrapulmonary

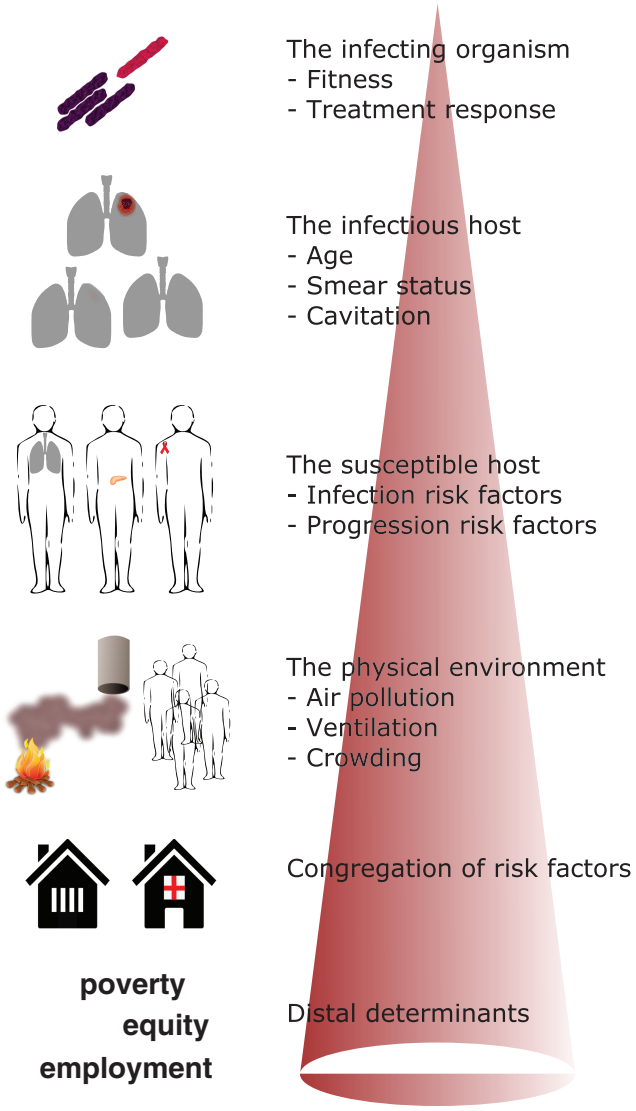


Figure 1. Conceptual framework for understanding heterogeneity in tuberculosis epidemiology. The cone indicates that the most local drivers are positioned toward the top of the figure and the broadest drivers toward the bottom, rather than reflecting the importance of these factors.

involvement, may infect no one. Although children and persons with HIV are less likely to transmit, the degree of infectiousness is variable, with children aged >10 years more often manifesting adult forms of TB [14, 15]. Despite its limitations, smear microscopy remains the mainstay of TB diagnosis worldwide with advantages that include its ability to identify highly infectious individuals. Social factors such as mixing patterns also influence spread by modifying the number of contacts exposed and these patterns also differ by setting (eg, household, workplace, general community). Importantly, social mixing patterns may act differently for *Mtb* than for other infections, given that *Mtb*, unlike many other major pathogens, is airborne and so can be transmitted without the need for direct person-to-person contact. However, the rate of transmission per day infectious is considerably lower than for other respiratory pathogens (eg, measles, influenza) [16], meaning that amplifying factors such as cough characteristics, ability to generate aerosols of appropriate size [17], and environmental factors may strongly influence whether infection occurs. Finally, myriad programmatic and social factors delay diagnosis and so prolong the infectious period and increase the duration of exposure [18], thereby potentiating heterogeneity through their impact on the most marginalized groups.

The Infecting Organism

Mtb is a clonal pathogen that displays variable fitness and a complex interaction with its human host [19]. Its multiple lineages differ in their genomic makeup and in several aspects of their clinical and epidemiological behavior, including disease progression, disease severity, transmissibility, and geographic distribution (Supplementary Bibliography). With recent advances in molecular epidemiology, the influence of *Mtb* genetic diversity on the outcomes of TB infection and disease is increasingly recognized. Strains are thought to have adapted to the human population they affect [20], resulting in a sympatric relationship whereby co-evolved host populations show high rates of TB due to certain strains, but concentration within high-risk groups elsewhere [21]. However, the discordance in findings between settings and the complex interaction between pathogen, host, and environment remain challenges to understanding these processes.

Arguably, the most critical form of pathogen-related heterogeneity is drug resistance, which makes clinical management considerably more challenging and expensive. Epidemiologically, transmission cycles of drug-resistant TB (DR-TB) differ from those of drug-susceptible TB because of limited access to the diagnostics available for determining drug resistance, the long duration of DR-TB treatment, and clustering of DR-TB patients in high-risk settings. All of these factors may act to prolong the infectious period, sustaining transmission chains of DR-TB [22]. Resistance-conferring mutations may be offset by associated physiological impairments in the organism that limit its ability to survive and reproduce ("fitness costs"), although sustained drug exposure may select for bacteria with compensatory mutations [23]. Moreover, fitness costs are likely to vary according to the drug in question (eg, higher for rifampicin resistance than for isoniazid or streptomycin) [24], while both modeling studies and large-scale outbreaks highlight the potential for DR-TB to proliferate [25].

The Susceptible Host

Characteristics of the susceptible host also markedly influence the likelihood of disease following exposure, which may reflect both susceptibility to infection or greater risk of progression to disease for those infected. Patterns of reactivation differ markedly by age, and comorbidities such as HIV,

Table 1. Examples of Specific Forms of Heterogeneity and Ways Forward

Source of Heterogeneity	Examples of Existing Evidence	Data Needs	Analytic Needs	Intervention Needs
Infectious host	Sequencing and social network analysis suggest that some individuals may act as "superspreaders" [4]	Importance of biological variables, eg, aerosolization, cough frequency	Implications of hosts with differential infectiousness and superspreading	Tools to identify the most infectious patients
	Available data on contact patterns (principally from low-burden settings) suggest age-specific (assortative) mixing	Data on contact patterns from high-burden settings and for risk factors relevant to TB (eg, HIV status)	Importance of population groups to sustaining transmission relative to their burden of disease	Case-finding efforts designed to identify patients with high-risk mixing patterns for broader dissemination of infection
Infecting organism	Strain responsible for extensive community spread confirmed to be highly virulent in mouse model [5]	Mechanisms of strain diversity and virulence	Implications of selecting for strains of greater fitness	Interventions to limit infectiousness of difficult-to-treat strains
	Highly resistant forms of TB causing extensive outbreaks, eg, XDR-TB in Tugela Ferry, South Africa [6]	Fitness costs associated with drug resistance	Likely future trajectory of drug resistance	Improved identification and treatment of highly transmissible strains of drug-resistant TB
Susceptible host	Individuals previously treated for TB had higher rates of recurrent TB due to reinfection than the general population in Cape Town, South Africa [7]	Protection or susceptibility afforded by past TB episodes and whether this is attributable to infection or progression risk	Distinguish the individual-level effect of increased susceptibility post–disease episode from the effect of selecting for a more susceptible cohort through infection	Protection of highest-risk individuals from infection or progression to disease
	Specific risk groups may experience polyclonal outbreaks [8]	Better estimates of disease prevalence in risk groups	Anticipated effects of trends in comorbid risk factors on TB	TB control interventions that link with systems for other high-risk conditions
Physical environment	Incarceration may have been a significant driver of community transmission [9]	Better estimates of location-specific TB transmission risk	Valid models for translating environmental heterogeneity into transmission risk	Active case finding targeted at high-risk environments (eg, prisons, transit)
	Greater proportion of infected contacts in less well-ventilated hospital wards [10]	Ability of specific interventions (eg, improved ventilation) to reduce that risk	Projected population-level impact of targeted environmental interventions	Mitigation of TB transmission through modification of high-risk built environments
Distal determinants	Ecological observation of declining TB rates during times of improvements in living standards [11]	Mechanistic linkages between poverty alleviation and TB transmission	Projected ability of social protection and similar efforts to reduce heterogeneity	Linkage between TB control programs and schemes to alleviate poverty and/or address other distal determinants
	Association between coverage of Brazil's conditional cash transfer program and improved TB control [12]	TB-specific effects of broader interventions	Models of the impact of TB on other outcomes in vulnerable populations	Implementation of TB interventions in a fashion that mitigates burden on the highest-risk populations, thus promoting equity and reducing disparities in risk

Abbreviations: HIV, human immunodeficiency virus; TB, tuberculosis; XDR-TB, extensively drug-resistant tuberculosis.

diabetes, malnutrition, and heavy alcohol use are critical considerations in the variation of risk of disease progression observed (Supplementary Bibliography). For example, HIV is the strongest individual-level risk factor and a major driver of the TB epidemic in many parts of Africa, while the rising global prevalence of noncommunicable diseases (eg, diabetes) may hinder our ability to achieve control targets by impairing host immunity at the population level [26]. History of exposure and disease are also important, as people who are latently infected may have partial protection against reinfection with the pathogen [27], whereas previously treated persons are likely to be at substantially increased risk for recurrent disease [28]. This latter increase in risk may reflect repeated exposure, incomplete treatment, or underlying immunological vulnerability [29].

The Physical Environment

The setting in which TB is transmitted is also an important modifier of spread—either due to increased population density or congregation of individuals with higher rates of specific risk factors, or directly through environmental features that facilitate airborne transmission. Characteristics of the physical environment that may contribute to transmission include crowding, poor ventilation, and high levels of indoor air pollution [30]. Furthermore, locations with these characteristics (eg, clinics,

public transit, churches, prisons, mines, and informal drinking spaces) are often frequented by the same high-risk individuals, further fueling heterogeneous transmission in these sites. These locations are themselves likely to be in close proximity, enhancing transmission in impoverished areas [31] and sustaining the epidemic [32].

Structural and Social Determinants

Heterogeneity at the community level is driven by a complex network of proximal and distal determinants that may not always be fully explained by quantifiable risk factors. Migration, urbanization, demographic transition, and other broad global trends combined with weak and inequitable policy and planning lead to pockets of poverty, unhealthy behaviors, and weak health systems in which TB thrives [33]. Social or spatial clustering of the individual-level characteristics described in the preceding sections may magnify the effect of these risk factors through transmission, as persons contact one another more if they share similar characteristics (assortative mixing). However, understanding of the effect of the various upstream determinants responsible for driving heterogeneity in TB burden is limited by the relative paucity of modeling studies in this area [34].

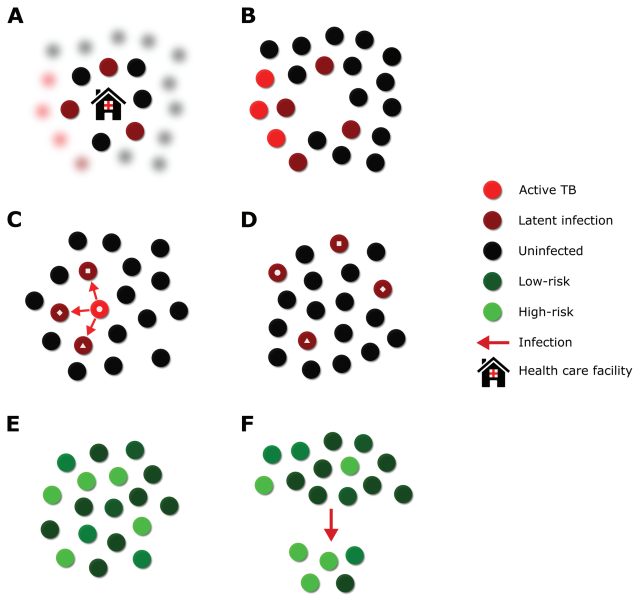


Figure 2. Illustration of some selected concepts from the text. *A*, Degree of heterogeneity that might be observed among individuals with good access to the healthcare system (unblurred discs) compared to those with poor access (blurred discs). This may be substantially less than the heterogeneity that exists in the population as a whole (*B*). *C*, Series of transmission events. *D*, Subsequent relocation of infected and uninfected individuals. This results in a more homogeneous distribution of infection across the population at this later time point, even though transmission was highly heterogeneous. *E*, Series of individuals at variable risk of infection. *F*, Selection of higher-risk individuals through the infection process. Although infection is the selecting illustrated process here, similar principles would apply to progression from infection to disease, through stages of the disease process and to interaction with the health system. Abbreviation: TB, tuberculosis.

CHALLENGES IN QUANTIFYING HETEROGENEITY

Although substantial between- and within-country differences in burden are frequently reported, challenges exist in interpreting the differences observed between demographic, geographical, or other subdivisions of the population. Our understanding of the population-level epidemiology of TB disease relies to a large extent on cases that have sought care, received a diagnosis, and been recorded through surveillance systems or local studies. The substantial proportion of cases that does not reach this stage in many settings [1] means that our estimates of heterogeneity in burden are prone to bias (Figure 2A and 2B). A particular consequence of relying on data from detected cases arises from the negative correlation between TB burden and access to care, which may mask heterogeneity in disease. For example, TB prevalence surveys consistently show a male predominance among adult TB cases, but this gender gap is much smaller in notifications—suggesting that men experience a higher burden but seek or access care at a lower rate than women [35]. Similar and even stronger unobserved effects, whereby mechanisms that increase risk of TB also decrease the probability of detection, may exist for features such as socioeconomic status or locality. Moreover, even if bias could be eliminated from health information systems, routinely collected data are not typically disaggregated beyond broad age categories, geographic regions, and drug resistance profiles, thereby limiting our ability to observe heterogeneity between smaller subpopulations without specifically designed studies.

Much less biased measures of disease burden are available from the recent increase in TB prevalence surveys. However, prevalence surveys in the general population are expensive undertakings and typically designed to yield a relative precision of 20%–25% [36], limiting their ability to discern patterns among subgroups or at the district/local level. Moreover, prevalence surveys are by design cross-sectional, meaning that they cannot provide information on heterogeneity through time without additional assumptions or repeated data collection.

One important consequence of detection bias is that clusters of notifications are difficult to interpret. Apparent hotspots of TB disease may represent either true areas of intense transmission or better diagnosis (via targeted campaigns or differential access to care), such that the areas of most intense transmission may be those with the highest notification rates in some settings and the lowest in others. Travel to access care may further exaggerate this process, creating artefactual aggregations of notifications. By contrast, heterogeneity in transmission may be masked by the often substantial latent period between infection and disease onset, during which infected individuals may relocate (Figure 2C and 2D). This process smooths disease distribution and obscures transmission chains, while the distribution of transmission and latent infection are even harder to observe in

an era when population-wide surveys of infection are no longer undertaken.

IMPLICATIONS FOR UNDERSTANDING AND MODELING TRANSMISSION

The impact of heterogeneity of infectiousness is influenced by characteristics of the infectious host and the organism being transmitted, and can be explored through its specific effects on the basic reproduction number, R_0 [37]. While the point estimate of R_0 is often emphasized as a measure of the expected number of secondary cases caused by an average index case in an infection-naive population, infectiousness may more appropriately be viewed as a probability distribution across a population of individuals, each with their own expected number of secondary cases. While superspreading is clearly observable in TB genomic studies [38], saturation of close contacts, whereby contacts occur primarily among individuals who have already been infected, may increase the importance of community transmission in high-burden settings [39].

When heterogeneity in susceptibility to TB exists, concerns regarding the assumption of a homogeneous population parallel concepts familiar in noncommunicable diseases, such as cohort selection and frailty models in survival analysis. As higher-risk individuals develop incident disease [28, 40], the incidence rate of a cohort may decline simply because those who remain susceptible have a lower average risk (Figure 2E and 2F). This process is disabled in models that collapse risk distributions to their mean values, leading to inaccurate simulations and biased predictions. Population-level heterogeneity in susceptibility can also induce thresholds near which small epidemiological changes will cause dramatic shifts in disease burden, leading to unanticipated effects of preventive interventions [41] and faster emergence of drug-resistant strains [42].

Any transition rate can be affected by cohort selection, as illustrated in Figure 3. Instead of the disease incidence process discussed above, consider a cohort of individuals with active TB comprised of 2 groups: fast and slow care seekers. As the faster care seekers leave the cohort earlier, the overall care-seeking rate will decline over time, even though it remains constant in each group. This process complicates estimation procedures and can be especially problematic in relation to rates of infection, which are proportional to the prevalence of infectious individuals and so part of a feedback loop. Moreover, epidemiological uncertainty around the most appropriate parameter values for transmission models means that multiple parameter sets may superficially replicate observed burden [43], which is particularly problematic for an endemic infection with a prolonged and unpredictable latency period.

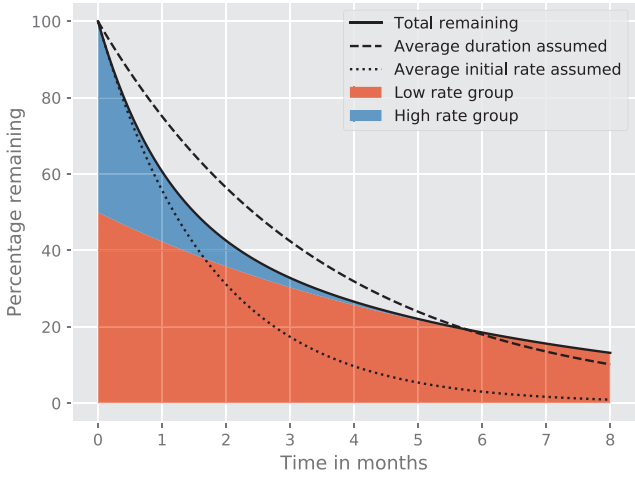


Figure 3. Composition of a simple 2-stratum heterogeneous cohort over time from entry to an epidemiological state (active undiagnosed tuberculosis). Plot displays the percentage of patients with active tuberculosis remaining undiagnosed after the onset of infectiousness (time 0 on the horizontal axis), under the assumption that 50% of the initial cohort has an average duration of infectiousness of 1 month (high-rate group), and 50% of the cohort has a duration of infectiousness of 6 months (low-rate group). The true total percentage of patients remaining infectious with time since onset of infectiousness (solid line) is compared against the proportion that would be expected to remain if the whole cohort was assumed to have the average time to diagnosis (3.5 months), and the proportion that would be expected to remain if the whole cohort was assumed to have a rate of diagnosis that is the average of the rates of the 2 groups (dotted line). The amount of the total population comprised of high-rate and low-rate persons at each time point is indicated by colored shading, demonstrating that the remaining cohort is increasingly comprised of low-rate individuals over time.

IMPLICATIONS FOR CONTROL

Targeting Risk Groups

A consequence of the heterogeneity in transmission, infection, incidence, and mortality is that benefits of interventions will differ depending on the groups targeted and the distribution of the risk factors introduced above. This consideration motivates much current TB policy, with groups at higher risk of infection, disease, or poor outcomes from TB episodes, such as household contacts, children, persons living with HIV, individuals with end-stage renal disease, and previously treated people identified as high-priority groups for screening and treatment of latent and active TB (Supplementary Bibliography). Heterogeneity in historical TB exposure is also a focus of interventions, with many low-incidence countries targeting services to foreign-born individuals [44], given their higher prevalence of latent TB and consequent risk of reactivation. However, interventions targeted at high-risk populations have not always been successful: A trial of mass screening and preventive treatment in South African miners had no impact on TB rates [45], because of reactivation of noncured infections and reinfection in the context of insufficient treatment and ongoing high environmental transmission risk [46].

Synergies With Non-TB Interventions

Regular interactions with the healthcare system for the management of chronic and noncommunicable diseases offer the opportunity for intensified case finding efforts, given that many such conditions increase TB risk or co-occur in populations

with such increased risk. More broadly, strengthening health systems for both TB and noncommunicable disease control provides the potential for synergistic interventions across diseases [47], while improving control by addressing distal determinants should also be a high priority [48]. The observation that both historical and more recent declines [11, 33] in TB burden have usually been achieved in the context of improvements in socioeconomic indicators highlights the importance of such upstream determinants and is particularly relevant in the era of the United Nations Sustainable Development Goals.

Geographical Targeting

TB incidence shows considerable geographical clustering at multiple resolutions [49], and spatial targeting of interventions has the potential to achieve major reductions in burden through focusing on geographically discernible TB hotspots [50]. However, the extent of mixing between hotspots and the broader population is important to quantify as it will modify the impact of such interventions [51]. Intensive TB control interventions targeted at Inuit communities in northern Canada, Alaska, and Greenland were effective at substantially reducing the extreme rates of TB incidence and mortality observed in the 1950s [52]. New and emerging analytic tools offer opportunities to identify and quantify TB hotspots, such as a recent genomic analysis in Peru that highlighted the spatial aggregation of multidrug-resistant genotypes [32].

Effect of Interventions on Heterogeneity

Where substantial reductions in TB burden are achieved, heterogeneity in TB distribution may increase, as transmission becomes more localized to remaining regions and population groups with fewer resources, limited healthcare access, and insufficient adherence to policy. However, even when fully implemented, control efforts may increase or decrease transmission heterogeneity depending on the intervention design. Interventions directed at those with poor access to care and thus high burden of disease may reduce heterogeneity, whereas interventions that strengthen routine programmatic management may increase heterogeneity even while decreasing overall burden. Heterogeneity may modify the impact of both targeted and untargeted interventions depending on the background burden of disease. For example, successful detection and treatment of a single active case may eliminate transmission from a community in a low-burden setting, whereas this would be harder to achieve in a high-burden setting. This may lead to unexpected relationships between control efforts and consequent reduction in the annual risk of *Mtb* infection [53].

Economic and Equity Concerns

The targeting of TB control interventions to those with high rates of infection or disease is expected to increase the effectiveness of interventions. Consequent gains in efficiency will depend on coverage levels, accessibility, disease prevalence, and contribution to transmission in the wider population of the target group. There are economies of scale to be achieved when increasing coverage, yet at high levels of coverage or for difficult-to-reach populations, targeted strategies may require additional supporting activities and so increase resource needs. For example, the cost-effectiveness of active case finding strategies is driven by both the heterogeneity in disease rates and in the cost of reaching different subgroups [54]. While maximizing impact within a given budget is a key objective in priority setting, heterogeneity in burden, healthcare access, and financial resources are linked to equity concerns in resource allocation for TB control strategies. Conceptually, the difference between inequalities and inequities is a value judgement about whether the observed heterogeneity is considered fair. Policy makers should seek to ensure that populations already experiencing increases in risk due to socioeconomic or other conditions (eg, crowding, incarceration) do not experience additional disparities in access to TB diagnosis and treatment, financial burden of illness, or unwarranted exposure to infection. While the reduction of such disparities is a key policy objective, there are situations in which achieving it may imply trade-offs in efficiency gains. For example, interventions aiming to place new technologies at decentralized locations may not be as cost-effective as placement at higher levels of the health system, yet may still be prioritized to reduce social inequities in financial burden, health outcomes, and access to health services [55].

WAYS FORWARD AND CONCLUSIONS

Causes of heterogeneity in TB epidemiology are diverse and include characteristics of the infectious host, pathogen, susceptible host, environment, and distal determinants—factors that may interact to amplify or reduce heterogeneity. Observed heterogeneity may not reflect reality and targeted epidemiological studies to quantify disease burden in more detail would be valuable, for example, prevalence surveys powered to obtain precise estimates of disease burden in specific population risk groups and age groups.

All TB modeling studies must judge which aspects of heterogeneity are sufficiently important to include given the question posed and the local context, and those which should not be specifically incorporated for parsimony. This highlights the importance of (1) detailed, context-specific data; (2) refining parameter estimation through epidemiological research; (3) communicating uncertainty in predictive modeling; and (4) confirmation of the predicted effectiveness and cost of interventions through operational research.

Heterogeneity has implications for the effectiveness and efficiency of control interventions. Targeting of interventions is an appropriate consideration in designing intervention strategies, although evidence to support specific targeted approaches is sometimes weak or contradictory. Therefore, such strategies must be considered in the context of resource availability and

the ethical imperative to ensure universal access to high-quality care. Moreover, it is also important to balance the need for clear guidelines that can facilitate the broad implementation of interventions at a national or global level against the importance of developing interventions that are targeted toward specific characteristics of regional or local epidemics.

As the global TB control community looks toward ending TB, understanding and harnessing heterogeneity to improve control will become increasingly important. Key considerations in addressing heterogeneity include better assessment of disease burden in population subgroups, context-specific modeling, targeting of interventions, and a focus on distal determinants of inequities in health status.

Supplementary Data

Supplementary materials are available at *Clinical Infectious Diseases* online. Consisting of data provided by the authors to benefit the reader, the posted materials are not copyedited and are the sole responsibility of the authors, so questions or comments should be addressed to the corresponding author.

Notes

Author contributions. J. M. T., P. J. D., Y. A. M., and D. W. D. coordinated this project. M. G. M. G. drafted the "Implications for Transmission Dynamics section". G. G. drafted the "Economic and Equity Concerns section". R. M. G. J. H. drafted the "Physical Environment section". N. A. M. drafted the "Effect of Interventions on Heterogeneity section". N. A. drafted the "Targeting Risk Groups section". S. S. drafted the "Infectious Host section". Y. A. M. drafted the "Ways Forward and Conclusions section". J. M. T. wrote the first draft of the full manuscript, which all authors reviewed and edited before submission.

Financial support. This perspective piece arose out of discussions of the Modelling Research Group at the annual meeting of the TB Modelling and Analysis Consortium, which is funded by the Bill & Melinda Gates Foundation. J. M. T. is a recipient of an Early Career Fellowship from the National Health and Medical Research Council, Australia (fellowship number APP1142638). R. M. G. J. H. received funding from the European Research Council under the European Union's Horizon 2020 research and innovation programme (grant agreement number 757699). P. J. D. is supported by the Medical Research Council, United Kingdom (fellowship number MR/P022081/1).

Potential conflicts of interest. All authors: No reported conflicts of interest. All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Conflicts that the editors consider relevant to the content of the manuscript have been disclosed.

References

1. World Health Organization. Global tuberculosis report 2017. Available at: http://www.who.int/tb/publications/global_report/en/. Accessed 15 January 2018.
2. Cohen T, Colijn C, Finklea B, Murray M. Exogenous re-infection and the dynamics of tuberculosis epidemics: local effects in a network model of transmission. J R Soc Interface **2007**; 4:523–31.
3. Feachem RGA, Phillips AA, Hwang J, et al. Shrinking the malaria map: progress and prospects. Lancet **2010**; 376:1566–78.
4. Gardy JL, Johnston JC, Sui SJH, et al. Whole-genome sequencing and social-network analysis of a tuberculosis outbreak. N Engl J Med **2011**; 364:730–9.
5. Valway SE, Sanchez MPC, Shinnick TF, et al. An outbreak involving extensive transmission of a virulent strain of *Mycobacterium tuberculosis*. N Engl J Med **1998**; 338:633–9.
6. Gandhi NR, Weissman D, Moodley P, et al. Nosocomial transmission of extensively drug-resistant tuberculosis in a rural hospital in South Africa. J Infect Dis **2013**; 207:9–17.
7. Verver S, Warren RM, Beyers N, et al. Rate of reinfection tuberculosis after successful treatment is higher than rate of new tuberculosis. Am J Respir Crit Care Med **2005**; 171:1430–5.
8. Nolan CM, Elarth AM, Barr H, Saeed AM, Risser DR. An outbreak of tuberculosis in a shelter for homeless men: a description of its evolution and control. Am Rev Respir Dis **1991**; 143:257–61.
9. Jones TF, Craig AS, Valway SE, Woodley CL, Schaffner W. Transmission of tuberculosis in a jail. Ann Intern Med **1999**; 131:557–63.
10. Ehrenkranz NJ, Kicklighter JL. Tuberculosis outbreak in a general hospital: evidence for airborne spread of infection. Ann Intern Med **1972**; 77:377.
11. Lienhardt C, Glaziou P, Uplekar M, Lönnroth K, Getahun H, Raviglione M. Global tuberculosis control: lessons learnt and future prospects. Nat Rev Microbiol **2012**; 10:407–16.
12. Nery JS, Rodrigues LC, Rasella D, et al. Effect of Brazil's conditional cash transfer programme on tuberculosis incidence. Int J Tuberc Lung Dis **2017**; 21:790–6.
13. Melsew YA, Doan TN, Gambhir M, Cheng AC, McBryde E, Trauer JM. Risk factors for infectiousness of patients with tuberculosis: a systematic review and meta-analysis. Epidemiol Infect **2018**; 146:345–53.
14. Marais BJ, Gie RP, Schaaf HS, et al. The natural history of childhood intra-thoracic tuberculosis: a critical review of literature from the pre-chemotherapy era. Int J Tuberc Lung Dis **2004**; 8:392–402.
15. Carvalho AC, DeRiemer K, Nunes ZB, et al. Transmission of *Mycobacterium tuberculosis* to contacts of HIV-infected tuberculosis patients. Am J Respir Crit Care Med **2001**; 164:2166–71.
16. Sepkowitz KA. How contagious is tuberculosis? Clin Infect Dis **1996**; 23:954–62.
17. Turner RD, Bothamley GH. Cough and the transmission of tuberculosis. J Infect Dis **2015**; 211:1367–72.
18. Lin X, Chongsuvivatwong V, Lin L, Geater A, Lijuan R. Dose–response relationship between treatment delay of smear-positive tuberculosis patients and intra-household transmission: a cross-sectional study. Trans R Soc Trop Med Hyg **2008**; 102:797–804.
19. Coscolla M. Biological and epidemiological consequences of MTBC diversity. Adv Exp Med Biol **2017**; 1019:95–116.
20. Gagneux S. Host-pathogen coevolution in human tuberculosis. Philos Trans R Soc B Biol Sci **2012**; 367:850–9.
21. Gagneux S, DeRiemer K, Van T, et al. Variable host-pathogen compatibility in *Mycobacterium tuberculosis*. Proc Natl Acad Sci **2006**; 103:2869–73.
22. Shrestha S, Knight GM, Fofana M, et al. Drivers and trajectories of resistance to new first-line drug regimens for tuberculosis. Open Forum Infect Dis **2014**; 1:ofu073.
23. Gagneux S, Long CD, Small PM, Van T, Schoolnik GK, Bohannan BJM. The competitive cost of antibiotic resistance in *Mycobacterium tuberculosis*. Science **2006**; 312:1944–46.
24. Denkinger CM, Pai M, Dowdy DW. Do we need to detect isoniazid resistance in addition to rifampicin resistance in diagnostic tests for tuberculosis? PLoS One **2014**; 9:e84197.
25. Knight GM, Colijn C, Shrestha S, et al. The distribution of fitness costs of resistance-conferring mutations is a key determinant for the future burden of drug-resistant tuberculosis: a model-based analysis. Clin Infect Dis **2015**; 61:S147–54.
26. Odone A, Houben RMGJ, White RG, Lönnroth K. The effect of diabetes and undernutrition trends on reaching 2035 global tuberculosis targets. Lancet Diabetes Endocrinol **2014**; 2:754–64.
27. Andrews JR, Noubary F, Walensky RP, Cerda R, Losina E, Horsburgh CR. Risk of progression to active tuberculosis following reinfection with *Mycobacterium tuberculosis*. Clin Infect Dis **2012**; 54:784–91.
28. Gomes MGM, Aguas R, Lopes JS, et al. How host heterogeneity governs tuberculosis reinfection? Proc Biol Sci **2012**; 279:2473–78.
29. Marx FM, Floyd S, Ayles H, Godfrey-Faussett P, Beyers N, Cohen T. High burden of prevalent tuberculosis among previously treated people in southern Africa suggests potential for targeted control interventions. Eur Respir J **2016**; 48:1227–30.
30. Lin H- H, Ezzati M, Murray M. Tobacco smoke, indoor air pollution and tuberculosis: a systematic review and meta-analysis. PLoS Med **2007**; 4:e20.
31. Murray EJ, Marais BJ, Mans G, et al. A multidisciplinary method to map potential tuberculosis transmission 'hot spots' in high-burden communities. Int J Tuberc Lung Dis **2009**; 13:767–74.
32. Zelner JL, Murray MB, Becerra MC, et al. Identifying hotspots of multidrug-resistant tuberculosis transmission using spatial and molecular genetic data. J Infect Dis **2016**; 213:287–94.
33. Lonnroth K, Jaramillo E, Williams BG, Dye C, Raviglione M. Drivers of tuberculosis epidemics: the role of risk factors and social determinants. Soc Sci Med **2009**; 68:2240–46.
34. Pedrazzoli D, Boccia D, Dodd PJ, et al. Modelling the social and structural determinants of tuberculosis: opportunities and challenges. Int J Tuberc Lung Dis **2017**; 21:957–64.
35. Horton KC, MacPherson P, Houben RMGJ, White RG, Corbett EL. Sex differences in tuberculosis burden and notifications in low- and middle-income countries: a systematic review and meta-analysis. PLoS Med **2016**; 13:e1002119.

36. World Health Organization. Tuberculosis prevalence surveys: a handbook. Geneva, Switzerland: WHO, **2011**.
37. Colijn C, Cohen T, Murray M. Emergent heterogeneity in declining tuberculosis epidemics. J Theor Biol **2007**; 247:765–74.
38. Ypma RJ, Altes HK, van Soolingen D, Wallinga J, van Ballegooijen WM. A sign of superspreading in tuberculosis: highly skewed distribution of genotypic cluster sizes. Epidemiology **2013**; 24:395–400.
39. McCreesh N, White RG. An explanation for the low proportion of tuberculosis that results from transmission between household and known social contacts. Sci Rep **2018**; 8:5382.
40. Gomes MGM, Barreto ML, Glaziou P, et al. End TB strategy: the need to reduce risk inequalities. BMC Infect Dis **2016**; 16:132.
41. Gomes MGM, Rodrigues P, Hilker FM, et al. Implications of partial immunity on the prospects for tuberculosis control by post-exposure interventions. J Theor Biol **2007**; 248:608–17.
42. Cohen T, Lipsitch M, Walensky RP, Murray M. Beneficial and perverse effects of isoniazid preventive therapy for latent tuberculosis infection in HIV-tuberculosis coinfected populations. Proc Natl Acad Sci U S A **2006**; 103:7042–7.
43. Ragonnet R, Trauer JM, Scott N, Meehan MT, Denholm JT, McBryde ES. Optimally capturing latency dynamics in models of tuberculosis transmission. Epidemics **2017**; 21:39–47.
44. Centers for Disease Control and Prevention. Reported tuberculosis in the United States. **2016**. Available at: https://www.cdc.gov/tb/statistics/reports/2016/default.htm. Accessed 26 March 2018.
45. Churchyard GJ, Fielding KL, Lewis JJ, et al. A trial of mass isoniazid preventive therapy for tuberculosis control. N Engl J Med **2014**; 370:301–10.
46. Vynnycky E, Sumner T, Fielding KL, et al. Tuberculosis control in South African gold mines: mathematical modeling of a trial of community-wide isoniazid preventive therapy. Am J Epidemiol **2015**; 181:619–32.
47. Pan S-C, Ku C-C, Kao D, Ezzati M, Fang C-T, Lin H-H. Effect of diabetes on tuberculosis control in 13 countries with high tuberculosis: a modelling study. Lancet Diabetes Endocrinol **2015**; 3:323–30.
48. Carter DJ, Glaziou P, Lönnroth K, et al. The impact of social protection and poverty elimination on global tuberculosis incidence: a statistical modelling analysis of Sustainable Development Goal 1. Lancet Glob Heal **2018**; 6:e514–22.
49. Shaweno D, Karmakar M, Alene KA, et al. Methods used in the spatial analysis of tuberculosis epidemiology: a systematic review. BMC Med **2018**; 16:193.
50. Dowdy DW, Golub JE, Chaisson RE, Saraceni V. Heterogeneity in tuberculosis transmission and the role of geographic hotspots in propagating epidemics. Proc Natl Acad Sci U S A **2012**; 109:9557–62.
51. Shaweno D, Trauer JM, Denholm JT, McBryde ES. The role of geospatial hotspots in the spatial spread of tuberculosis in rural Ethiopia: a mathematical model. Royal Soc Open Sci **2018**; 5:180887. doi:10.1098/rsos.180887
52. Grzybowski S, Dorken E. Tuberculosis in Inuit. Ecol Dis **1983**; 2:145–8.
53. Uys P, Marais BJ, Johnstone-Robertson S, Hargrove J, Wood R. Transmission elasticity in communities hyperendemic for tuberculosis. Clin Infect Dis **2011**; 52:1399–1404.
54. Azman AS, Golub JE, Dowdy DW. How much is tuberculosis screening worth? Estimating the value of active case finding for tuberculosis in South Africa, China, and India. BMC Med **2014**; 12:216.
55. Sachdeva KS, Raizada N, Sreenivas A, et al. Use of Xpert MTB/RIF in decentralized public health settings and its effect on pulmonary TB and DR-TB case finding in India. PLoS One **2015**; 10:e0126065.

END